Environmental Research Infrastructures in the Command and Control Anthropocene

Inauguraldissertation
zur Erlangung des Doktorgrades der Philosophie
an der Ludwig-Maximilians-Universität München

vorgelegt von
Nils Christopher Hanwahr

aus Göttingen

2018

Erstgutachter: Prof. Dr. Helmuth Trischler

Zweitgutachter: Prof. Dr. Christof Mauch

Tag der mündlichen Prüfung: 15. Februar 2019

ENVIRONMENTAL RESEARCH INFRASTRUCTURES IN THE COMMAND AND CONTROL ANTHROPOCENE

Acknowledgements

I want to thank Prof. Helmuth Trischler for his supervision. I also want to thank Prof. Christof Mauch for enabling me to participate in the Oberseminar in Schönwag.

Thanks to Gregg Mitman, Marco Armiero, and Robert Emmett for enabling me to participate in the Anthropocene Cabinet of Curiosities Slam in Madison, Wisconsin, which resulted a contribution to an edited volume:

Hanwahr, Nils C. 2018. Marine Animal Satellite Tags. In: Mitman, Gregg, Marco Armiero & Robert S. Emmett (eds.) *Future Remains*. A Cabinet of Curiosities for the Anthropocene. Chicago: University of Chicago Press.

Thanks again to Marco Armiero who allowed me to spend time at the Environmental Humanities Lab at KTH Stockholm in October 2017.

My research at the University of California, Berkeley, in October 2015, was supported by the LMU Fellowship program LMU-UCB Research in the Humanities.

This dissertation was supported by a scholarship of Studienstiftung des Deutschen Volkes.

A version of chapter 2 of this dissertation has been published as a journal article in the special issue "Big Data" of the NTM Journal, see:

Hanwahr, Nils C. 2017. „Mr. Database". Jim Gray and the History of Database Technology. *NTM Zeitschrift für Geschichte der Wissenschaften, Technik und Medizin* (25): 519-542.

Zusammenfassung in deutscher Sprache

Die vorliegenden Doktorarbeit mit dem Titel „Environmental Research Infrastructures in the Command and Control Anthropocene" betrachtet Forschungsinfrastrukturen in der Umweltforschung in den USA und ihr Potential, einen neuen digitalen Zugang zur Umwelt zu ermöglichen und so den Umgang mit des Auswirkungen des Klimawandels politisch zu gestalten und technologisch kontrollierbar zu machen. Dabei stellt sich die Frage, welche diskursiven und technologischen Hintergründe Narrative eines vernetzten digitalen Umweltbewusstseins ermöglicht haben und wie diese jenseits des Command and Control für politische Partizipation nutzbar gemacht werden könnten.

Das erste Kapitel beginnt mit dem Hack der Emails von Wissenschaftlern der University of East Anglia, der eine von ‚Klimaskeptikern' als „Climategate" bezeichnete Kontroverse im United Kingdom ausgelöst hatte. Climategate zeigte dabei, wie leicht die auf großen Datenmengen und Forschungsinfrastrukturen basierende Großforschung von der Öffentlichkeit missverstanden werden kann, die oft Vorstellungen von Forschung hat, die nicht der Wissenschaft im Zeitalter der Digitalisierung entsprechen. Die nationale Wissenschaftsakademie in England, die Royal Society, fürchtete daraufhin einen sich ausbreitenden Mangel an Verständnis und Akzeptanz für Wissenschaft, insbesondere die Klimaforschung, und berief eine Arbeitsgruppe ein, die Empfehlungen in einem Report mit dem Titel „Science as an Open Enterprise" veröffentlichte. Der Report befasste sich mit den Herausforderungen, die eine Wissenschaft basierend auf großen digitalen Infrastrukturen und massiven Datenbanken mit sich bringt. Aber datenbasierte Großforschung bring auch Chancen mit sich, wie die Open Science Bewegung es sich erhofft, in der Form von effizienter Verwendung von Forschungsgeldern sowie neuen Erkenntnissen, die sich durch eine neue Form der Datenanalyse, Big Data, gewinnen lassen. Die Studie zeichnet den Diskurs um Big Data nach sowie die verwandte Debatte darum, ob im

Zeitalter von Big Data eine neue Wissenschaft ohne Hypothesen möglich sei. Das Konzept der Fourth Paradigm Science des Datenbankeningenieurs Jim Gray wird vorgestellt, dem wir im Kapitel 2 noch begegnen werden. Die vielfältigen Metaphern, die sich an die Diskussion um Big Data Forschung hefteten verleiteten Ursula Heise, aus der Disziplin der Environmental Humanities, dazu, Datenbanken als die neuen ökologischen Epen des 21. Jahrhunderts zu bezeichnen. Es wird diskutiert, inwiefern Datenbanken überhaupt als Narrative fungieren könnten und welche historischen Voraussetzungen gegeben sein mussten, damit ein Vertrauen in Zahlen an sich entstand. Ein Beispiel einer Idee, mit einer Datenbank ein neues Narrativ zu schaffen wird mit Al Gores Idee der Digital Earth vorgestellt. Gore wollte mit dieser Digital Earth Infrastruktur sowohl die Digitalisierung voranbringen als auch globale Umweltprobleme lösen. Diese Idee, man könne ein neues politisches Bewusstsein durch die Engführung von digitalem, ökologischem und sozialem Netzwerk erschaffen wir zurückverfolgt zur Counterculture der 1960er Jahre, insbesondere dem Visionär Stewart Brand und seinem Whole Earth Catalogue.

Kapitel 2: „Mr. Database“ und die Geschichte der Datenbanktechnologien

Der Informatiker Jim Gray, im Silicon Valley bekannt als „Mr. Database“ bevor er 2007 auf See verschollen ging, war an vielen entscheidenden Entwicklungen seit den 1970er Jahren beteiligt, die die Basis für immer größere, schnellere und dezentralisierte Datenbanken bilden. Auf Grundlage der von Edgar F. Codd bei IBM konzipierten Prinzipien war Jim Gray an der Entwicklung von Relational Database Systemen beteiligt, und entwickelte später selbst Standards des Transaction Processing. Außerdem wirkte er daran mit, Austauschforen zwischen Wissenschaft und Industrie zu schaffen, die Funktionsstandards und Forschungsprogramme beeinflussten. Als Mitbegründer von *Microsoft Research* in San Francisco wandte sich Gray der wissenschaftlichen Anwendung von Datenbanktechnologien zu, etwa im TerraServer Projekt, einer Onlinedatenbank von Satellitenbildern. Inspiriert von Vannevar Bushs Idee des Memex entwickelte Gray seine Vision eines Personal Memex sowie eines World Memex, und postulierte letztlich ein neues Zeitalter der auf Daten basierenden wissenschaftlichen Entdeckung genannt „Fourth Paradigm Science". Dieses Kapitel gibt einen Überblick über Grays Beitrag zur Entwicklung von Datenbanktechnologien sowie seiner Forschungsagenda und zeigt, dass zentrale Ideen rund um Big Data die Akteure der technologischen Entwicklung schon sehr viel länger beschäftigten als der Begriff selbst in Verwendung ist. Die Idee der Fourth Paradigm Science war auch Inspiration für die Infrastrukturprojekte, welche in den folgenden Fallstudien genauer untersucht werden.

Kapitel 3: Fallstudie Ocean Observatories Initiative (OOI)

Die erste von zwei Fallstudien ist die Forschungsinfrastruktur Ocean Observatories Initiative (OOI) in den USA, ein Großprojekt der ozeanografischen Forschung finanziert durch die National Science Foundation. Zur Einführung wird der methodische Rahmen der Untersuchung vorgestellt, der sich auf die Arbeiten von Paul Edwards im Bereich ‚Knowledge Infrastructures' stützt sowie das Konzept der „Infrastructure Inversion" von Geoffrey Bowker. Aufbauend auf den Begriffen von Edwards und Bowker wird die vom Autor erweiterte ‚Infrastructure Matrix' vorgestellt, welche die Hauptakteure (Individuen, Institutionen und Artefakte) sowie deren Tätigkeiten (Knowledge Generation, Sharing und Maintenance), die Erzeugung, Verteilung und Erhaltung von Wissen in Infrastrukturen vorstellt. Als Hintergrund für die Fallstudie der Ocean Observatories Initiative werden dann historische Bezüge der Ozeanographie reflektiert, insbesondere die Verflechtungen dieser wissenschaftlichen Disziplin mit militärischer und nuklearer Forschung während des Kalten Krieges. Die Planung der Ocean Observatories Initiative wird anhand von Reports, Workshopberichten und Protokollen nachgezeichnet. Besondere Aufmerksamkeit gilt dabei dem Widerstreit zwischen dem Ziel der Infrastruktur, langfristige Dynamiken der Umweltveränderungen zu erfassen und zugleich, im besten Falle medienwirksame, singuläre Ereignisse zu beobachten und über ein digitales Observatorium der Öffentlichkeit zugänglich zu machen. Im Falle eines bevorstehenden Ausbruchs eines submarinen Vulkans lässt sich der Konflikt der Motivationen der verschiedenen Akteure besonders gut nachzeichnen. Abschließend wird diskutiert, inwiefern die OOI Forschungsinfrastruktur tatsächlich als „Fourth Paradigm Science" im Sinne von Jim Gray betrachtet werden könnte und welche Herausforderungen dies auch mit sich bringt.

Kapitel 4: Fallstudie National Ecological Observatory Network (NEON)

Die zweite der untersuchten Fallstudien ist das National Ecological Observatory Network (NEON), eine Forschungsinfrastruktur für Ökologie in den USA. Auch diese Fallstudie betrachtet die Hauptakteure in der Genese der Forschungsinfrastruktur und deren Tätigkeiten der Erzeugung, Verteilung und Erhaltung von Wissen über Ökologie und den Zustand der Umwelt in den Vereinigten Staaten. NEON wurde als Infrastruktur mit historischer Ambition und Ausmaß auf Initiative der National Science Foundation hin geplant. Das Netzwerk umfasst zahlreiche Beobachtungsstationen mit verschiedenen Messeinrichtungen über 20 klimatische Regionen der Vereinigten Staaten verteilt und hat den Anspruch, als Messgerät des Zustands der nationalen Ökosysteme der USA zu fungieren. Schon während Konzeption und Planung gab es Unstimmigkeiten zwischen den individuellen und institutionellen Akteuren. Deren Debatte um eine dezentrale Netzwerk-Struktur im Unterschied zu einer zentralisierten System-Struktur wir anhand von Quellen wie Workshop Berichten und Arbeitspapieren aber auch Anhörungen im US Kongress herausgearbeitet. Letztlich führten die Spannungen innerhalb der Planungsorganisation von NEON zum Bruch, und das Projektmanagement der Infrastruktur wurde einem privaten Unternehmen aus der Rüstungsbranche übertragen. Anhand von NEON zeigen sich so besonders gut die Konflikte zwischen dezentralen individuellen Akteuren und zentralisierten institutionellen Akteuren, mit denen Infrastrukturen der Fourth Paradigm Science allgemein zu kämpfen haben können.

Kapitel 5: Neue Narrative und Politik für Forschungsinfrastrukturen in der Umweltforschung

Das Schlusskapitel beschäftigt sich mit Überlegungen zu politischen Empfehlungen für Forschungsinfrastrukturen in der Umweltforschung. Es wurde gezeigt, dass viele dieser Infrastrukturen mit Narrativen über politische Partizipation und öffentliche Sicherheit aufgeladen sind, was dieses Kapitel zunächst erneut in Frage stellt. Es soll reflektiert werden, wie das aus Forschungsinfrastrukturen der Fourth Paradigm Science generierte Wissen, d.h. Daten aus virtuellen Observatorien, politisch nutzbar gemacht werden kann jenseits von Konzepten, die nach dem ‚Market Model' hauptsächlich maximalen wirtschaftlichen Wert generieren wollen. Zuerst stellt sich die Frage, in welcher Form eine Forschungsinfrastruktur überhaupt in den Prozess der evidenzbasierten Politikberatung eingebunden werden könnte. Dazu werden Konzepte aus der ‚Science Policy' vorgestellt, insbesondere die Idee des ‚Honest Broker' von Roger Pielke. Des weiteren stellt sich die Frage, welchen Standpunkt der Beobachtung, und damit welchen politischen Standpunkt, eine Forschungsinfrastruktur überhaupt einnehmen kann: generiert diese lokales oder globales Wissen? Die Möglichkeit, digitale Vermittlung könne überhaupt eine politische Bewegung formieren wird dabei in Frage gestellt und nochmals die Verwandtschaft der Forschungsinfrastrukturen zu Konzepten von Command und Control und ‚paranoiden' Überwachungstechnologien herausgestellt. Dem marktorientierten Modell des Informationsmarkts wird abschließend das ‚Polis Model' gegenübergestellt, um Ideen und Empfehlungen zu entwickeln, wie man Forschungsinfrastrukturen in der Umweltforschung auch tatsächlich für eine Gesellschaft nutzbar und zugänglich machen kann, die sich in den kommenden Jahrzehnten mit den Herausforderungen einer sich durch den Klimawandel rasch wandelnden Umwelt auseinanderzusetzen muss.

List of Abbreviations

AEC	Atomic Energy Commission
AIBS	American Institute of Biological Sciences
AOP	Airborne Observation Platform
ARENA	Advanced Real-time Earth Monitoring Network in the Area
ARGO	Array for Real-Time Geostrophic Oceanography
ATOC	Acoustic Tomography of Ocean Climate
AUV	Autonomous Underwater Vehicle
BCI	Biological Collections Institution
BDEI	Biodiversity and Ecosystem Informatics
BFS	Biological Field Station
BON	Biodiversity Observation Network
BTASC	Biodiversity Technology and Analysis Support Center
CI	Cyberinfrastructure
CSTB	Computer Science and Telecommunications Board
CyberPoP	Cyberinfrastructure Points of Presence
DARPA	Defense Advanced Research Projects Agency
DB	Database
DBTG	Data Base Task Group
DCAA	Defense Contract Audit Agency
EMSO	European Multidisciplinary Seafloor & Water Column Observatory
ERIC	European Research Infrastructure Consortium
ESONET	European Seafloor Observatory Network
FIU	Fundamental Instrumentation Unit
FSU	Fundamental Sentinel Unit
GARP	Global Atmospheric Research Program
GOOS	Global Ocean Observing System
GPS	Global Positioning System
HPTS	High Performance Transaction Systems
IBP	International Biological Program
IBRICS	Infrastructure for Biology at Regional to Continental Scales
IDOE	International Decade of Ocean Exploration
IGY	International Geophysical Year
IOC	Intergovernmental Oceanographic Commission
ION	International Ocean Network
IOOS	Integrated Ocean Observing System

ISEP	Integrated Science and Education Plan
LTER	Long Term Ecological Research
LUP	Land Use Package
MARS	Monterey Accelerated Research Systems
MBARI	Monterey Bay Aquarium Research Institute
MREFC	Major Research Equipment and Facilities Construction
MRP	Mobile Relocatable Platform
NASA	National Aeronautics and Space Administration
NCAB	National Center for the Analysis of Biodiversity
NEON	National Ecological Observatory Network
NEPTUNE	North-East Pacific Time series Undersea Networked Experiments
NOAA	National Oceanic and Atmospheric Administration
NOPP	National Oceanographic Partnership Program
NRAO	National Radio Astronomy Observatory
NRC	National Research Council
NSF	National Science Foundation
OOI	Ocean Observatories Initiative
OSB	Ocean Studies Board
PI	Principal Investigator
ROV	Remotely Operated Vehicle
SERDP	Strategic Environmental Research and Development Program
SOSUS	Sound Surveillance System
SQL	Structured Query Language
STEAC	Science, Technology and Education Advisory Committee
TOGA/TAO	Tropical Ocean Global Atmosphere/Tropical Atmosphere Ocean
TPC	Transaction Processing Performance Council
UNESCO	United Nations Educational, Scientific and Cultural Organization
UNIVAC	Universal Automatic Computer
USGS	United States Geological Survey
VPN	Virtual Private Network

List of Figures

Introduction

Command and Control in the Anthropocene

Command and Control was a central military principle of the Cold War specifying how the United States nuclear arsenal was at once to be made secure from failure and attack and always ready to retaliate against an enemy's nuclear strike.[1] In the 21st Century, the looming threat is no longer human or ideological, but how we respond to the impacts of a changing climate and a warming planet threatening more than just the unprecedented prosperity still prevalent in Western developed nations. Can one, and should one, command and control environments and manage changing ecosystems to ensure continuing prosperity and economic growth on this finite planet?

Since atmospheric chemist and Nobel Prize winner Paul Crutzen suggested the term "Anthropocene" in 2000 to define a new global era in which all environments are irrespectively impacted by human activity, emissions, waste, and technology, scholars from historians to geologists have debated how to measure the 'golden spike' of the onset of this new era called the "Anthropocene".[2] Humans are impacting the environment as well as all lifeforms and humans dwelling in it every day. While Command and Control was supposed to safeguard the United States from imminent nuclear annihilation, climate science has been at the forefront of deploying large scientific infrastructures, from satellite networks to buoys floating in the ocean,

---

[1] See Eric Schlosser, *Command and Control. Nuclear Weapons, the Damascus Accident, and the Illusion of Safety* (New York: Penguin Books, 2013).page 93. Originally, the concept of Command and Control was introduced by the infamous Air Force General Curtis LeMay.

[2] See e.g. Paul J. Crutzen and Eugene F. Stoermer, "The Anthropocene," *IGBP Newsletter* 41 (2000); Paul J. Crutzen, "Geology of Mankind," *Nature* 415 (2002). And numerous more recent publications on the Anthropocene. This study assumes no particular definition of the Anthropocene beyond the assumption that all terrestrial environments are now subject to anthropogenic influence which, in turn, affects human and non-human entities dependent on these environments.

to assess the state of the climate and predict its course of change due to human activity.[3]

Other environmental sciences such as ecology and oceanography have followed suit in planning large environmental research infrastructures to monitor the Anthropocene environment. Economists and ecologists meanwhile are suggesting responses to a changing climate in the Anthropocene to manage environments as systems and wield the powers of digital technologies to collect ever more data to surveil and control entire environments.[4]

How anthropogenic climate change impacts the environment is aimed to be measured with ever more accuracy to manage the impact these changes, in turn, are likely to have on humans living and working in environments. Environmental research infrastructures such as the Ocean Observatories Initiative and the National Ecological Observatory Network were constructed with the aim of surveilling and controlling Anthropocene environments and managing their continuing human use as well as the environmental risks arising from a changing climate. Just as notions of an ecological web of life arose from cybernetics ideas of the Cold War era, technologies deployed in current environmental research infrastructures build on military technologies and principles of surveillance and Command and Control emerging from the Cold War military-industrial complex.[5]

---

[3] Paul N. Edwards, *A Vast Machine. Computer Models, Climate Data, and the Politics of Global Warming* (Cambridge, MA: MIT Press, 2010).

[4] See e.g. Sarah E. Cornell et al., eds., *Understanding the Earth System. Global Change Science for Application* (Cambridge: Cambridge University Press, 2012).

[5] For the history of cybernetics see e.g. T. Rid, *Rise of the Machines: The Lost History of Cybernetics* (Scribe Publications Pty Limited, 2016).

The environmental critic Rob Nixon has called this kind of approach to managing environments under the influence of humans the "Command and Control Anthropocene".[6] After the setback of the United States exit from the promising Paris Agreement to limit climate change to 2 degrees Celsius, talk of 'preventing, mitigating, and adapting to' climate change has ceased and we appear to have settled for mitigation while slowly realizing that adaptation to the impacts of anthropogenic climate change is what we need to prepare for.[7] This attempted adaptation is meant to secure an inhabitable world well into the future and entails the ambition to manage environments and ecosystem services by constantly surveilling and monitoring their changing state. This ambition is rooted in the entanglement of cybernetics, ecology, computer science, military surveillance, neoliberal economics, and counterculture environmentalism. Historicizing environmental research infrastructures, as this study strives to do, is thus a key entry point to an understanding of the drivers of Command and Control Anthropocene, of the underlying ideologies as well as of the implications that managing the environment in such a framework has for policy making in democracies.

Over the course of this study, I will give an account of how the rationale for environmental research infrastructures has developed around ideas of Command and Control as well as narratives of national security and public participation. A closer analysis of the intersections of digital technology,

---

[6] "Species thinking, particularly when partnered with Silicon Valley-style technoexuberance, tends to sidestep thorny questions of representative governance. That tendency is evident in those we might call command-and-control Anthropocene optimists," see p. 12 in Rob Nixon, "The Anthropocene. The Promise and Pitfalls of an Epochal Idea," in *Future Remains. A Cabinet of Curiosities for the Anthropocene*, ed. Gregg Mitman, Marco Armiero, and Robert S. Emmett (Chicago: The University of Chicago Press, 2018).

[7] "Global Warming of 1,5 °C. An Ipcc Special Report on the Impacts of Global Warming of 1.5 °C above Preindustrial Levels and Related Global Greenhouse Gas Emission Pathways, in the Context of Strengthening the Global Response to the Threat of Climate Change, Sustainable Development, and Efforts to Eradicate Poverty," (Intergovernmental Panel on Climate Change, 2018).

Carolyn Kormann, "The Dire Warnings of the United Nations' Latest Climate-Change Report," *The New Yorker*, 8 October 2018.

especially database technology, and ideas of ecological systems can serve to interrogate the current development and discursive claims of both database technologists and systems ecologists as well as their relation to notions from economics, public health, and military surveillance.

Motivations for Studying Environmental Research Infrastructures

There are two major motivations for historicizing environmental research infrastructures in the Command and Control Anthropocene. First, history of science and technology helps us imagine alternate pathways ahead by being aware of contingent past choices that have led up to the present, yet that were themselves never strictly determined. Thus, history can act as an "honest broker" in informing policy by imagining alternative futures and unearthing the contingencies of the present's past.[8] In science policy, i.e. science aiming to inform policy, an erudite historical cherry-picking is often carried out.[9] Policy reports cite examples from the history of science and technology that are supposed to show how deeply rooted in the past a contemporary proposal is, while alternative pathways and failures are swept under the rug.

The second motivation for studying data-intensive research infrastructures arose out of my experience of working on the 2012 policy report "Science as an Open Enterprise" at the Royal Society Science Policy Centre in London.[10] The policy report was a response to the controversy surrounding the hack of climate scientists' email correspondence, 'Climategate', which was construed by 'climate skeptics' to have revealed systematic manipulation of data by scientists aiming to convey the impression of accelerated climate change. It made recommendations for the opening up of scientific data to the public to increase

---

[8] Roger A. Jr. Pielke, *The Honest Broker. Making Sense of Science in Policy and Politics* (Cambridge: Cambridge University Press, 2007).

[9] At least, that is my experience from working for Royal Society Science Policy Centre and the Union of German Academies of Science.

[10] "Science as an Open Enterprise," (London: Royal Society, 2012).

both transparency and the potential value generated from such data. "Science as an Open Enterprise" assumed that scientific research data was continuously expanding, needed to be managed by scientific institutions, and would, under the condition of maximum transparency, increase not just research output, but also bolster the public understanding and acceptance of scientific research results.

With this research project I aim to put some of the technological developments covered in "Science as an Open Enterprise" into a broader context. It gives an historical background of environmental research infrastructure projects in the United States, traces the history of database technology since the 1970s along the life and work of one particular computer scientist, Jim Gray, and provides case studies of two specific large-scale environmental research infrastructure projects covering the time since the 1990s. Yet this dissertation is not only a study in the history of a specific technology but also an opening statement to a debate on science policy options that are informed by the history of science and technology.

Outline and Central Questions

While the Command and Control Anthropocene is built on the convergence of notions from ecology and earth systems science with military principles of surveillance and control, it was the rapid development of digital technologies and database storage technology in particular that has made large environmental research infrastructures possible. The first relational database systems were developed in the 1970s. Since around the year 2000, the marketing term 'Big Data' has established itself as a label for an allegedly new kind of knowledge generation that leverages large databases, statistical analytics, machine learning, and digital observatories to create insights for businesses as well as scientists and policy-makers. Over the past decades, the ambition to generate this new kind of knowledge has informed the construction of large-scale infrastructures in climate science and environmental research. Some have called these technologically complex, decentralized, and digitally networked ways of gaining knowledge 'eScience', others like the database pioneer Jim Gray have called it "Fourth Paradigm Science".[11]

The first chapter of this dissertation is going to motivate the question of how data and environmental research infrastructures became a main tool in the Command and Control Anthropocene. I will ask how the notion of a converging ecological and digital network arose out of an entanglement of political and scientific ideas. The convergence of ecological and technological ideas has served as a foundation for the justification of investments in large environmental research infrastructures in the United States. Notions of Open Science, Big Data, and Fourth Paradigm Science were part of an emerging narrative framing how to address the impacts of a changing climate in the Anthropocene.

The second chapter focuses on Jim Gray's idea of Fourth Paradigm Science and outlines the development of database technologies since the 1970s. Gray's work, e.g. on the Microsoft TerraServer project, foreshadows the ambitions of the environmental research infrastructures in their aim to build an all-encompassing virtual observatory of the environment. Indeed, the ideas that inform the construction of digital environments for environmental assessment are best embodied in the person of Jim Gray. The course of his life traces the locations and developments that were central in leading up to two large data-driven environmental research projects, the Ocean Observatories Initiative

---

[11] Gabriele Gramelsberger, *Computerexperimente. Wandel Der Wissenschaft Im Zeitalter Des Computers*, Science Studies (Bielefeld: Transcript, 2010).

From Science to Computational Studies. Studies in the History of Computing and Its Influence on Today's Sciences (Zürich: diaphanes, 2011).

Tony Hey, Stewart Tansley, and Kristin Tolle, eds., *The Fourth Paradigm. Data-Intensive Scientific Discovery* (Redmont, Washington: Microsoft Research, 2009).

Cornelius Borck, "Big Data. Praktiken Und Theorien Der Datenverarbeitung Im Historischen Querschnitt.," *NTM Zeitschrift für Geschichte der Wissenschaften, Technik und Medizin* 25, no. 4 (2017).

(OOI) and the National Ecological Observatory Network (NEON), that constitute the two case studies of this dissertation.

The case study chapters examine two environmental research infrastructures more closely. The central question in the case studies is how the planners of knowledge infrastructures such as the Ocean Observatories Initiative (OOI) and the National Ecological Observatory Network (NEON) deployed narratives of what Jim Gray has called Fourth Paradigm Science to legitimize the use of their projects in promoting national security, public health, and political participation. By scrutinizing the genesis of these knowledge infrastructures, we are going to see conflicts between the various actors involved. I will trace the justificatory narratives of environmental research infrastructures serving as the main tools of a Command and Control Anthropocene in contrast to the way the planning, implementation, and running of the infrastructures played out in practice.

The concluding chapter turns to the question of public and political participation. What kind of policy-making dominates in the Command and Control Anthropocene? Are the hopes for national security and public participation promised by environmental research infrastructure projects and inspired by Fourth Paradigm Science really implying a model of policy-making that promotes neoliberal economics and reduces the individual to a user of a virtual environmental observatory. I will contrast this model with a more collaborative model of policy-making and ask how knowledge from environmental research infrastructures could be deployed in narratives that go beyond the paranoid logic of Command and Control.

The final chapter, therefore, is inspired by Jo Guldi and David Armitage's *History Manifesto*, in which they postulate:

> Put to the service of the public future, history can cut through the fundamentalisms of scientists and economists who preach elite control of wealth or scientific monitoring of all earth systems as the only possible way to avoid catastrophe. History can open up other options, and involve the public in the dialogue and reimagination of many possible sustainabilities.[12]

Following up on this challenge, this dissertation's concluding chapter attempts to outline and open up for alternatives the modern idea of avoiding, managing, and mitigating climate change by extensively monitoring earth system environments, a notion I call using Rob Nixon's term the "Command and Control Anthropocene". Thus, the normative project of this study is to open up a field of imagination for other forms of dealing with threatened ecosystems in a world where climate change mitigation and adaption to climate change are an undeniable fact.

The sources used in this study differ according to the focus of the chapters. The first chapter traces the discourse around Big Data, database technologies, and Jim Gray's notion of a Fourth Paradigm drawing on popular accounts, marketing reports, and newspaper articles. The second chapter outlines the history of database technologies along the life of Jim Gray based on sources from Gray's personal online archive as well as oral history accounts and working reports as well as established research in computing history.

The case studies on the Ocean Observatories Initiative and the National Ecological Observatory Network build, mainly, on comprehensive online archives. The project archives contain all major policy reports, workshop and meeting minutes, marketing brochures, and even slides of presentations given by infrastructure managers. In addition, reports by funding agencies such as the National Science Foundation were examined. In the case of NEON, testimonies from Congressional committees were also consulted. Lastly,

---

[12] J. Guldi and D. Armitage, *The History Manifesto* (Cambridge University Press, 2014).p. 56

magazine articles and newspaper reports concerning the environmental research infrastructures are considered as well and provide some critical assessment of the case study projects.

In sum, this dissertation thus sets out to consider three overarching questions. What were the discursive and technological prerequisites of notions of the Command and Control Anthropocene and Fourth Paradigm Science? How did narratives based on Fourth Paradigm Science play out in practice in the case studies of environmental research infrastructures? And, how can we go beyond the frame of a Command and Control Anthropocene in the realm of policy-making to reimagine the use of environmental research infrastructures for society in the future?

Chapter 1: Knowledge Infrastructures and the Command and Control Anthropocene

1.1 Open Science and Environmental Research Infrastructures

Climategate and Science as an Open Enterprise

On the 17th of November 2009, a server at the Climate Research Unit of the University of East Anglia was hacked and email correspondence between climate scientists was disclosed.[13] The hacked climate scientists' emails allegedly showed how researchers were falsifying data to mislead the world about the existence and impact of climate change. While the emails contained nothing more scandalous than scientists at work discussing how to wrangle the large amounts of data, 'skeptics' claimed to have found the 'smoking gun' to prove that climate change was nothing but an elaborate hoax. The public debate that followed came to be labeled as Climategate. While Climategate did not change anything about the scientific consensus on climate science, it was a valuable lesson in how the public could misunderstand and be led to misconstrue the process by which science turns data derived from vast environmental research infrastructures into valid scientific results.

Several investigations about the processing of data by researchers followed, although no scientist was eventually blamed for any wrongdoing.[14] What this episode showed, however, was that the ways in which science is conducted is easily misconstrued by so-called skeptics and frequently misunderstood by the lay public. Especially in the United Kingdom, the Climategate controversy led to a lot of soul-searching by climate scientists and the scientific community. The Royal Society national academy of sciences subsequently formed a working group to consider the challenges of public transparency and accountability in the face of massively data- and infrastructure-based science.

---

[13] See e.g. "Closing the Climategate," *Nature*, 18 November 2010 2010.

[14] See House of Commons Science and Technology Committee, "The Reviews into the University of East Anglia's Climatic Research Unit's E-Mails," ed. Government Office for Science (London2011).

Thus, a project initially called "Science as a Public Enterprise" was conceived that was to inquire how the scientific system could deal with the increase in research data while at the same time maintaining transparency and accountability towards the lay public and public research funders. The report was eventually published in 2012 under the title "Science as an Open Enterprise".[15]

The report title refers to a movement that calls itself "Open Science".[16] Open Science is an extension of ideas of the Open Source community and an example of how ideas and utopias of the hacker realm with respect to cyberspace based on a sovereignty-free social contract have spilled over into academic science.[17] The ideal of Open Science is total openness, the ability of everyone to access all scientific data, not just the final publication as well as computer and model code to be able to reproduce research results with full transparency. Open Science would also enable scientists to work in unprecedented collaborative ways by sharing their research data via online platforms, potentially ushering in a whole new way of doing science, as the quantum computing expert Michael Nielsen has argued in his book *Reinventing Discovery*.[18] This idea poses major challenges in practice. Open Science would require large investments in data storage, data curation, and data quality assurance measures to ensure storage and universal access. Thus, Open Science remains an idealized aspiration that has nevertheless raised many questions about the relation of science and society in an era of research based on large research infrastructures generating massive amounts of information.

---

15 "Science as an Open Enterprise."

16 See for example John Willinsky, "The Unacknowledged Convergence of Open Source, Open Access, and Open Science," *2005* (2005).

17 See for example John Perry Barlow, "A Declaration of the Independence of Cyberspace," https://www.eff.org/de/cyberspace-independence. Barlow published his declaration online on February 8, 1996.

18 Michael Nielsen, *Reinventing Discovery. The New Era of Networked Science* (Princeton: Princeton University Press, 2011).

From Open Science to Big Data Hype

While the Royal Society policy report "Science as an Open Enterprise" was not explicitly about Big Data, the phenomena of an unprecedented growth of scientific data and data-based research are closely related. The report recommended new investments in infrastructures for scientific data and data access, in the hope of both systemically strengthening the integrity of science through transparency as well as deriving the maximum possible economic value from public investment in research. These ideas found their echo on the European level as well. The EU Commission and its data commissioner Neelie Kroes picked up the recommendations and started to promote data infrastructures as a key component of the future competitiveness of the European research area.[19]

The idea behind Open Science, however, was not only transparency and openness. Open Science also aimed for the maximum return from public investments in scientific research and data generation. Since sharing data was supposed to maximize the value and knowledge that could be extracted from that data it echoes notions that were first thought up by data-driven businesses such as Google, as we will discuss later on.

Businesses had come to realize early on that all their transactions create heaps of data that offer insights into their own operations, their customers, and the environment that they operated in. The idea of Business Intelligence, just like Big Data, is not entirely new; for example, one of the leading global logistics companies, the United Parcel Service (UPS), founded an operation research unit analyzing supply route optimization as early as 1954.[20] However, the idea of

---

[19] "Riding the Wave. How Europe Can Gain from the Rising Tide of Scientific Data," in *High Level Expert Group on Scientific Data* (European Commission, 2010).

[20] Thomas H. Davenport and Jeanne G. Harris, *Competing on Analytics. The New Science of Winning* (Boston, MA: Harvard Business Review Press, 2007). p. 101

turning all available data into "data products"[21] was a new notion that many businesses have attempted to adapt from technology companies.[22] A major report that promised billions of dollars of untapped market potential was a 2011 publication by the analytics division of the McKinsey Global Institute.[23] This report presented Big Data as the 'new new thing', similar to the New Economy around the turn of the century, a technology that would enable everyone to tap vast market potentials.

The technology marketing consultancy firm Gartner publishes an annual 'hype cycle' of new and emerging technologies. The first time a technology related to Big Data appeared in their hype cycle was in 2008, when Gartner listed cloud computing as an emerging technology.[24] The following year, cloud computing was deemed to be at the "peak of inflated expectations" by Gartner, yet there was still no mention of Big Data.[25] The first time the term appeared was in 2011 when Gartner listed "'Big Data' and Extreme Information Processing and Management" as an emerging technology expected to take 2 to 5 years to market maturity.[26] In 2012, Big Data was also nearing the "peak of inflated

---

[21] Data products are products and services that result from mining and analyzing unstructured data to produce economic value. (e.g., Robert W. Gehl, "Sharing, Knowledge Management and Big Data: A Partial Genealogy of the Data Scientist," *European Journal of Cultural Studies* 18, no. 4-5 (2015).). As DJ Patil, former Head of Data Products and Chief Scientist of LinkedIn, remarks, creating data products is at the heart of social network businesses and the major corporate asset. He claims that "When the company sees what can be created with data, when it sees the power of being data enabled, you'll see data products appearing everywhere. That's how you know when you've won." (Dhanurjay Patil, *Building Data Science Teams* (Sebastopol, CA O'Reilly Media, 2011). p.18)

[22] Thomas H. Davenport, *Big Data @ Work. Dispelling the Myths, Uncovering the Opportunities* (Boston: Harvard Business Review Press, 2014); *Enterprise Analytics : Optimize Performance, Process and Decisions through Big Data* (Upper Saddle River, N.J.: Financial Times/Prentice Hall, 2013).

[23] James Manyika, Michael Chui, and Brad Brown, "Big Data: The Next Frontier for Innovation, Competition, and Productivity," (McKinsey Global Institute, 2011).

An updated report has also been published: Nicolaus Henke, Jacques Bughin, and Michael Chui, "The Age of Analytics: Competing in a Data-Driven World," (McKinsey Global Institute, 2016).

[24] Gartner, "Hype Cycle for Emerging Technologies, 2008," (Gartner, Inc., 2008).

[25] "Hype Cycle for Emerging Technologies, 2009," (Gartner, Inc., 2009).

[26] "Hype Cycle for Emerging Technologies, 2011," (Gartner, Inc., 2011).

expectations" that it should summit in Gartner's 2013 technology hype report.[27] The last time Big Data appeared in Gartner's hype cycle was in 2014 when the technology was entering a so-called "trough of disillusionment."[28]

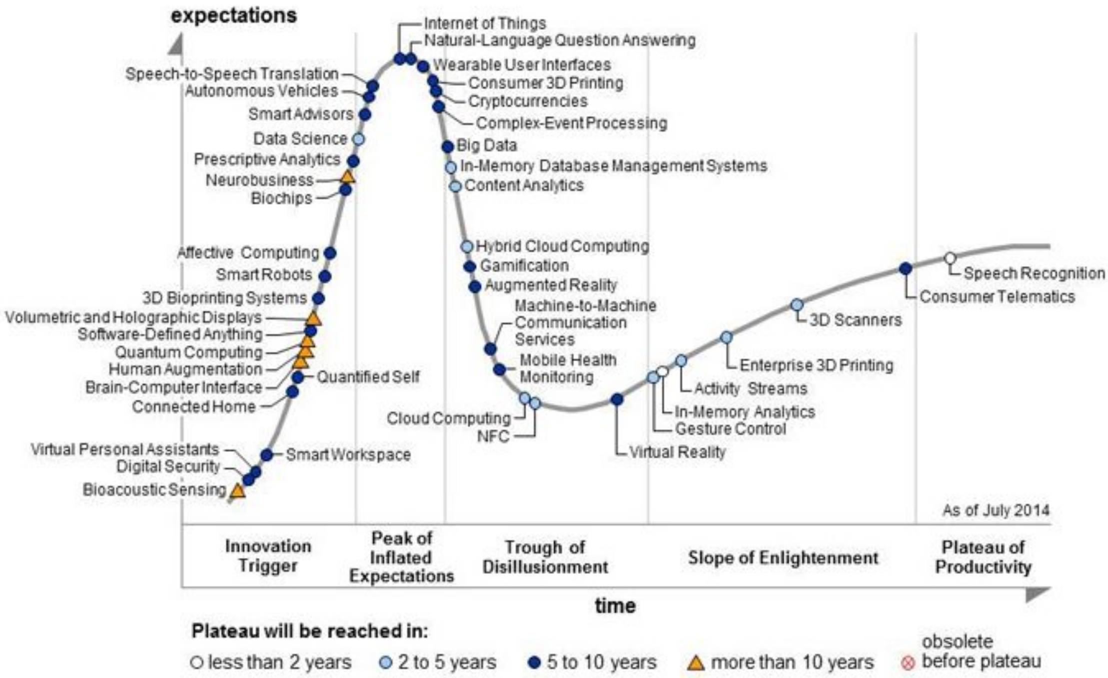


Figure 1 Gartner Hype Cycle 2014

The Gartner marketing reports show that Big Data was hard to identify as any one specific technology, rather it was a catch-all term for a development in the process of taking shape. The origin of the use of the term Big Data is also hard to pin down, although some trace it back to work by a certain John Mashey at a Californian company called Silicon Graphics in the late 1990s.[29] An economist named Francis Diebold has also claimed to have coined the term and has published several versions of a paper, attempting to track the term Big Data

[27] "Hype Cycle for Emerging Technologies, 2012," (Gartner, Inc., 2012); "Hype Cycle for Emerging Technologies, 2013," (Gartner, Inc., 2013).

[28] "Hype Cycle for Emerging Technologies, 2014," (Gartner, Inc., 2014).

[29] Steve Lohr, "The Origins of 'Big Data': An Etymological Detective Story," *The New York Times*, 1 February 2013 2013.

back to the year 2000.[30] And yet, it was only around the "year 2008, according to several computer scientists and industry executives, [...] when the term 'Big Data' began gaining currency in tech circles."[31]

How could a term that was at the peak of the technology marketing hype in 2012, less than ten years ago, disappear so completely from Gartner's marketing analysis? Big Data has never been a clear-cut concept but a term comprising various technologies such as databases, cloud storage, and data analytics. Some of these related technologies have eclipsed the term Big Data since 2014 and were already included in that year's hype cycle: data science, in-memory database management systems and analytics, and cloud computing.

Yet, it was progress made in database technologies since the 1970s (which will be outlined in chapter 2), cloud computing, and analytics that led many to believe that the early years of the 21st Century were witnessing a data-intensive information 'revolution' leading to a whole new way of knowing the world and doing science.

Should Science learn from Google?

In 2007, Mark Zuckerberg founded Facebook, while software engineers at Google developed a tool called *Hadoop* that enabled them to manage massive amounts of data distributed over physically disparate datacenters.[32] The following year, 2008, an article by former editor-in-chief of *Wired* Magazine, Chris Anderson, titled "The End of Theory"[33] captured the mood of the moment by making the provocative claim that science could now be done

---

[30] Francis X. Diebold, "A Personal Perspective on the Origin(S) and Development of "Big Data": The Phenomenon, the Term, and the Discipline," (2012).

[31] Steve Lohr, "How Data Became So Big," *The New York Times*, 11 August 2012 2012.

[32] Thomas Friedman has argued that 2007 was the 'beginning' of the tech-dominated world we live in now. Thomas L. Friedman, *Thank You for Being Late. An Optimist's Guide to Thriving in the Age of Accelerations* (New York: Farrar, Straus and Giroux, 2016).

[33] Chris Anderson, "The End of Theory: The Data Deluge Makes the Scientific Method Obsolete," *Wired Magazine*, no. 16.07.2008 (2008).

without theory, knowledge needed no hypotheses, and that massive amounts of data would speak for itself. "This is a world where massive amounts of data and applied mathematics replace every other tool that might be brought to bear," Anderson claimed. Pointing out that he was not simply making a point about how the advertising targeting algorithms at Google were functioning, but was asking for a new scientific epistemology, Anderson stated: "It's time to ask: What can science learn from Google?"[34]

Chris Anderson's provocation about the 'end of theory' caused a big splash in the popular discourse on technology, science, and business. Information scholar David Weinberger announced in his book *Too Big to Know,* that "knowledge is now a property of the network."[35] He joined Anderson in claiming that science was going to be transformed by the new data-intensive technologies and methods, musing that "we should expect the next Darwin [...] more likely to be a data wonk than a naturalist wandering through an exotic landscape."[36] Viktor Meyer-Schönberger and Kenneth Cukier from the Oxford Internet Institute chimed in and announced that "the world of big data is poised to shake up everything from businesses and the sciences to healthcare, government, education, economics, the humanities, and every other aspect of society."[37]

---

[34] Ibid.

[35] David Weinberger, *Too Big to Know. Rethinking Knowledge Now That the Facts Aren't the Facts, Experts Are Everywhere, and the Smartest Person in the Room Is the Room.* (New York: Basic Books, 2011). p. xiii

[36] Ibid. p. 195

[37] Viktor Mayer-Schönberger and Kenneth Cukier, *Big Data. A Revolution That Will Transform How We Live, Work and Think* (London: John Murray, 2013). p. 11

See also A. McAfee and E. Brynjolfsson, "Big Data: The Management Revolution," *Harv Bus Rev* 90, no. 10 (2012).

There were more cautious voices as well, tech insiders such as danah boyd and Katherine Crawford, or the tech industry critic Evgeny Morozov.[38] Eventually, even gilded publications such as *The Economist* and *Nature* published special issues on the "Data Deluge" and Big Data.[39] Everyone seemed to agree, something new was happening to the way we know things, yet, it seemed hard to grasp where this alleged revolution had come from.[40]

In fact, the ideas about applying large database systems to scientific inquiry were not without precedent. Taking a closer historical look at the life and legacy of a little known but highly influential computer scientist, Jim Gray, I will discuss next how someone who built much of the technology behind 'Big Data' has shaped the discourse about the concurrence of scientific knowledge and data-intensive methods of inquiry early on.

1.2 Jim Gray's Idea of Fourth Paradigm Science

Fourth Paradigm Science as the Convergence of Open Science and Big Data
On January 11, 2007, during the National Research Council's Computer Science and Telecommunications Board (CSTB) meeting at the heart of Silicon Valley in Mountain View, California, an appropriately nerdy looking 63 year-old man called Jim Gray, gave a talk titled "eScience: A Transformed Scientific Method". Gray was a central figure in the development of modern

---

38 danah boyd and Kate Crawford, "Critical Questions for Big Data," *Information, Communication & Society* 15, no. 5 (2012); danah boyd, "Privacy and Publicity in the Context of Big Data," in *WWW 2010* (Raleigh, North Carolina2010).

Evgeny Morozov, *To Save Everythink, Click Here. The Folly of Technological Solutionism* (New York: Public Affairs, 2013).

Kate Crawford, "Think Again: Big Data. Why the Rise of Machines Isn't All It's Cracked up to Be.," *Foreign Policy*, 10 May 2013 2013.

39 "The Data Deluge. And How to Handle It: A 14-Page Special Report," *The Economist*,, February 27 2010; "Big Data," *Nature* 455, no. Special Issue (2008).

40 Since then, the discourse has continued and is now much more critical, see e.g. C. O'Neil, *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy* (Crown/Archetype, 2016). and F. Foer, *World without Mind: The Existential Threat of Big Tech* (Penguin Publishing Group, 2017).

database technology and transaction processing. He received the prestigious Turing Award in 1998, and developed the TerraServer project at Microsoft, the largest database at the time and predecessor of contemporary mapping applications such as GoogleMaps.[41]

In his talk, Gray claimed that the development of database technologies had advanced so far that it would enable researchers to do science in a whole new way which he called "Fourth Paradigm Science".[42] Gray diagnosed that science was facing a 'data deluge', in a similar way that Chris Anderson would later suggest. Gray recommended extended funding by the CSTB for research into new tools for data capture, data curation, and data analysis. An extended version of Gray's talk was later published in an edited book by Gray's employer Microsoft Research titled *The Fourth Paradigm*. In the published version of the talk, he elaborated his ideas into a narrative of the 'paradigms' of the history of science and argued why eScience or Fourth Paradigm Science constituted a new paradigm.

According to Jim Gray, the history of science has seen four distinct paradigms in research, which we will discuss in more detail in chapter 2. Gray concluded in his talk: "The techniques and technologies for such data-intensive science are so different that it is worth distinguishing data-intensive science from computational science as a new, *fourth paradigm* for scientific exploration."[43] Jim Gray characterized The Fourth Paradigm of data exploration and e-Science as follows:

[41] The details of Gray's ideas on Fourth Paradigm Science as well as the crucial ways in which he shaped the development of database technologies deployed in the case studies on environmental research infrastructures are going to be covered in chapter 2. The chapter has also been published in the NTM special issue on Big Data, see Nils C. Hanwahr, "„Mr Database". Jim Gray and the History of Database Technology," *NTM Zeitschrift für Geschichte der Wissenschaften, Technik und Medizin* 25, no. 4 (2017).

[42] The presentation slides from Gray's talk can be found on his personal website: http://jimgray.azurewebsites.net/talks/NRC-CSTB_eScience.ppt

[43] Hey, Tansley, and Tolle, *The Fourth Paradigm. Data-Intensive Scientific Discovery*.p. xix

- "Data captured by instruments or generated by simulator
- Processed by software
- Information/knowledge stored in computer
- Scientist analyzes database / files using data management and statistics"[44]

Whether science as such has actually entered an era of a new Fourth Paradigm is a question beyond the scope of this study. However, Gray's description of Fourth Paradigm Science serves as an idealized model of doing science in what Rob Nixon has called the "Command and Control Anthropocene".[45]

The case study chapters 3 and 4 are going to consider two environmental research infrastructures, the Ocean Observatories Initiative (OOI) and the National Ecological Observatory Network (NEON), in the light of Jim Gray's ideas. Both of these environmental research infrastructures were described as instances of Fourth Paradigm Science in Gray's Microsoft Research book and much of his work laid the technological foundation for these infrastructure projects. In some instances, Jim Gray was personally involved in the conception and planning of the virtual observatories and cyberinfrastructures supporting these model instances of environmental research infrastructures in the Command and Control Anthropocene.

Together, the discourses around Big Data and Fourth Paradigm Science have not just postulated a new method of knowledge generation. A researcher doing science based on data exploration is still going to formulate an explanation afterwards, i.e. data can never really speak for themselves. Yet we need to consider how a new relationship between knowledge and data may still be lending itself to new narratives about what knowledge in Fourth Paradigm Science could be. Thus, before we examine the work and life of Jim Gray in more detail in chapter 2 and delve into the environmental research infrastructure case studies in chapters 3 and 4, we need to more thoroughly

---

[44] Ibid. p. xviii

[45] Nixon, "The Anthropocene. The Promise and Pitfalls of an Epochal Idea." p. 12

discuss how notions of data-intensive science originated and how 'data' was invested with so much power of command and control over the natural world.

1.3 Narratives of Data and Quantification

Data as a new Environmental Epic?

Daniel Rosenberg has retraced the usage of the term "data" in the English language and found that the rise of the concept of data "in the seventeenth and eighteenth centuries is tightly linked to the development of modern concepts of knowledge and argumentation."[46] Rosenberg stresses that data, originating from mathematics and theology, "was always a rhetorical concept" aiming to establish "that which is given prior to argument."[47] There are various competing rhetorical uses of 'data' in order to establish evidence as a given and, as Lisa Gitelman points out, every "disciplinary institution has its own norms and standards for the imagination of data, just as every field has its accepted methodologies."[48] Thus, data as a rhetorical strategy to establish what is given before the argument even begins can also be the foundation of a narration about what is the case and what is knowable.

Information studies scholars and information managers often relate the concept of data to a hierarchical order of data, information, knowledge, and wisdom.[49] Considering the many environmental metaphors for data such as

---

[46] Daniel Rosenberg, "Data before the Fact," in *„Raw Data" Is and Oxymoron*, ed. Lisa Gitelman (Cambridge, MA: MIT Press, 2013). p. 15

[47] Ibid. p. 36

[48] Lisa Gitelman, *"Raw Data" Is an Oxymoron* (Cambridge, MA: MIT Press, 2013). p. 3

[49] See Yaron Ezrahi, "Science and the Political Imagination in Contemporary Democracies," in *States of Knowledge. The Co-Production of Science and the Social Order*, ed. Sheila Jasanoff (London: Routledge, 2004).

Bruno J. Strasser and Paul N. Edwards, "Big Data Is the Answer ... but What Is the Question?," *Osiris* 32, no. 1 (2017).

"data deluge" or "data avalanche"[50], it should be noted that data is construed as a given beyond the comprehension of human knowledge. Big Data is often said to have sublime qualities, since it is not comprehensible to a human analytical mind and can only be tackled by computational methods.[51] According to the idea of data 'speaking for itself' by Anderson, once a threshold level of sublime massive size is reached, data research turns into a kind of 'revelation via computation', a 'burning bush of data' telling a myth about a deeper truth that a human mind may not have access to.[52]

The manifold metaphors used when talking about what we do with database systems has led the Environmental Humanities scholar Ursula Heise to call databases in the life sciences the "environmental epic" of our times.[53] However, a database alone does not tell a story, it is the practice in its genesis and use that provide a narrative. Contrary to claims about Big Data's potential for new ways of knowing, databases are set up with a hypothesis in mind, or at least an aim of what one wants to be able to do. Furthermore, design decisions about what to include and which technology and software to deploy frame the range of hypotheses, theories, and narratives that one can extract from the database. Although this narrative extraction may in turn be carried out by a computer

---

[50] The discourse on Big Data comprises a remarkable number of metaphors relating to data and data analysis that draw on images from the natural environment. Data is often described as a "field", a "mountain" or an "ocean" that is to be "mined", "dug through" or "surveyed". Large amounts of data are construed as natural disasters such as a "wave", a "deluge", and "avalanche".

[51] On sublime data see https://thenewinquiry.com/the-data-sublime/

Regarding data as sublime could be seen as an updated belief of the notion of the technological sublime: D.E. Nye, *American Technological Sublime* (MIT Press, 1996).

[52] For the relation between science and religion see also Yuval Noah Harari, "Dataism Is Our New God," *New Perspectives Quarterly* 34, no. 2 (2017).

[53] Compare Ursula K. Heise, *Sense of Place and Sense of Planet: The Environmental Imagination of the Global* (New York: Oxford University Press, 2008).

And see *Nach Der Natur. Das Artensterben Und Die Moderne Kultur* (Berlin: Suhrkamp Verlag, 2010). "Die Epen verschiedener Kulturkreise [...] suchten stets die Gesamtheit der ihnen bekannten Welt zu erfassen und in diesem Sinne sind die Datenbanken gewissermaßen die ökologischen Epen unserer Gegenwart.", p. 89

program that brings with it its own technological black box, as could be the case with automated data analytics.[54]

Nevertheless, data as a rhetorical strategy can form the basis of explanatory narratives promising to reveal a deeper level of truth about nature and the environment. Thus, we could expand Ursula Heise's notion of databases as narratives and conceive of databases as an electronic form of the old topos "The Book of Nature" – the ubiquitous metaphor that nature was written in an intelligible language from which to infer universal or divine knowledge.[55] Databases enable us, metaphorically, to read Nature as a kind of e-Book. One could call this the "remediation"[56] of a metaphorical medium, the Book of Nature, in a digital environment. In a similar vein, Bruno Strasser has suggested that large databases in the Life Sciences are a remediation of a museum of natural history, since they classify and order our taxonomies of the natural world. Strasser examined the first large scientific digital database in the Life Sciences, GenBank, founded in 1982, pointing out that a database can also be a site of experimentation and research, rendering new narratives about nature and ecology.[57]

Yet, nature is not always an intelligible Book of Nature, at times it can be chaotic and unfathomably wild. Technologies of observation and tracking in even the most remote settings providing the large volumes of continuous, long-term data are also attempts to make 'environments' accessible to human control

---

[54] A good example is the automated business intelligence and analytics software platform Tableau, see www.tableau.com

[55] On historically changing environmental metaphors see William J. Mills, "Metaphorical Vision: Changes in Western Attitudes to the Environment," *Annals of the Association of American Geographers* 72, no. 2 (1982).

On another use of the Book of Nature topos in science, see L.E. Kay, *Who Wrote the Book of Life?: A History of the Genetic Code* (Stanford University Press, 2000).

[56] Jay David Bolter and Richard Grusin, *Remediation. Understanding New Media* (Cambridge, MA: MIT Press, 2000).

[57] Bruno J. Strasser, "The Experimenter's Museum: Genbank, Natural History, and the Moral Economies of Biomedicine.," *Isis* 102 (2011).

and management.[58] Two central aspects of wild nature are that it is either not entirely known or not entirely manageable.[59] An area may be fully mapped but not entirely manageable because it is so wild and dangerous in landscape, flora, and fauna. An example of this might be the contemporary Amazonian jungle. Or, an environment may be a wilderness because it is not mapped, although the landscape and surrounding may not be unmanageably threatening in themselves. Here an example would be an unexplored grassland or prairie. Today, we increasingly see environments not as threatening wilderness to conquer but as resources to exploit and manage, especially in the school of thought of so-called "environmental services" and notions of sustainability.[60]

Can any environment still be a wilderness if it is monitored in real-time and not only entirely explored, but continually surveilled and always accessible via a virtual observatory? Only the physical threats seem to remain as a definition of wilderness while the representations of our planet have achieved to paint a picture of the 'whole earth'. Just as Cold War military commanders strived to command and control the nuclear threat, environmental research infrastructures seek to command and control the 'wild' effects of a changing climate.

[58] See Gregg Mitman, "When Nature Is the Zoo: Vision and Power in the Art and Science of Natural History," *Osiris* 11 (1996).

See also "Big Data and Natural Disasters: New Approaches for Spatial and Temporal Massive Data Analysis," *Computers & Geosciences* 115, no. Special Issue (2018).

[59] On ideas of wilderness see Roderick F. Nash, *Wilderness and the American Mind* (New Haven: Yale University Press, 1967).

For the relation of technology to wild animals see Etienne Benson, *Wired Wilderness: Technologies of Tracking and the Making of Modern Wildlife* (Baltimore: The Johns Hopkins University Press, 2010).

[60] See for example Stephen J. Jordan et al., "Accounting for Natural Resources and Environmental Sustainability: Linking Ecosystem Services to Human Well-Being," *Environmental Science & Technology* 44, no. 5 (2010); Richard B. Howarth and Richard B. Norgaard, "Environmental Valuation under Sustainable Development," *The American Economic Review* 82, no. 2 (1992).

The question of how we came to trust data and quantification as the basis for narratives about the state of the world as well as the controllability of the wild and chaotic, however, is the question about why we came to trust quantification in the first place.

Trust in Numbers

A hierarchy of data, information, knowledge (and sometimes wisdom) is often implicitly or explicitly assumed in information studies and computer science.[61] According to this assumption, data at the lowest level of the hierarchy is the basis of information; information creates knowledge and knowledge is eventually transformed into wisdom.[62]

Yet, how did we come to trust the leaps from quantified data to information which is then regarded as knowledge? When data, information or knowledge is transferred from one domain into another, from one discourse to another, if it passes through the engine of a scientific database, it is always undergoing a transformation that includes an incommensurable leap that requires a good deal of trust in the practice of the scientific method. There is a tendency among the proponents of data-intensive science to deny this transformational leap.[63] Bruno Latour calls this denial *Doppelklick*.[64] Double Click is the enticing notion of direct knowledge, of direct reference, correspondence, and access. We allegedly gain access to an entity merely by commanding it to unveil itself,

---

[61] For a critical discussion of the data-information-knowledge-wisdom (DIKW) hierarchy, see Jennifer Rowley, "The Wisdom Hierarchy: Representations of the Dikw Hierarchy," *Journal of Information Science* 33, no. 2 (2007).

[62] Compare Strasser and Edwards, "Big Data Is the Answer ... but What Is the Question?." Edwards and Strasser also cite examples of the use of the data, information, knowledge hierarchy from T.S. Eliot's poem *The Rock* as well as lyrics by Frank Zappa: "information is not knowledge, knowledge is not wisdom . . . and music is the best." p.1

[63] See for example C. L. Philip Chen and Chun-Yang Zhang, "Data-Intensive Applications, Challenges, Techniques and Technologies: A Survey on Big Data," *Information Sciences* 275 (2014).

[64] Bruno Latour, *Existenzweisen: Eine Anthropologie Der Modernen* (Frankfurt am Main: Suhrkamp, 2014).

as if opening a data file on a desktop computer by double clicking on the icon that represents the file.

Theodore M. Porter, a historian of science, discusses the problem of scientific objectivity and expert elites in his book *Trust in* Numbers. Outlining the history of how we came to trust quantification, numbers and data as epitomes of objectivity, Porter states that "the ideal of objectivity is a political as well as a scientific one." Don Worster and Evelyn Fox Keller are cited by Porter as saying that striving for objectivity implies "no small degree of alienation from nature," and that "the control of nature is also the control of the self."[65] Thus, there is a relation between the ideas of a scientifically and quantitatively informed political process and the idea of quantification of the self.[66]

In order to be able to quantify oneself as well as nature one has to take a step back and become an outsider, as Porter writes: "Unless you become like outsiders, you shall never enter the domain of quantitative science. The ultimate outsider is the machine, and it is rapidly becoming the greatest in the kingdom of quantification."[67] Following Porter's narrative of the rise of 'trust in numbers', statistics, and math as tools in shaping modern societies and politics as well as policies, we can think of the technologies of environmental research infrastructures as purportedly 'objective authority' that are nevertheless instantiations of certain political and social values. "For quantification is not an unmovable mover, or the product of a conspiracy, by which a culture has been overturned. It reflects values before it created them, and its massive expansion in recent times has grown out of a changing political culture."[68]

---

[65] Theodore M. Porter, *Trust in Numbers. The Pursuit of Objectivity in Science and Public Life* (Princeton: Princeton University Press, 1995).p. 74 f.

Another important source on the history of quantification is Alain Derosières, *Die Politik Der Großen Zahlen. Eine Geschichte Der Statistischen Denkweise* (Berlin: Springer, 2005).

[66] Compare on Gordon Bell and self-archiving in chapter 2 Alec Wilkinson, "Remember This? A Project to Record Everything We Do in Life," *The New Yorker*, May 28 2007.

[67] Porter, *Trust in Numbers. The Pursuit of Objectivity in Science and Public Life*.p. 85

[68] Ibid. p. 86

Indeed, political culture in the United States has changed since the invention of database management systems in the 1970s. Porter suggests that specific views of quantification as a legitimate safeguard of objectivity correspond to particular political views: "Critics, especially on the left, present the quantitative mentality as morally indefensible, an obstacle to utopia. Advocates have sometimes answered their opponents, but usually by defending the legitimacy of quantification as a way of knowing, not of organizing a polity and a culture."[69] Thus, is quantification, generating knowledge from gathering large amounts of quantified data, merely a tool to generate objective knowledge or is it also a tool to organize a society and shape policy according to its inherent notions?

Let us now take a closer look at how political ideas about databases and environmental issues came to converge first in the counterculture of the 1960s and later in the politician Al Gore who championed both environmentalism and an extension of digital infrastructures in the United States in the 1990s.

[69] Ibid. p. 73

1.4 Only Connect: Digital Earth and Environmental Consciousness

Whole Earth Watched over by Machines of Loving Grace[70]

The idea of 'whole-earth', the first photographs of the planet Earth in its entirety, taken from space by an Apollo mission, is often seen as an important turning point in the environmental imagination and a founding document of the countercultural environmental movement.[71] The credo of countercultural environmentalists in the 1960s United States was that everything is connected, whether it was ecosystems and environments, minds on psychedelic drugs, or digital networks of computers.

This systems thinking and utopian promise of counterculture environmentalism fused with the early hacker movement and culminated in the San Francisco Bay Area around Stewart Brand. Brand began publishing the 'Whole Earth Catalog' in 1968, a manual that encouraged an environmentally conscious and libertarian approach to self-actualization,[72] or what post-Marxist sociologists Richard Barbrook and Andy Cameron have called "cybernetic libertarianism". Barbrook and Cameron defined this "Californian Ideology" as a "bizarre mish-mash of hippy anarchism and economic liberalism beefed up with lots of technological determinism."[73]

---

[70] The poem is referenced in Adam Curtis' BBC television documentary of the same name. Compare the title poem "All watched over by machines of loving grace" in Richard Brautigan's poetry collection published in 1976:

I like to think / (it has to be!) / of a cybernetic ecology / where we are free of our labors / and joined back to nature, / returned to our mammal / brothers and sisters, /and all watched over / by machines of loving grace.

[71] Among others, Ursula Heise makes this claim: Heise, *Sense of Place and Sense of Planet: The Environmental Imagination of the Global.*

[72] Stewart Brand 1968. The Whole Earth Catalog. And for the last issue of the Whole Earth Catalog see Portola Institute, *The Last Whole Earth Catalog: Access to Tools* (Portola Institute, 1971).

[73] Richard Barbrook and Andy Cameron, "Californian Ideology," in *Crypto Anarchy, Cyberstates, and Pirate Utopias*, ed. Ludlow, Peter (Cambridge, MA: The MIT Press, 2001).

The communications scholar Fred Turner traces Brands ideas back to a reaction to the Free Speech Movement at UC Berkeley in 1964. Back then, Mario Savio and many other Berkeley students were objecting to the conformist and military culture of the Cold War United States while dreaming of 'putting themselves onto the levers of the machine.'[74] The movement around Stewart Brand and the Whole Earth Catalogue, baptized 'New Communalists' by Fred Turner, however, did not exit the Cold War machine of Command and Control, but decided to build their own machine, a different world and a different level of consciousness.

Fred Turner has also argued in his book *From Counterculture to Cyberculture* that Stewart Brand's work on the Whole Earth Catalogue informed the inception of the techno-libertarianism of the magazine *Wired* in the 1990s. "Mind and computation, economy and nature, the corporation and the individual [...] all mirrored one another, linked by the universal logic of cybernetics and by the New Communalist hope that new, non-hierarchical social forms might arise thanks to technologies of consciousness," Turner concludes.[75] Stewart Brand himself wrote in a *Time* magazine article in 1995, titled "We Owe It All to the Hippies," that the "real legacy of the sixties generation is the computer revolution."[76]

But the sixties generation did not only lay the foundation of techno-libertarianism. It should also influence United States politics in the 1990s and inform large environmental research infrastructures to this day, namely by imagining "the world as a series of overlapping information systems" and the

---

[74] "There's a time when the operation of the machine becomes so odious, makes you so sick at heart, that you can't take part! You can't even passively take part! And you've got to put your bodies upon the gears and upon the wheels...upon the levers, upon all the apparatus, and you've got to make it stop! And you've got to indicate to the people who run it, to the people who own it, that unless you're free, the machine will be prevented from working at all!" cited after Fred Turner, *From Counterculture to Cyberculture. Stewart Brand, the Whole Earth Network, and the Rise of Digital Utopianism* (Chicago: The University of Chicago Press, 2006).

[75] Ibid. p. 234

[76] Cited after ibid. p. 103

belief "that social, technological, and biological systems were in fact mirrors of one another."[77]

The Atari Democrats and Al Gore's Digital Earth

In the late 1980s, a group of Democratic politicians came to be known as the "Atari Democrats", a group of members of Congress who wanted to reinvigorate American business by information technology, but also saw digital technology as a way of addressing environmental problems and raise awareness through environmental knowledge. A *New York Times* article from 1989 titled "Greening the Democrats" described how the Atari Democrats led by then Congressman Al Gore took up the environmentalist cause.[78] Gore described his environmentalist background stating that "as I was entering high school, my mother was reading 'Silent Spring [...] The year I graduated from college the momentum was building for Earth Day. After Vietnam [...] the Club of Rome report came out and the limits to growth was a main issue."[79] The Atari Democrats believed "that their commitment to innovative use of markets and to the environment are complementary."[80] Market forces were to ensure the reconciliation of economic growth and environmental protection and to integrate environments with the rationality of the markets 'the environment' had to be turned into information.

The extension of digital infrastructures in the US in the 1990s promised two things at once: economic growth and environmental management. Al Gore was the key figure at the overlap of this nexus of digital infrastructure expansion and environmentalism. In a speech Gore gave in 1998 at the California Science Center, Los Angeles, he laid out his vision of digital access to the environment.

---

[77] Ibid. p. 250

[78] E. J. Dionne, "Greening of Democrats: An 80's Mix of Idealism and Shrewd Politics," *The New York Times*, June 14, 1989 1989.

[79] Ibid.

[80] Ibid.

His talk titled "The Digital Earth: Understanding our planet in the 21st Century" connected the promises of cognitive expansion via vast databases and remote-sensing to the social hope of protecting the environment. Yet, Gore also problematized the new level of data available about the global environment and the need to use it and make it available: "The hard part of taking advantage of this flood of geospatial information will be making sense of it - turning raw data into understandable information. Today, we often find that we have more information than we know what to do with."[81]

This was an expression of 'Big Data' before the term was even used in public discourse. Gore went on to claim that this new amount of data urged us to come up with new ways of dealing with the flood of data: "I believe we need a "Digital Earth". A multi-resolution, three-dimensional representation of the planet, into which we can embed vast quantities of geo-referenced data." He then continued to develop ideas about collaborative data use and even foreshadowing some of the ideas that Jim Gray would later call "Fourth Paradigm Science":

> Although some of the data for the Digital Earth would be in the public domain, it might also become a digital marketplace for companies selling a vast array of commercial imagery and value-added information services. It could also become a "collaboratory"-- a laboratory without walls — for research scientists seeking to understand the complex interaction between humanity and our environment.[82]

We will see in the case studies in chapters 3 and 4 how closely Gore's idea of Digital Earth mirrored the mission statements of environmental research infrastructures such as the Ocean Observatories Initiative and the National Ecological Observatory Network. How much Gore anticipated these infrastructural developments also becomes visible in Gore's list of technologies needed to realize the Digital Earth project, he names: computational science, mass storage, satellite imagery, broadband networks, interoperability, and

[81] Al Gore, "The Digital Earth: Understanding Our Planet in the 21st Century," (California Science Center, Los Angeles1998).

[82] Ibid.

metadata. Also, Gore elaborated that to realize the full potential of Digital Earth, "automatic interpretation of imagery, the fusion of data from multiple sources, and intelligent agents that could find and link information on the Web about a particular spot on the planet" would need to be developed.

In his book *Earth in the Balance* Gore expanded his idea of Digital Earth in his plans for a "Mission to Planet Earth". Gore's writing was steeped in metaphors that drew on the contrast of democracy and communism, on planned economies versus market capitalism, to explicate his notions of how a transformative new global environmental consciousness could arise. *Earth in the Balance* was originally published in 1992, and thus, the collapse of the Soviet Union would still have been present to Gore at the time. Gore's vision was a project that he promoted as a return of power to the people, "the Mission to Planet Earth should be a Mission by the people of Planet Earth," he wrote.[83] Just like democracy needed a free press, Gore regarded the transparent and free dissemination of information as crucial to his project of connecting a networked global environmental consciousness. Gore thus drew an explicit analogy of a free press and the importance of building a globally interconnected network of environmental information.

In another analogy, Gore stated that the Mission to Planet Earth involved the monitoring of environmental change in the same crucial way that emergency intensive care relies on monitoring the vital signs of a critical patient. Gore wrote, "the first step is collecting the kind of rudimentary information necessary to monitor the environment closely, just as hospital emergency rooms monitor the vital signs of patients receiving intensive care [...], and the instruments themselves could be designed to facilitate daily electronic 'polling' or data collection."[84]

[83] *Earth in the Balance. Forging a New Common Purpose.* (London: Earthscan, 2007). p. 356

[84] Ibid. p. 356

Gore's ideas drew heavily on cybernetics and Buckminster Fuller's notion of the Spaceship Earth.[85] Just like the artificial environment of a spaceship needs to be maintained within certain parameters that are constantly monitored to ensure human survival, Earth's life support systems have to be kept within a range suitable to human survival. Furthermore, the metaphor of "polling" with reference to environmental monitoring constructs the image of a 'parliament of things' that is consulted like an electorate of hybrid actors, as Bruno Latour has suggested, and yet automatically bypasses the kind of political deliberation Latour envisioned.[86] Thus, universal monitoring was supposed to achieve a kind of total information on two levels, concerning the state of the endangered environment, on the one hand, and the state of consciousness of the people and entities interacting with this environment, on the other hand.

Yet, Gore himself had little idea how to realize his Digital Earth ideas in practice. "Another difficulty with the current design of the Mission to Planet Earth is that no one yet knows how to cope with the enormous volume of data that will be routinely beamed down from orbit."[87] On the one hand, Gore envisioned a centralization system that would gather all available data on the global environment. On the other hand, the large volumes of data may have required decentralization to distribute storage capacities around a dispersed network. "Because of the unprecedented volume of data, it may also be necessary to disperse the means of storing and processing it much more widely," Gore stated.[88]

---

[85] On cybernetics and Spaceship Earth see N. Wiener, *Cybernetics or Control and Communication in the Animal and the Machine* (M.I.T. Press, 1961).

R.B. Fuller, *Operating Manual for Spaceship Earth* (Lars Muller Publishers, 2008).

S. Höhler, *Spaceship Earth in the Environmental Age, 1960-1990* (Taylor & Francis, 2015).

[86] Compare Bruno Latour, *Politics of Nature. How to Bring the Sciences into Democracy*, trans. Catherine Porter (Cambridge, MA: Harvard Univerity Press, 2004). and B. Latour, *Das Parlament Der Dinge: Für Eine Politische Ökologie* (Suhrkamp, 2010).

[87] Gore, *Earth in the Balance. Forging a New Common Purpose.* p. 358

[88] Ibid. p. 358

We are going to see in the case study chapters that the conflict between proponents of a centralized system and a dispersed network approach to building environmental research infrastructures is not easily resolved. In fact, Gore may even have touched upon a dilemma inherent to large knowledge infrastructures in general. Personally, he seemed to favor the distributed networks model, writing:

> The current plan is to bring all the data to a few large centers where they will be processed; somehow the results will then be translated into policy changes that are in turn shared around the world. [...] The alternative approach – or architecture – that I am recommending here is to distribute the information collecting and processing capability in a 'massively parallel' way throughout the world.[89]

Whether or not such a "massively parallel" architecture could work in practice was unclear. However, the reason for why Gore favored this model was that he regarded a distributed information network as the perfect fit for a free market economy, in contrast to a centralized planning system. A Digital Earth, in Gore's view, would foster a global democratic environmental consciousness by virtue of being an unregulated information processing infrastructure just as he believed "democracy, as a political system, and capitalism, as an economic system, work on the same principle and have the same inherent 'design advantage' because of the way they process information."[90]

Gore's ideas about the concurrence of capitalism, democracy, and the free flow of environmental information in a digital network were obviously informed by notions of liberal market capitalism. Yet, his plans for a Digital Earth infrastructure also ignored the practical reality of infrastructures.

Knowledge infrastructures, like all infrastructures, become most noticeable upon breakdown, when a failure reveals the complex underbelly of a service that was otherwise humming along just out of sight. The notion of 'polling' public opinion or users of Digital Earth would have to face the frictions created

---

[89] Ibid. p. 359 f.

[90] Ibid. p. 359

by the people, technologies, and institutions comprising that infrastructure. A distributed or 'massively parallel' infrastructure setup does not only empower users, as Paul Edwards objects, it "also gives you tensions: friction among the many connected systems; a parliament of instruments, individuals, and institutions, all with their potentials for disagreement, resistance, and revolt."[91] It is these kinds of frictions that we are going to trace in the case studies of two environmental research infrastructures to interrogate and understand notions of Fourth Paradigm Science and their relevance to policy-making. Next, we need to consider some key methodological concepts for the analysis of knowledge infrastructures.

1.5 Knowledge Infrastructures and Infrastructure Inversion

Historian of Science Paul N. Edwards has argued in his seminal book *A Vast Machine* that data is never entirely separable from the model and the infrastructure deployed to generate it. Even researchers themselves need so-called "infrastructure inversion", because if they do not understand the infrastructure, they do not understand the data.[92] Thus, Edwards argues, climate scientists have had to become historians of their own technological infrastructures in order to conduct "infrastructure inversion" and comprehend, interpret, and assemble their research data.

"Infrastructure inversion" is a notion coined by Geoffrey Bowker and refers to the process of understanding and questioning a knowledge infrastructure in order to 'understand' and validate its output. "Infrastructural inversion means recognizing the depths of interdependence of technical networks and standards, on the one hand, and the real work of politics and knowledge

[91] Edwards, *A Vast Machine. Computer Models, Climate Data, and the Politics of Global Warming*. p. 229

[92] Ibid.

production on the other," Bowker writes in his book *Sorting Things Out.*[93] Thus, one cannot understand the data produced by a knowledge infrastructure without putting it in the context of technologies, standard practices, political discourse, and the actors involved in running an infrastructure. Without infrastructure inversion, one can never fully evaluate the validity and meaning of data produced by, in our case, a large environmental research infrastructure.

Similarly, a central assumption of Edwards' book *A Vast Machine* is that without an infrastructure there is no data, and that, inversely, there are no data without retroactively interwoven models and infrastructures, "without models, there are no data."[94] Similar to Lisa Gitelman in her book *Raw Data is an Oxymoron*, Edwards upholds that "data aren't data until you have turned their infrastructure upside down to find out how it works."[95]

However, before going any further, we need to discuss the crucial definition of a "knowledge infrastructure," which is defined as follows in ny Edwards: "*Knowledge infrastructures comprise robust networks of people, artifacts, and institutions that generate, share, and maintain specific knowledge about the human and natural worlds."*[96]

The key components of a knowledge infrastructure are thus threefold: people, artifacts, and institutions. These three kinds of actors will be the central categories of analysis in the case studies to be considered in chapter 3 and 4. Who were the people to promote an idea in the discourse, build the networks of people to support it, and ultimately build an infrastructure? What were the artifacts, i.e. mostly technologies and machines, that were essential in realizing the network of technologies to actually construct the knowledge infrastructure?

---

[93] Geoffrey C. Bowker and Susan Leigh Star, *Sorting Things Out. Classification and Its Consequences* (Cambridge, MA: MIT Press, 1999). p. 34

[94] Edwards, *A Vast Machine. Computer Models, Climate Data, and the Politics of Global Warming*.p. xiii

[95] Ibid. p. 20

[96] Ibid. p. 17 (italics in original)

And finally, what were the institutions that were essential to the infrastructure not only coming into being, but, which institutions' long-term existence is central to the maintenance and running of a specific knowledge infrastructure?

For Edwards, these networks are central to the claim to objectivity that data and knowledge produced by knowledge infrastructures can put forth, since if you "[g]et rid of the infrastructure [...] you are left with claims you can't back up, facts you can't verify, comprehension you can't share, and data you can't trust."[97] However, knowledge infrastructures do not only produce claims of objectivity, they also produce narratives that justify their existence and the funding invested in them. In particular, we are going to see how the environmental research infrastructures described in the case study chapters are highly invested with narratives of public participation in political decision-making as well as national security in the United States' response and adaptation to the impacts of climate change on national environments. Infrastructure inversion is going to help unearth these narratives, question their validity, and outline entailing challenges in the construction of large knowledge infrastructures.[98]

Stressing the importance of infrastructure inversion and pointing out that the idea of "raw data" is oxymoronic also implies a critique of the advocates of so-called Open Science.[99] Campaigning for access to all raw data would not make sense if one accepted the claim that there is no such thing as raw data. This underlines the importance of the issue of metadata, how data is provided,

---

[97] Ibid. p. 19

[98] Strasser and Edwards, "Big Data Is the Answer ... but What Is the Question?."

Paul N. Edwards et al., "Knowledge Infrastructures: Intellectual Frameworks and Research Challenges. Report of a Workshop Sponsored by the National Science Foundation and the Sloan Foundation, 25-28 May 2012," (**University of Michigan School of Information** 2012).

The MPIWG in Berlin also has a working group called ‚Historicizing Big Data'. A special journal issue with the work has been published: Elena Aronova, Christine von Oertzen, and David Sepkoski, eds., *Data Histories*, vol. 32, Osiris (Chicago: University of Chicago Press, 2017).

[99] Gitelman, *"Raw Data" Is an Oxymoron.*

curated, annotated, and standardized.[100] Yet, one might also ask how so many technologists come to belief in the dissipation and transparency of 'raw data' when they should be aware of how many processes are involved in any kind of data generation. As Paul Edwards claims, it should be the constructors of knowledge infrastructures who are most aware of the 'constructedness' of data, since deconstructing infrastructures, i.e. infrastructure inversion, is essential to understanding and interpreting the data they themselves have generated.

Before we move on to apply Edward's ideas in the two case study chapters covering the Ocean Observatories Initiative (OOI) and the National Ecological Observatory Network (NEON), we are going to take a closer look at the history of database technologies. The notion of Fourth Paradigm Science had been suggested by the database engineer Jim Gray, who was involved in developing many of the technologies that form the basis for operating large environmental research infrastructures today. Well-connected in the scientific community, Gray also provided input to some aspects of the planning of OOI and NEON, and thus forms the link between our discussion of Fourth Paradigm Science in the Command and Control Anthropocene and the case studies on environmental research infrastructures.

---

[100] "Science as an Open Enterprise."

Chapter 2: "Mr. Database" Jim Gray and the History of Database Technologies

The database engineer Jim Gray, known as "Mr. Database" in Silicon Valley before his disappearance at sea in 2007, was involved in many of the crucial developments since the 1970s that constitute the foundation of exceedingly large and distributed databases as well as many of the technologies used in the environmental research infrastructures described in the case study chapters 3 and 4.

Jim Gray, whom we have already encountered in chapter 1, was involved in the development of relational database systems based on the concepts of Edgar F. Codd at IBM in the 1970s before he went on to develop principles of Transaction Processing that enable the parallel and highly distributed performance of databases today. He was also involved in creating forums for discourse between academia and industry, which influenced industry performance standards as well as database research agendas. As a co-founder of the San Francisco branch of *Microsoft Research*, Gray increasingly turned toward scientific applications of database technologies, e.g. leading the TerraServer project, an online database of satellite images.

Inspired by Vannevar Bush's idea of the memex, Gray laid out his vision of a Personal Memex as well as a World Memex, eventually postulating a new era of data-based scientific discovery termed 'Fourth Paradigm Science'. This chapter gives an overview of Gray's contributions to the development of database technology as well as his research agendas and shows that central notions of Big Data and data-based ways of doing science have been occupying database engineers for much longer than the term has been in use.

In 2007, Jim Gray was lost at sea while sailing his ship *Tenacious* off the coast of San Francisco.[101] After the US Coast Guard had to abandon the search for his sailing yacht, many of Gray's friends in the database community attempted to find him using just the distributed database systems that Gray had helped to develop. The story of the search and the methods used are aptly discussed in an article by Gray's colleague Joe Hellerstein.[102] Yet, all the technology could not locate neither the man nor his ship, and Gray was eventually pronounced dead in absentia in 2012. A "Tribute to Jim Gray" was held at UC Berkeley in 2008 by Gray's family and former colleagues, whose contributions were published as a special issue of the journal SIGMOD Record.[103] Certainly, the contributions to this tribute volume have to be regarded as the eulogies that they are, yet nevertheless, they contain valuable and highly personal information. It is difficult to find sources that are directly critical of Gray and his work, which could indicate that it is still too soon for a truly critical assessment. This chapter attempts such an assessment or at least a contextualization of Gray's work and positions.

2.1 How do you know?

In January 2003, database engineer Jim Gray released a memo titled "How do you know?" to his colleagues at *Microsoft Research* in San Francisco. The memo was a meditation on what Gray's own work on database technologies had aimed to accomplish: "Wouldn't it be nice if we could just put all the books and journals in a library that would automatically organize them and start

---

[101] Katie Hafner, "Silicon Valley's High-Tech Hunt for Colleague," *The New York Times*, February 3 2007.

Steve Silberman, "Inside the High Tech Hunt for a Missing Silicon Valley Legend," *Wired*, July 24 2007.

[102] Joseph M. Hellerstein and David L. Tennenhouse, "Searching for Jim Gray: A Technical Overview," *Communications of the ACM* 54, no. 7 (2011).

[103] "Tribute to Honor Jim Gray," *ACM Sigmod Record* 37, no. 2 (2008).

producing new answers?"[104] Not only were databases supposed to store information in digital form, Jim Gray also wanted them to automatically generate creative ways of compiling the trove of knowledge into novel assemblages of insight.

He went on to ask: "How can knowledge be represented so that algorithms can make new inferences from the knowledge base? This problem has challenged philosophers for millennia. There has been progress."[105] While the claim that the representation of knowledge in a form that renders itself useful for computation has been an issue of philosophy for thousands of years is overstated, it has certainly been a challenge that led computer scientists to develop tools that are today assembled under the heading of Big Data. Progress has indeed been made in representing knowledge in forms accessible to algorithms, and yet this progress has a history that is closely related to the life of the author of the "How do you know?" memo, Jim Gray.

"Database researchers have labored to make it easy to define the schema, easy to add data to the database, and easy to pose questions to the database," Gray went on to write in his memo.[106] By 2003, the issues of sorting, indexing, and organizing information had essentially been solved by deploying relational database management systems that are widely used in science and business applications to this day.[107] Jim Gray summed up the development of relational

---

[104] Jim Gray, "How Do You Know?," in *http://jimgray.azurewebsites.net/default.htm*, ed. Microsoft Research (Microsoft Research, 2003).

[105] Ibid.

[106] Ibid.

[107] A good introduction to data mining and its history is Matthew L. Jones' 2017, "Querying the Archive: Data Mining from Apriori to PageRank" in Lorraine Lorraine Daston, ed. *Science in the Archives. Pasts, Presents, Futures* (Chicago: The University of Chicago Press, 2017).

A seminal reference on databases in science is Geoffrey C. Bowker, *Memory Practices in the Sciences* (Cambridge, MA: MIT Press, 2005). Useful overviews are e.g. Thomas J. Bergin and Thomas Haigh, "The Commercialization of Database Management Systems, 1969-1983," *Annals of the History of Computing* IEEE 31, no. 4 (2009). as well as Thomas Thomas Haigh, "How Data Gots Its Base: Information Storage Software in the 1950s and 1960s," ibid. and Avi Avi Silberschatz, Michael Stonebraker, and Jeffrey D. Ullman, "Database Systems: Achievements and Opportunities," *ACM Sigmod Record* 19, no. 4 (1990).

databases based on the ideas postulated by Edgar F. Codd in 1970, "the research community embraced the relational data model championed by Ted Codd. [...] After a decade of experimentation, these research ideas evolved into the SQL database language."[108] Next, database engineers had to address the issues of how to build a database that could be spread out over various storage media, be accessed by multiple queries in parallel, and still be reliable at a level that enables one to put trust in making purchases online or carrying out financial transactions via online banking. Yet, Gray's framing of progress in database technology overlooks a more complicated history than his memo suggests.

This chapter tells the story of how Jim Gray was involved in creating database technologies that allow us to sort, index, and organize information, and then went on to develop principles of transaction processing that ensure the concurrency and reliability of databases. Concluding his memo, Jim Gray wrote "Over the last decade, the traditional database systems have grown to include analytics (data cubes), and also data mining algorithms borrowed from the machine learning and statistics communities."[109] Eventually, the aim of creating databases that allow for new knowledge to be gained by applying algorithms began to be realized by deploying a combination of machine learning and database technology that we often call Big Data.

Jim Gray was also actively involved in selling a narrative of linear progress in the development of database technologies that he deployed to influence research agendas, omitting the frustrations and dead-ends of research and technology development. To trace both Jim Gray's work as well as his influence on the discourse among the database technology community, this chapter draws on several original sources. Gray himself made available much of his personal and professional communication such as memos, technical reports, workshop presentations, and conference talks on his personal website.

---

[108] Gray, "How Do You Know?."

[109] Ibid.

Furthermore, as a tireless networker and mentor of many computer scientists, he was connected to numerous people in Silicon Valley, who, in turn, make frequent references to Gray in both oral history interviews as well as interviews in newspapers and the trade press.

To assess Gray's impact and influence on discourse and technology development, I draw on several of these public sources. Jim Gray believed that database technology held the promise to change the way knowledge comes into the world, an idea he called "Fourth Paradigm Science". I will also attempt to trace some steps in Gray's work such as his concepts of Transaction Processing, his work on the Microsoft TerraServer, and his ideas of eScience to put current debates and claims about the powers and promises of Big Data, whether in commerce or environmental science, into a broader perspective.

2.2 Mr. Database and Mr. Memex

First, we should remind ourselves of the source of the idea to create a universal library comprising automated knowledge, accessible to everyone, and capable of generating new insights algorithmically, which was echoed in Jim Gray's "How do you know?" memo. Many computer scientists have been fascinated by a concept that Vannevar Bush, the Director of the Office of Scientific Research and Development during the US postwar years, had developed in an article titled "As We May Think" in the July 1945 Issue of *The Atlantic* magazine: the memex.[110]

Jim Gray included Vannevar Bush's article at the top of a recommended readings list on his personal website and frequently referred to Bush's ideas.[111] "As We May Think" addressed the swift expansion of information and information technology that had taken place during the Second World War: "Science [...] has provided a record of ideas and has enabled man to manipulate

[110] Vannevar Bush, "As We May Think," *The Atlantic*, July 1945.

[111] http://jimgray.azurewebsites.net/ section "Recommended Articles"

and to make extracts from that record so that knowledge evolves and endures throughout the life of a race rather than that of an individual."[112] The scientific record, of course, can also be expanded via the media of writing, books, and libraries. However, Bush envisioned such a rapid growth of information, that new technological means are necessary to store and consult the ever-expanding record of knowledge.[113]

Bush was focused on analog storage media such as microphotography rather than digital storage media, and yet, his idea has inspired much of the work of Jim Gray up to his talk on Fourth Paradigm Science in 2007,which will be discussed later on.[114] Remarkably, the pitfalls of Big Data were already formulated in "As We May Think", when Bush wrote "we seem to be worse off than before – for we can enormously extend the record; yet even in its present bulk we can hardly consult it."[115] This is to be achieved by the personal and associative indexing of the memex that each individual uses to trace her path through the universal record of knowledge.

Despite envisioning the memex to take the form of a wooden desk-like contraption including levers, Bush had sketched out not just what drove the development of the personal computer in the 1970s, but also what could be called eScience.[116] Notably, this appears to call for historical research as much as for scientific inquiry, thus also foreshadowing what we have come to call Digital Humanities. Reminding us of how "As We May Think" was published just weeks before the dropping of the first nuclear bombs on Hiroshima and

---

[112] Bush, "As We May Think."

[113] Ibid.

For further context on Vannevar Bush's influence see G. Pascal Zachary, *Endless Frontier: Vannevar Bush, Engineer of the American Century* (New York: The Free Press, 1997).

[114] Hey, Tansley, and Tolle, *The Fourth Paradigm. Data-Intensive Scientific Discovery.*

[115] Bush, "As We May Think."

[116] Gramelsberger, *Computerexperimente. Wandel Der Wissenschaft Im Zeitalter Des Computers*; *From Science to Computational Studies. Studies in the History of Computing and Its Influence on Today's Sciences.*

Nagasaki, Bush closes on a cautionary note, remarking that man "may perish in conflict before he learns to wield that record for his true good."[117]

Postwar computing in the United States, however, was dominated by the military concerns of the Cold War and focused on cryptography and cybernetic control of ballistic missiles. Thus, the power of supercomputers was taken to be a measure of progress in computing, more so than database technologies.[118] Yet, the predominance of supercomputing as the main concern of national digital infrastructure projects should also be challenged. Today, Big Data is not about the amazing speed and power of supercomputing centers, but about the amount of data in distributed systems and the kinds of novel analytics employed to mine this trove of information.

In May 2003, the National Research Council's Computer Science and Telecommunications Board (CSTB) met at Stanford University to listen to a presentation by Gordon Bell and Jim Gray, both of them working at *Microsoft Research* at the time. "Gordon and I have been arguing that today's supercomputer centers will become superdata centers in the future," Jim Gray was quoted by *New York Times* technology correspondent John Markoff.[119] While United States science policy had funded immense supercomputer infrastructure programs since the 1980s, the two IT engineers were arguing that it was no longer computing capacity but data storage capacity and ease of access that was crucial to scientific computing.[120] "Central to the Bell-Gray argument is the vast amount of data now being created by a new class of

---

[117] Bush, "As We May Think."

[118] Compare Paul N. Edwards, *The Closed World. Computers and the Politics of Discourse in Cold War America* (Cambridge, MA: MIT Press, 1996). As well as Naomi Naomi Oreskes and John Krige, eds., *Science and Technology in the Global Cold War* (Cambridge, MA: The MIT Press, 2014). and George Dyson, *Turing's Cathedral. The Origins of the Digital Universe* (London: Penguin, 2012).

[119] John Markoff, "In Computing, Weighing Sheer Power against Vast Pools of Data," *The New York Times*, 2 June 2003 2003.

[120] Compare National Research Council Committee on Innovations in Computing and Communications: Lessons from History, "Funding a Revolution: Government Support for Computing Research," (Washington, D.C.: National Research Council, 1999).

scientific instruments that integrate sensors and high-speed computers," writes Markoff. Basically, Bell and Gray argued for a reorientation of strategy for US scientific infrastructure policy, turning away from the focus on powerful supercomputers able to run intricate simulations of weather or war, toward an infrastructure that forms the foundation of computers and sensor networks by providing database technologies for the entire scientific community, reminiscent of Vannevar Bush's notion of the Memex.

Jim Gray's colleague and friend Gordon Bell was another central figure in the history of database technology and had been personally involved in establishing what became the World Wide Web through his participation at the National Science Foundation's Computing and Information Sciences and Engineering Directorate and his work on the National Research and Education Network in the late 1980s.[121] Keenly aware of technology history, he introduced one of computer science's 'laws' in a 2007 paper titled "Bell's Law for the birth and death of computer classes," postulating that roughly every ten years, a new kind of computing device would come along that rendered previous systems obsolete.[122] For example, personal desktop computers have eventually come to be replaced by various mobile and connected computing devices such as tablets and smartphones.

And yet again, in 2003 Bell and Gray were announcing a new era, arguing that "data-storage technology is now significantly outpacing progress in computer processing power, [...] heralding a new era where vast pools of digital data are becoming the most crucial element in scientific research."[123] In essence, Bell and Gray were announcing nothing less than an era of Big Data in scientific infrastructures to the National Research Council in 2003, without actually mentioning "Big Data" by name.

---

[121] See Jane Jane Abbate, *Inventing the Internet* (Cambridge, MA: MIT Press, 1999).

[122] Gordon Bell, "Bell's Law for the Birth and Death of Computer Classes: A Theory of the Computer's Evolution," (San Francisco: Microsoft Research, 2007).

[123] Markoff, "In Computing, Weighing Sheer Power against Vast Pools of Data."

Jim Gray himself did not live to experience the height of the Big Data hype. He had, however, received the Turing Award, one of computer science's most prestigious awards, in 1998 for his contributions to the development of transaction processing. Transaction processing, which Grey introduced in the 1980s, has been called one of the most important algorithms of the modern world by the computer scientist and author John McCormick.[124] The following sections will trace Jim Gray's work and career as a central figure in database technology and seek to contextualize some of the developments that lead to the assumption of a Big Data era. It is especially noteworthy how Jim Gray frequently used reflection on the historical development of database technologies to contextualize his own work and thinking in various timelines of technological breakthroughs. As a keen networker, who was well connected in the Bay Area tech community, Gray deploys the narratives of an amateur historian to locate himself within technology history and harness the focus of a research community to rally around his predictions and research agendas.

We need to be aware of how *what* a database *is* has changed crucially over time. Not just the storage hardware has been transformed from punch-cards to magnetic tape, to hard-disks, and flash memory, but crucially the way databases were conceptualized and how one could query a database to get answers to specific questions was constantly evolving. Database technologies developed by computer scientists such as Jim Gray have enabled databases to be distributed and yet reliable, they are in ubiquitous use in the background of most digital applications, and yet the question of *where* a database is and *what* it consists of has become ever harder to pin down.[125]

[124] John MacCormick, *9 Algorithms That Changed the Future: The Ingenious Ideas That Drive Today's Computers* (Princeton: Princeton University Press, 2012).

[125] Paul E. Ceruzzi, *A History of Modern Computing* (Cambridge, MA: The MIT Press, 1998).

2.3 Relational Databases – Sort, Index, Organize

Born in San Francisco in 1944, Jim Gray trained as a mathematician and computer scientist, and spent practically his entire life in the San Francisco Bay Area and Silicon Valley. Following his undergraduate studies at UC Berkeley, Gray completed a PhD in computer science, allowing him to be exempt from the dreaded military draft during the ongoing Vietnam War. Following his doctorate, Michael A. Harrison, Gray's doctoral advisor at UC Berkeley encouraged him to stay in Berkeley for two more years as an IBM-affiliated post-doctoral researcher. Harrison later remarked on how spell checkers would have been a blessing for the young computer scientist, stating, "It was always surprising to me that, for someone so smart, Jim was so poor at spelling."[126]

Gray then went to work for IBM in 1971 at the *IBM Research* center in San Jose, where Edgar F. Codd had just developed the concept of relational databases.[127] "Jim Gray, who we all know, knows everybody," fellow database engineer Michael Stonebraker said of him in 2007, he "is the kind of guy that just pokes his nose into everything."[128] Although competitors while Gray was involved in developing IBM's first relational database management system, called System R, and Stonebraker was building the competing INGRES database system at UC Berkeley, Jim Gray appears to have had a talent for networking and was frequently in touch with the Berkeley competitors.

Jim Gray is widely recognized to have had a significant influence on the development of database technologies since the 1970s. Following his disappearance at sea in 2007, a colleague at *Microsoft* pointed out that „Jim was one of the fathers of the database industry as we know it today. While databases were invented, per se, in the late 60's and early 70's, those early

---

[126] Michael A. Harrison, "Jim Gray at Berkeley," *SIGMOD Record* 37, no. 2 (2008).

[127] E.F. Codd, "A Relational Model of Data for Large Shared Data Banks," *Communications of the ACM* 13, no. 6 (1970).

[128] Michael Stonebraker, interview by Burton Grad, 23 August 2007, 2007.

systems were not usable in most practical terms."[129] Michael Stonebraker points out in his textbook *Readings in Database Systems* that the most influential and enduring work on IBM's System R was Gray's contribution: "The transaction manager is probably the biggest legacy of the project, and it is clearly the work of the late Jim Gray. Much of his design endures to this day [2015] in commercial systems."[130]

Looking back at his own work at IBM in the 1970s, Gray published a technical report at *Microsoft Research* titled "Data Management: Past, Present, and Future" in 1996, in which he placed his own work in a broad historical context and traced what he believes to be six generations of data management in the history of technology:

> There have been six distinct phases in data management. Initially, data was *manually* processed. The next step used *punched-card equipment* and *electro-mechanical machines* to sort and tabulate millions of records. The third phase stored *data on magnetic tape* and used stored program computers to perform batch processing on sequential files. The fourth phase introduced the concept of a *database schema and online navigational access to the data.* The fifth step *automated access to relational databases* and added *distributed and client-server* processing. We are now in the early stages of sixth generation systems that store richer data types, notably *documents, images, voice, and video data.* These sixth generation systems are the storage engines for the emerging Internet and Intranets.[131]

By manual processing, Gray means any analogue media from Sumerian clay tablets to writing and printing on paper and in books. Whether a cultural capability such as speech and writing can be reduced to information processing in a manual way is questionable, however, for scientific and commercial purposes, writing and print were used for the same ends that are today

---

[129] David Vaskevitch, "Jim Gray: His Contribution to Industry," *SIGMOD Record* 37, no. 2 (2008).

[130] Peter Bailis, Joseph M. Hellerstein, and Michael Stonebraker, eds., *Readings in Database Systems*, Fifth Edition ed. (2015).

[131] Jim Gray, "Data Management: Past, Present, and Future," *IEEE Computer* 29, no. 10 (1996).

addressed by database technologies. Gray places the second era, the time of punch-cards, between Hollerith's use of them in the 1890 US census and roughly 1955. In 1951, the UNIVAC1 was delivered to the US Census Bureau and replaced thousands of punch-cards with it magnetic tape storage.[132] These databases, however, were file-oriented and used batch transaction processing, making the databases error-prone and slow to update. Online transaction processing overcame the limitations of this era to enable the use of direct access databases for applications such as stock-market trading or booking reservations by travel agents. The *Data Base Task Group* (DBTG) and General Electric engineer Charles Bachman developed this kind of new database, for which Bachman received the Turing Award in 1973.[133]

Throughout the 1970s, Jim Gray had worked on developing the fifth step of his genealogy of database technologies at IBM when he was involved in constructing the major relational database management system of the time, IBM's System R. To this day, basic relational databases use a programming language derived from the foundations of System R, the Structured Query Language, known as SQL. "In the context of the System R relational database project at IBM Research, Jim Gray developed and refined recovery techniques that ensure the reliability of the records and concurrency control methods to coordinate interactions among simultaneously executing programs accessing and modifying shared sets of records," Gray's former colleague Bruce Lindsay summed up his contribution.[134]

---

[132] See Jon Jon Agar, *The Government Machine. A Revolutionary History of the Computer* (Cambridge, MA: MIT Press, 2003)., and Dyson, *Turing's Cathedral. The Origins of the Digital Universe.*

[133] See https://amturing.acm.org/

[134] Bruce G. Lindsay, "Jim Gray at Ibm. The Transaction Processing Revolution," *SIGMOD Record* 37, no. 2 (2008).

However, IBM was unable to capitalize on the development of Jim and his colleagues. In fact, the company licensed the code for System R out to a company that is known today under the name *Oracle*, with its founder Larry Ellison. Gray comments on this technology transfer:

> Perhaps the most frustrating thing for me has been the technology transfer business. [...] However, our most successful transfer has been to Relational Systems, a company which sells a System R look-alike called Oracle. Oracle entered the market this year. It is nicer than System R in many ways. Why is it that IBM, to whom we gave both the code and years of consulting, is five years behind Oracle which started in 1977 with only the System R syntax and examples? To give another example, all our ideas about distributed database are being implemented by Tandem. They credit us with the design. IBM is not planning to use our ideas until the late eighties.[135]

In fact, IBM did not bring a relational database to market before 1982, naming their first commercial relational database product DB2. However, the main competitor of *Oracle's* relational database systems were not IBM's products but the group around Michael Stonebraker and Gene Wong at UC Berkeley, who developed a database system called INGRES. There had been, as was mentioned above, a spirit of collaboration between the rather academically inclined database engineers at *IBM Research* and the INGRES team, and Jim Gray frequently crossed the San Francisco Bay to meet with the INGRES developers at his Alma Mater. The competition between Michael Stonebraker's company *Relational Technology* and Larry Ellison's *Oracle*, who had licensed the technology that would become SQL from IBM, was fierce. By the early 1980s, Oracle had essentially taken over the market by aggressive marketing methods, which left Stonebraker with some resentment: "Larry Ellison had no qualms about lying to his customers", he commented in 2007.[136]

Yearning for a more dynamic and commercially oriented work environment, Jim Gray eventually quit his job at *IBM Research*: "I am resigning my position

---

[135] Jim Gray, "A Critique of Ibm's Computer Science Research," ed. IBM Research (San Jose: IBM Research, 1980).

[136] Stonebraker, "Oral History of Michael Stonebraker."

at IBM Research because it is seventy-five minutes from my home and I am a little tired of commuting," is how he started his resignation letter in 1980.[137] After several comments on commuting and IBM's apparent preference to locate their research centers far away from the urban centers that Gray seemed to prefer, he goes on to lay out his personal understanding of what it means to do research: "Perhaps I should begin with a very personal statement: I aspire to be a scholar of computer science. All fields of scholarship, from religion to medicine emphasize three aspects: meditation, teaching and service."[138]

His frustration appears to have been long in the making, since he had circulated memos in the company before, decrying the lack of computing infrastructure and commercial product orientation at IBM. "When I left UC Berkeley to join IBM, I was surprised to find that the university provided better computing services than IBM."[139] Not before he entered *Microsoft Research* in 1995 would Jim Gray be able to work full time as a scholar of computer science. Yet, for the moment, Gray moved on to one of the first Silicon Valley companies that were fostering the sort of experimental work environment that so many start-ups attempt to emulate today, *Tandem Computers* in Cupertino, California.

2.4 Transaction Processing – Setting Standards

Pat Helland, an early employee of *Tandem Computers*, said about his work on fault-tolerant database systems at *Tandem*: "We read LOTS of papers but the ones that mattered were written by this fellow named Jim Gray who worked at IBM."[140] *Tandem Computers* had been founded in 1974 and built commercial database applications that required an especially high level of reliability, such as

[137] Gray, "A Critique of Ibm's Computer Science Research."

[138] Ibid.

[139] "Mip Envy: A Programming Complex," ed. IBM Research (San Jose: IBM Research, 1980).

[140] Pat Helland, "Knowledge and Wisdom," *SIGMOD Record* 37, no. 2 (2008).

bank transactions, cash machines, stock exchanges, and airline booking centers.[141] Tandem's culture appears to have been the polar opposite of the corporate juggernaut IBM. As a young company, it was still run by its founders and had an "unusual [...] culture which has been adopted and adapted by many startup companies," stated former colleague John Nauman.[142]

The *Tandem* products were supposed to process database transactions without interruptions, and were thus called NonStop. Jim Gray arrived from work on IBM's relational databases, System R and DB2, including its query language SQL, and used his experience to combine SQL for relational databases with the fault-tolerant systems developed by *Tandem* to create NonStop SQL. This was a strategic pivot for *Tandem*, since most commercial users of databases did not use SQL-based systems for their crucial distributed systems. However, Gray was able to convince *Tandem* that an SQL-based version of their NonStop industrial product was the most cost-effective way to move forward. "NonStop SQL was developed by a relatively small team, many of whom Jim recruited from outside Tandem. He served as everything from architect to developer to cheerleader within the team while at the same time continuing to explain the benefits to *Tandem's* upper management," John Nauman elaborated.[143]

Gray is also credited with developing what is to this day known as the "ACID test" for database transactions. ACID is the acronym for atomicity, consistency, isolation, and durability. Atomicity postulates that one database transaction shall never be split or carried out only partly. One transaction has to be either carried out completely or it has to be rolled back in case of any faults. For example, a bank transfer has to comprise a change in both the origin and the destination account of the transfer, otherwise transferred money could either be

---

[141] Gaye I. Clemson, *Tandem Computers Unplugged. A People's History* (Fast Pencil Inc., 2012).

[142] John Nauman, "Jim Gray's Tandem Contributions," *SIGMOD Record* 37, no. 2 (2008).

[143] Ibid.

lost or generated out of the blue. Thus, atomicity ensures the consistency of the databases involved in the transaction, although different types of databases will require appropriate conditions of consistency. Isolation is a crucial condition when a large number of transactions are processed in parallel online or in a distributed system. To ensure the efficiency of the process, "each transaction must appear to be executed as if no other transaction is executing at the same time," even though in practice, many transactions are processed in parallel. Finally, durability means that it has to be ensured that after the completion of a transaction, changes in the database cannot somehow be corrupted, which would once again render the databases inconsistent.[144]

Furthermore, Gray was involved in introducing performance benchmarks for database transactions. Moving from software engineer into a product development role at *Tandem Computers*, he was more frequently in contact with customers. "Jim kept a suit hanging on the back of his office door. If someone needed a technical spokesperson to address a customer's concerns, Jim could transform himself from a dressed-down engineer/architect to a super-product-manager," a co-worker describes his evolving role at *Tandem*.[145] By 1985, Gray had also published his theoretical considerations of what transaction processing benchmarks could be in his papers "One Thousand Transactions per Second" and "A Measure of Transaction Processing Power".[146] The setting of standards and measures to make performance comparable seems to have appealed to Gray as a natural networker.

[144] Hector Garcia-Molina, Jeffrey D. Ullman, and Jennifer Widom, *Database Systems: The Complete Book*, New International Edition of Second Revised Edition ed. (New York: Pearson Education, 2013).

[145] Nauman, "Jim Gray's Tandem Contributions."

[146] See Jim Jim Gray et al., "One Thousand Transactions Per Second" (paper presented at the Proceedings of IEEE Compcon-85, San Francisco, 1984); Jim Gray, "A Measure of Transaction Processing Power," *Datamation* (1985).

Also in 1985, Gray started the High Performance Transaction Systems (HPTS) Workshop. The HPTS Workshop is still held every two years on the Asilomar Conference Grounds in Pacific Grove, California, and is currently being co-organized by Gray's former colleague at *Tandem*, Pat Helland. The workshops bring together computer science researchers from top universities with database engineers from the largest Silicon Valley companies, including *Amazon*, *Google*, IBM, and *Oracle*.[147]

Another yet more institutionalized forum for database hardware and software manufacturers to discuss industry standards was launched upon encouragement by Jim Gray in 1988, the Transaction Processing Performance Council (TPC).[148] All of the institutions have established themselves as joint forums for database technology researchers from academia and the private sector, enabling the practitioners to exchange their experiences and collaboratively adjust the research agenda to address issues encountered in commercial applications.

The 'linear narrative' of Big Data overlooks the importance of standards in measuring and comparing the performance of database systems. Without a common way of assessing the 'size' and velocity of a database, postulates of new achievements remain vacuous.[149] In an IBM whitepaper, five "Vs" of Big Data are described to characterize the phenomenon: volume, variety, velocity, viability, and value. Especially the volume and velocity parameters of a database

---

[147] See www.hpts.ws/index.html

[148] David J. DeWitt and Charles Levine, "Not Just Correct, but Correct and Fast. A Look at Jim Gray's Contributions to Database System Performance," *SIGMOD Record* 37, no. 2 (2008).

[149] While IBM uses these five V widely, as does Gartner, there is also a discussion about how useful these are in defining Big Data, see www.ibmbigdatahub.com/blog/why-only-one-5-vs-big-data-really-matters

cannot be measured without a form of standard to compare the performance of various database and transaction systems.[150]

In addition to networking in the commercial and academic database research community, Jim Gray also aimed to unify the field by creating common ground in the teaching of database technologies. Gray cited "meditation, teaching, and service" as his central career aims in his IBM resignation letter, however, immersed in research and involved in a commercial company such as *Tandem Computers*, Gray did not regularly teach. Yet still, as a networker and mentor, his desire to teach had not vanished. In 1987, he wrote in a letter to his wife Donna Carnes: "I bought a Mac to write the Great American Technical Novel. I was to start March 16, but now it is April 27th and I have yet to do anything on it. [...] So in June I'll take a leave of absence from Tandem and devote myself to writing."[151]

Gray had taught a one-week seminar on transaction processing in Berlin in collaboration with the German academic Andreas Reuter in early 1987, and the two decided to turn the slides of their workshop presentations into a textbook. Yet, the project stalled for several years until Gray and Reuter "decided to rent a house in a small village in Tuscany named Ripa (near Carrara) and spend February through April of 1990 there."[152] After another stint of focused writing, the textbook ended up being longer than a thousand pages and was published in 1992 under the title "Transaction Processing – Concepts and Techniques".[153] Usually, textbooks in computer science have a short half-life. Yet, the textbook

---

[150] One discussion of the 5 Vs of Big Data from IBM can be found in B. Delibašić et al., *Decision Support Systems V – Big Data Analytics for Decision Making: First International Conference, Icdsst 2015, Belgrade, Serbia, May 27-29, 2015, Proceedings* (Springer International Publishing, 2015).

[151] Donna Carnes, "Ode to a Sailor," *SIGMOD Record* 37, no. 2 (2008).

[152] Andreas Reuter, "Is There Life Outside Transactions? Writing the Transaction Processing Book," ibid.

[153] Jim Gray and Andreas Reuter, *Transaction Processing: Concepts and Techniques*, The Morgan Kaufmann Series in Data Management Systems (San Francsico, California: Morgan Kaufmann Publishers, 1992).

was well received and is still in print as one of the major texts on Transaction Processing nearly twenty-five years after its publication.

By the early 1990s, when Jim Gray left *Tandem Computers* to work for *Digital Equipment Corporation*, he had not only contributed to major developments in relational database technology and transaction processing, but had established himself as a major figure in setting standards for database performance measures as well as in teaching following generations of database engineers.

2.5 Microsoft TerraServer – a Virtual Earth

In 1995, Gordon Bell, another former employee of *Digital Equipment Corporation*, and Jim Gray were the founding directors of the *Microsoft Research* center in the Bay Area.[154] Just after Gray had arrived at *Microsoft Research*, the company envisioned to launch a project that was supposed to impressively display to their competitors that they were capable of creating the largest online database ever conceived at the time. According to his colleague Tom Barclay, Gray was initially reluctant to work on a project that was merely a scaled-up version of an old technology, questioning the research value of such an endeavor.[155] Yet, he appears to have been convinced by the challenge to construct an online database that exceeded one terabyte of data, postulating that the team should aim to "find both an interesting tera-byte *and* a cheap tera-byte."[156] Eventually, *Microsoft* chose the goal of providing images of the surface of the globe for its terabyte database ambitions and christened the project TerraServer.

---

[154] For more on Gordon Bell's work on self-archiving see Wilkinson, "Remember This? A Project to Record Everything We Do in Life."

[155] Tom Barclay, "Terraserver and the Russian Adventure," *SIGMOD Record* 37, no. 2 (2008).

[156] Ibid.

Jim Gray led the TerraServer project and was able to establish a co-operation with the United States Geological Survey (USGS) to incorporate more than 2.3 Terabytes of their grayscale images. To acquire satellite images, Gray and several colleagues went on a trip to Russia, where they were able to forge a co-operation with *Sovinformsputnik*, who provided more than 1 terabyte of recently declassified Russian military satellite images at a resolution of about two meters. The co-operation had been established via the small firm Aerial Images that was attempting to capitalize on the opening up of regulation concerning the distribution of high resolution satellite images following the collapse of the Soviet Union.[157]

However, the Russians from *Sovinformsputnik* were only willing to provide the satellite images on the condition of personally meeting with the project's directors. The Russians wanted *Microsoft* to guarantee data security as well as the promise to construct and online platform for the commercial distribution of images by their US partner *Aerial Images*. Furthermore, they wanted to publicly announce the co-operation with *Microsoft* during a press conference with the Russian Space Agency. Eventually, an agreement was reached and Jim Gray participated in a press conference in Moscow announcing the co-operation between *Microsoft* and *Sovinformsputnik*. Before the Americans returned to California, the agreement was celebrated with a "nine-course meal and [we] participated in 27 vodka toasts [...] We didn't sober up until we arrived back in the US two days later," Tom Barclay recollected.[158]

Thus, Gray and his team were able to begin constructing TerraServer in late 1996, and the online database of satellite images and aerial photographs was eventually launched on 22 June 1998. According to a *New York Times* article covering the launch of TerraServer, *Microsoft* had initially "considered creating a database for major league baseball statistics, or of every trade in the history of

[157] Tom Barclay, Jim Gray, and Don Slutz, "Microsoft Terraserver: A Spatial Data Warehouse," in *Technical Report* (Microsoft Research, 1999).

[158] Barclay, "Terraserver and the Russian Adventure."

the New York Stock Exchange, but neither project provided enough data to suit its goals."[159] While *Microsoft* was dominating the operating systems market with Windows and the consumer software market with its MS Office products, the market for commercial business databases was firmly held by the old rivals IBM and *Oracle*, and not Microsoft's SQL Server software. IBM spokespeople were quick to denounce *Microsoft's* claim to the largest existing database, stating "We've been at this for a while. It's good to see other companies learning to put large databases on the Internet."[160]

Of course, to reliably test scalability, the project would not only have to include a very large database, but would also have to attract millions of users to access the database and prove its capabilities. The TerraServer team had initially estimated a demand of about 250,000 page views per day, which was later expanded to an estimate of one million daily views. However, once TerraServer went officially online on 24 June 1998, there was a demand of more than eight million views a day, which forced the team to expand their capacity from one to ten webservers, just to be able to deliver the content at a reasonable bandwidth.[161] Eventually, TerraServer was integrated into follow-up projects such as Microsoft Virtual Earth and Bing Maps, while the TerraServer website itself is no longer available.

2.6 Setting Research Agendas

In 1998, Jim Gray received the most prestigious award of the computer science community, the Turing Award. His acceptance speech was later released as a technical report at *Microsoft Research*, titled "What Next? A Dozen Information-Technology Research Goals". In his speech, Gray speaks of cyberspace as a new frontier, a "New World": "One way to think of the

---

[159] Matt Richtel, "Huge Microsoft Photo File Is Part of a Bigger Picture," *The New York Times*, June 25, 1998 1998.

[160] Ibid.

[161] Barclay, Gray, and Slutz, "Microsoft Terraserver: A Spatial Data Warehouse."

Information Technology revolution is to think of cyberspace as a new continent – equivalent to discovery of the Americas 500 years ago."[162] Referring to his work as a member of the Presidential IT Advisory Committee, Gray called for a "Lewis and Clark style expedition into cyberspace."[163]

On the one hand, databases are supposed to create a representation of the world, which is supposed to render new insights into the physical world. On the other hand, Gray construes information technology as a new continent unto itself, which one is supposed to explore. This is a striking inversion of world and database, a construal that hints at Jim Gray's ideas of a new kind of epistemology associated with database interfaces that we will discuss further in the section on Gray's idea of the Fourth Paradigm.

Gray's talk also hits upon a central dilemma of past and current science policy, the question of whether the results of publicly funded research should be available for free to the public that has funded it in the first place. Furthermore, shouldn't the public profit from the gains made by the commercialization of products based on such publicly funded research?[164]

The unresolved problems arising from the new ubiquitous storage were the issues of privacy and intellectual property in cyberspace. "So, why isn't everything in Cyberspace? Well, the simple answer is that most information is valuable property and currently, cyberspace does not have much respect for property rights."[165] While the amount of information available online today has skyrocketed even in comparison to twenty years ago, the issues of privacy and intellectual property remain unresolved and have only become more pressing.

Gray acknowledges the issue of privacy when he posits the creation of a "Personal Memex" technology, a "box that records everything you see, hear, or

---

[162] Jim Gray, "What's Next? A Dozen Information-Technology Research Goals,"ibid.

[163] Ibid.

[164] See questions of Open Science in "Science as an Open Enterprise."

[165] "What's Next? A Dozen Information-Technology Research Goals."

read", as a research goal. Similarly, the idea of a "World Memex", Vannevar Bush's "vision of putting *all* professionally produced information into Memex" appears to Gray as a research goal within close reach since "we are getting close to the time when we can record most of what exists very inexpensively."[166] Not only is a World Memex supposed to be able to store text and other media, it is also supposed to "answer questions about the text and summarize the text as precisely and quickly as a human expert."[167] At the time of Gray's Turing Award speech, the term Big Data was not in wide use yet, in fact Gray did not mention it at all, and yet, the Personal Memex and the World Memex as Gray construes them are a Big Data vision *avant la lettre*. An obvious model of what a Personal Memex could be are current smartphones that include various sensors and enough storage to carry around media such as pictures, music, and video files. However, the launch of a smartphone such as the iPhone in 2007 required another step in storage technology: flash storage.

While the first iPod music player, introduced in 2001, still contained a small hard disk drive, flash storage is more suitable to mobile devices, since it consumes less energy, creates no noise, and cannot be disrupted by motion. Jim Gray once again anticipated the rise of flash storage technology and summed it up in a talk given in 2006, stating "Tape is Dead, Disk is Tape, Flash is Disk, RAM Locality is King."[168] Magnetic tape and its smaller offspring floppy disks had long since been out of use, while the market for hard disk drives had been growing and innovating relentlessly since the 1980s.[169] By 1995, flash storage chips with a capacity of up to 16 Megabytes were available, and the capacity had risen to 16 Gigabytes by the year 2005, which is essentially what many smartphones contain. However, there were some shortcomings of the flash

[166] Ibid.

[167] Ibid.

[168] "Tape Is Dead, Disk Is Tape, Flash Is Disk, Ram Locality Is King," (Microsoft Research, 2006).

[169] Clayton M. Christensen, *The Innovator's Dilemma. The Revolutionary Book That Will Change the Way You Do Business* (New York: Harper Business, 2000).

technology, since it was still comparatively expensive, and fairly slow at reading data.

In "Tape is Dead", Gray also proposed that one could construct an entire file system out of flash storage that would take up less energy and space and also be faster, because the separation between main memory and random access memory in computations is broken down. In 2012 SAP co-founder Hasso Plattner called Jim Gray's program for flash storage "100% true – every single word. He predicts what is happening and [it] will happen. And we just work along."[170] In conjunction with the power of multi-core central processing units (CPUs), this kind of memory technology is what actually enables real time 'Big Data' applications. Plattner remarks "we can do things now we couldn't do before," such as "instant calculation of pricing based on the current situation in the market. Wall Street does that every single second."[171]

2.7 Fourth Paradigm Science

Thus, by 2007, the components of what is data-intensive machine learning as it had been envisioned in Jim Gray's 1998 Turing Award lecture, were eventually coming together. According to *New York Times* author Thomas Friedman, 2007 was the year that the era of digitalization of the 1990s entered the next level of acceleration. "In 2007, storage capacity for computing exploded thanks to the emergence that year of a company called Hadoop, making 'big data' possible for all."[172] At the time of writing, even database giants such as IBM and *Oracle* are deploying *Hadoop* to perform analytics on unstructured data. Yet, I do not want to leap too far ahead and focus on the state of discourse around Big Data in 2007.

[170] Hasso Plattner, interview by John Hollar, 2012.

[171] Ibid.

[172] Friedman, *Thank You for Being Late. An Optimist's Guide to Thriving in the Age of Accelerations.*

One practitioner in the field of database technology in particular, and what Gabriele Gramelsberger has called "eScience"[173] generally, had taken Chris Anderson's cue even before his infamous article on the "End of Theory", Jim Gray, who made his own contribution to the business of announcing new scientific eras and 'paradigms' in his speech "eScience: A Transformed Scientific Method" on 11 January 2007 at the annual convention of the Computer Science and Telecommunications Board (CSTB) of the US National Research Council.[174]

Gray's talk on "eScience: A Transformed Scientific Method" is instructive in laying bare the rhetorical strategies deployed in order to construct a continuity between what Gray thinks of as a new way of doing data-driven science and the history of science. His talk is also the introduction to a book published by *Microsoft Research* titled "The Fourth Paradigm". Obviously, the announcement of a fourth paradigm implies the existence of three previous paradigms that are somehow being superseded by the new method of eScience. In fact, Gray mostly focuses on the locus of calculation and hypotheses testing rather than discussing characteristics of scientific paradigms in detail. He starts out speaking about scientific paradigms, presented as largely continuous rather than incommensurable and, over and over, ends up much closer to home, discussing digital scientific infrastructures.

Crucially, it is not the sheer amount of data that Gray takes to be the central aspect of any new paradigm, it is the technology deployed in knowledge creation: digital knowledge infrastructures. Most importantly, also for Gray's work as a technologist, he is concerned with the question of where data 'meets' software. The engineer of scientific infrastructures has to address the question of whether to transport the data to the calculation or carry the calculating

[173] Gramelsberger, *Computerexperimente. Wandel Der Wissenschaft Im Zeitalter Des Computers*; *From Science to Computational Studies. Studies in the History of Computing and Its Influence on Today's Sciences.*

[174] See Jim Gray 2009, "eScience: A Transformed Scientific Method" in Hey, Tansley, and Tolle, *The Fourth Paradigm. Data-Intensive Scientific Discovery.*

power to the data. Thus, the size of Big Data becomes as crucial as the speed of data transfer. In fact, the 'size' of data is completely relative to the speed at which it can be transferred

According to Jim Gray, the history of science has seen four distinct paradigms in research. Since Gray was by no means a historian of science, and probably did not aspire to be one, we should not understand his ideas as part of an historian's argument. In fact, the four paradigms may exist simultaneously or coexist as a plurality of methods within disciplines. For Gray, the first 'paradigm' is empirical science that supposedly has been practiced since the time of the ancient Greeks. This is supposed to be the kind of science that describes empirical phenomena and observations. It is unclear how much quantification and hypothesizing is supposed to be involved in this kind of science, since Gray entirely disregards both philosophical origins and non-western scientific traditions. The second 'paradigm' is the "theoretical branch" of science that employs generalizations and models in order to derive general knowledge about the world. Saying that this kind of science has been going on for the "last few hundred years", Gray may be thinking of the kind of mathematically driven inquiry in the natural sciences since the time of Newton and Leibniz.

The third 'paradigm' according to Gray is then the use of computational simulations in science during the past few decades. Under this paradigm, complex phenomena are simulated, which requires at least some digital computational capacity. This has been feasible only since the Second World War and was not deployed on a larger scale until the expansion of scientific computing in the 1960s and 1970s. But even then, computational capacity was only accessible to a selective few, since the resources of supercomputing centers were limited and exclusively available in a few developed countries. Finally, the fourth 'paradigm' according to Jim Gray is that of data exploration and eScience, which he characterizes as follows: "data captured by instruments or generated by simulator, processed by software, information/knowledge stored in computer, scientist analyzes database / files using data management and statistics."[175]

Yet, one should ask how any of these characteristics constitutes a fundamental difference from the kind of research conducted under the third paradigm. Data has been captured by instruments since the development of the experimental method in science. Also, data generated by simulators is nothing exclusively used in computational sciences at the beginning of the twenty-first century. Data being processed by software also does not seem to be anything fundamentally new, in fact, one might argue that there is no such thing as digital data that has not been processed by software. Issues with the notion of 'raw data' are insightfully discussed in Lisa Gitelman's book *Raw Data is an Oxymoron*, which argues that data cannot be conceptualized independently of its infrastructure, storage hardware and database management software.[176]

The third point in Gray's enumeration is that knowledge and information are stored in a computer. There is no definition of what knowledge and information are in this context. Information is sometimes defined as contextualized and meaningful data, while knowledge is applied and practiced information to a specific end. Thus, one might question generally whether knowledge as such and not just data and information can be stored in a computer or database at all, independently of any knowing subject. Most importantly, however, how is the storage of information on a computer anything new in comparison to the era of computational science since the Second World War, when more and more information was stored on a variety of media? Gray fails to convince here that storage alone is a sufficient and not just a necessary characteristic of Fourth Paradigm Science.

---

[175] Ibid.

[176] Gitelman, *"Raw Data" Is an Oxymoron.*

Nevertheless, Gray concluded his talk on eScience: "The techniques and technologies for such data-intensive science are so different that it is worth distinguishing data-intensive science from computational science as a new, fourth paradigm for scientific exploration."[177] Looking back at the ways in which Jim Gray has deployed internal memos, professional forums, extensive networking with colleagues, science policy advisory, and public appearances such as his Turing Award lecture, his 2007 talk on eScience also represents another instance of agenda setting by Gray. Gray's reputation as a prescient visionary of database technology development can in part be ascribed to the fact that he has been quite influential in shaping the course of database technology research throughout his career. Fourth Paradigm Science is still used as a marketing term by *Microsoft Research*, as well as Gray's colleague Gordon Bell, to promote their technological capabilities. And yet, Fourth Paradigm Science is absent from current discourse, while the technologies it connotes have been lumped in with Big Data.

2.8 "Don't replace me with a person"

This chapter has traced Jim Gray's involvement in the development of relational database systems at IBM in the 1970s as well as his work on principles and standards of Transaction Processing, laying the groundwork for the highly distributed performance of high-volume, high-velocity databases today. Platforms of discourse between academia and industry were important during the 1980s in setting industry performance standards and database research agendas. As a co-founder of the San Francisco branch of *Microsoft Research*, Gray had turned toward scientific applications of database technology in the late 1990s. His work on the *Microsoft* TerraServer was followed by further scientific collaborations constructing virtual observatories such as the Worldwide Telescope, the Sloane Digital Sky Survey, and the

[177] Hey, Tansley, and Tolle, *The Fourth Paradigm. Data-Intensive Scientific Discovery.*

Ocean Observatories Initiative. Vannevar Bush's idea of the Memex, coined in 1945, informed Gray's vision of a database technology research agenda when he laid out his vision of creating a Personal Memex as well as a World Memex in his Turing Award speech in 1998.

Although Gray did not coin the term Big Data, his work, his activities in standardization and science policy, as well as his advocacy of research agendas show him to be a major trailblazer of what we are today discussing under the heading of 'Big Data'. Just as 'the Internet' was the new promised land for entrepreneurs, which eventually failed to deliver for most but a few moguls, data itself became a new frontier for the American entrepreneurial spirit. This initiated a new space race for the data gold; the metaphor of 'data mining' should actually be taken very seriously in this case.[178] Jim Gray also construed information technology as a new continent to be explored, calling for a "Lewis and Clark style expedition" into database technology research. Yet, the metaphor of exploration has two aspects to it. On the one hand, it is the technology to be explored, while, on the other hand, developments of database technology are allegedly enabling one to explore the world in an entirely new way via database interfaces and virtual observatories, as in projects such as TerraServer.

In conclusion, we have seen how Jim Gray's talk on the Fourth Paradigm as well as his previous statements on research goals incorporate two curious developments of the recent past, one epistemological and one in public culture. In epistemology, large databases, Big Data, and Fourth Paradigm Science, promise an allegedly new scientific method that will finally lend us the tools for an immediate representation of the empirical world, the plain of truth in reality that is accessed by extensive automated measurements, thus getting rid of the subjective and soft human factors of knowledge infrastructures. This is a promise that needs to be considered with reservations. Scholars of scientific

[178] Compare e.g. Jaron Lanier, *Who Owns the Future?* (London: Allen Lane, 2013).

infrastructures such as Paul Edwards and Geoffrey Bowker remind us of the need for infrastructure inversion.[179] Only through a turning upside down of the scientific infrastructure are we going to be able to fully comprehend the knowledge derived from data-intensive science. This poses a challenge to many of the claims circulated in the Big Data discourse and stresses the importance of scholarship in communicating and contextualizing the results of any sort of alleged Fourth Paradigm Science.

In public discourse, on the other hand, Big Data is connected to a movement of American popular culture that has, more or less, succeeded in announcing the next endless frontier for exploration and expansion: data. Data is the new space, both literally and figuratively, that entrepreneurs and government agencies scramble to control, sometimes with very real aims of control and surveillance, as in the case of the National Security Agency, at other times with more hazy and commercial aims such as in the most massive advertising operations history has ever witnessed, Google and Facebook.[180] The narratives supporting this 'data frontier' discourse have their origins in discussions of research agendas and science policy reaching back to Vannevar Bush in 1945, and have been transported, among others, by well-connected prolific database engineers such as Jim Gray.

---

[179] See Bowker, *Memory Practices in the Sciences*; Bowker and Star, *Sorting Things Out. Classification and Its Consequences.*

[180] For a cultural history of these ideas see Turner, *From Counterculture to Cyberculture. Stewart Brand, the Whole Earth Network, and the Rise of Digital Utopianism.*

For a popular critique see e.g. Evgeny Morozov, *To Save Everythink, Click Here. The Folly of Technological Solutionism.*

“Don't replace me with a person, replace me with a fully configured 4341 for the exclusive use of the R* project,” is how Jim Gray concludes his resignation letter at IBM.[181] One might equally well ask what the use of humans as people remains to be when our knowledge is automated in a World Memex and the personal memories and assembled narratives that comprise any individual are rendered digitally immortal by a Personal Memex.

[181] Gray, "A Critique of Ibm's Computer Science Research."

Chapter 3: Case Study Ocean Observatories Initiative (OOI)

3.1 Introduction – Context and Methods

Chapter 1 introduced us to Jim Gray, Big Data discourse, and Al Gore's idea that creating a digital environmental network would foster a more connected, even global, environmental consciousness. Chapter 2 gave a short history of the life and work of computer scientist Jim Gray and the development of database technologies since the 1970s. This third chapter is going to consider the first of two case studies that analyze the genesis of an environmental research infrastructure in the Command and Control Anthropocene in more detail, the Ocean Observatories Initiative (OOI). The following chapter 4 is going to discuss the second case study, the National Ecological Observatory Network (NEON).

The case studies were selected to provide an exemplary view of the planning and implementation of a large environmental research infrastructure inspired by notions of Fourth Paradigm Science and the Command and Control Anthropocene. Both OOI and NEON were discussed in the Microsoft Research publication *The Fourth Paradigm* even before they were completely constructed. Thus, both case studies build on the discourse around data-intensive science, as discussed in chapter 1, and the developments in database technologies, as discussed in chapter 2. Furthermore, both research infrastructures were funded by the National Science Foundation (NSF) under their Major Research Equipment and Facilities Construction (MREFC) program.

Yet, this is where the similarities end and it makes sense to look at both programs in conjunction. OOI and NEON cover two very different environments at a different scope, oceanic and terrestrial. Whereas there were similar strategies in justifying the use the environmental research infrastructure in both cases, planning and implementation turned out to be different since the scientific communities and institutions in oceanography (for OOI) and ecology (for NEON) were entirely different.

Scientific and political institutions, individuals, and technological artifacts have all played a crucial role in the process of attempting to make a digital copy of the world, to build a Digital Earth. Climate science, ecology, and genetics have steadily grown with the expansion of technological capabilities to store data and are today at the forefront of deploying such technology.[182] In the case of climate science, this has been historicized by Paul Edwards in his book *A Vast Machine.*[183]

As Paul Edwards points out, 'the climate' is a difficult object to experience and 'know'. "No one lives in a 'global' climate,"[184] and neither does anyone live in 'the environment'. After proclamations of the 'end of nature', critiques of wilderness and constructions of naturalness as such, the environment is, likewise, hard to locate, define, or pin down. [185] The oceans, in particular, are a kind of environment that is even more remote to human contexts than e.g. forests or

---

182 See Strasser and Edwards, "Big Data Is the Answer ... but What Is the Question?."

Niki Vermeulen, "Big Biology: Supersizing Science During the Emergence of the 21st Century," *NTM* 24, no. 2 (2016).

Gabriele Gramelsberger, "Big Data-Revolution Oder Datenhybris? Überlegungen Zum Datenpositivismus Der Molekularbiologie," *NTM Zeitschrift für Geschichte der Wissenschaften, Technik und Medizin* 25, no. 4 (2017).

183 Edwards, *A Vast Machine. Computer Models, Climate Data, and the Politics of Global Warming.*

184 Ibid. p. 4

185 See e.g. Bill McKibben, *The End of Nature* (New York: Random House Trade Paperbacks, 2006); Timothy Morton, *Ecology without Nature: Rethinking Environmental Aesthetics* (Cambridge: Harvard University Press, 2009). and J. Purdy, *After Nature: A Politics for the Anthropocene* (Harvard University Press, 2015).

agricultural landscapes, which is why we are highly reliant on technology to mediate knowledge about the oceans.

Rachel Carson wrote in the introduction to the 1961 edition of her book *The Sea Around Us*: "The sea has always challenged the minds and imagination of men and even today it remains the last great frontier of Earth."[186] Indeed, one only needs to look at a world map, as they can often be found in schools, to realize that while landmasses are projected including various physical and political features, oceans are usually depicted as a simple, empty, light blue expanse.[187] This shows, that oceans and the deep sea are at the same time poorly explored and beyond the scope of our everyday interests. Thus, the oceans remain an empty 'blue spot' in our popular imaginations.

Analyzing Knowledge Infrastructures – Methodology

The analytic concepts deployed in this case study, as well as in that of the following chapter on NEON, are modelled after the approach used by Paul Edwards in his book on the history of climate sciences, *A Vast Machine*. The key concept to consider before we set out to explore the Ocean Observatories Initiative is that of a 'knowledge infrastructure'.

Based on an article by Susan Leigh and Karen Ruhleder, Paul Edwards enumerates features that characterize such a knowledge infrastructure.[188] These features can be found in all kinds of infrastructures, both public infrastructures such as a railway network or sewage system, as well as in a large scientific infrastructure such as the Ocean Observatories Initiative: "embeddedness, transparency, reach and scope, learned as part of membership and naturalized familiarity, links with conventions of a community of practice, embodiment of

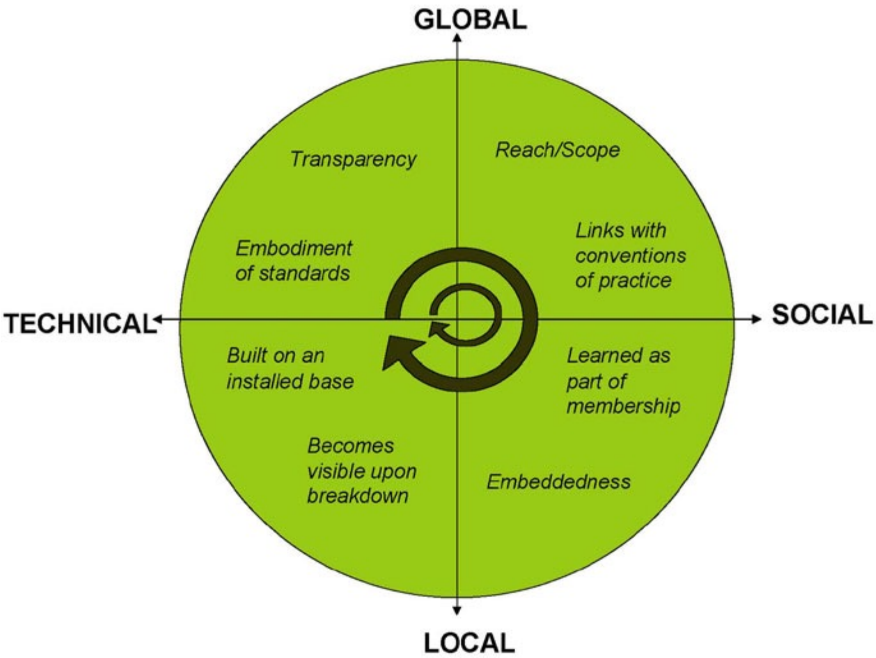
---

[186] Rachel Carson, *The Sea around Us* (Oxford University Press, 1989). p. vii

[187] Denis Cosgrove, *Apollo's Eye: A Cartography of the Earth in the Western Imagination* (Baltimore: Johns Hopkins University Press, 2001).

[188] Susan Leigh Star and Karen Ruhleder, "Steps toward an Ecology of Infrastructure: Design and Access for Large Information Spaces," *Information Systems Research* 7, no. 1 (1996).

standards, built on an installed base, becomes visible upon breakdown, and is fixed in modular increments, not all at once or globally."[189]



Cyberinfrastructure as *distributions* along technical/social & global/local axes [190]

Figure 2 Analyzing Knowledge Infrastructures

Obviously, these features of an infrastructure are more apparent when considering public infrastructures such as a domestic water supply. Yet, an environmental research infrastructure is also a kind of infrastructure, since it delivers a public service, more specifically, it generates public information and knowledge on the state of the environment.

---

[189] Edwards, *A Vast Machine. Computer Models, Climate Data, and the Politics of Global Warming*.p. 9

[190] From Geoffrey C. Bowker et al., "Toward Information Infrastructure Studies: Ways of Knowing in a Networked Environment," in *International Handbook of Internet Research*, ed. J. Hunsinger (New York: Springer). p. 101

However, we will see that in planning and implementing the Ocean Observatories Initiative, some of the characteristics of an actual infrastructure first had to be created. For example, the project was linked to the conventions of one specific research community of practice, and yet, OOI brought together various scientific communities and could not build on a coherent set of practices. Rather, the oceanographic community first had to come to terms over the kinds of standards and practices it wanted to implement. One could even say that a community of practice first had to be forged out of various groups of scientists, administrators, and computer engineers. Such an emergence of standards and practices is not unusual for an infrastructure under construction, however, a knowledge infrastructure linking together so many institutions and new technologies will face particular challenges with respect to standards and practices during rollout, maintenance, and operations. Furthermore, while OOI was built on an installed base of existing oceanographic technologies, research methods, and institutions, and is embedded in many other infrastructures, both physical and digital, the planning community had to make choices concerning on what base of predecessor projects to build. Individual technologies such as moored measurement buoys, autonomous underwater vehicles, or the collection of ocean data by ARGO floats connected to satellites were already available. The novel ambition to link all of these observation technologies to create a research infrastructure and an integrated virtual observatory, however, relied on database technologies that were developed in the late 1990s (see chapter 2), when OOI was conceived. Thus, while the oceanographic community was in the process of planning how to create an ocean observatory, digital technologies and database capabilities were rapidly expanding beyond what a marine scientist could foresee. Yet even though technology was steadily expanding, being able to connect numerous observation technologies still left oceanographers with the challenge of which existing observation systems to build on and weave into a digital ocean observatory.

In regarding the Ocean Observatories Initiative, we are going to see that the OOI was conceived as a scientific infrastructure embodying the characteristics listed above. We will not have the space to elaborate on all the infrastructure characteristics laid out by Edwards to prove that OOI is indeed a kind of infrastructure. The focus is going to be on individual, institutional, and technological actors. Communities of practice are particularly relevant since the OOI project has at times appeared to fall into fractions of individual experimental setups due to the entrenched practices within the scientific community of oceanography and geosciences. In fact, we are going to see that a scientific infrastructure such as the Ocean Observatories Initiative is a perpetual work-in-progress that is both driven and threatened by the dynamics within the scientific community, the public, and political institutions.

Paul Edwards describes this work-in-progress character as a "perpetual oscillation between the desire for smooth, system-like behavior and the need to combine capabilities no single system can provide," and then concludes that "*infrastructures are not systems* but networks or webs."[191] Thus, I am going to use Edwards definitions of a knowledge infrastructures as "robust networks of people, artifacts, and institutions that generate, share, and maintain specific knowledge about the human and natural worlds."[192]

It is important to note, however, that the systems character and the network character of the knowledge infrastructure form a continuum, with centralized top-down systematic forces on the one hand, and decentralized bottom-up network forces on the other. The distinction between a system and a network infrastructure is central to the case study chapters. A system is a kind of infrastructure technology that is conceived and built as a complete and closed-up structure. While the system can be expanded, it can only be expanded as such, and any incorporation of essentially different kinds of technologies is

---

[191] Edwards, *A Vast Machine. Computer Models, Climate Data, and the Politics of Global Warming*.p. 12

[192] Ibid.p. 17

difficult. In contrast, a network is a linkage of technological infrastructure components that can be much more heterogeneous, built on top of already existing networks, and is open to expansion at most nodes of the network. The oscillation between the network character and the systems aspects of the Ocean Observatories Initiative as a scientific infrastructure will become apparent, as its dual nature entails both virtues and challenges for such an environmental research infrastructure.

Following Paul Edwards, the case study is going to consider three kinds of actors: people, i.e. the scientists, engineers, and administrators involved; artifacts, the physical and digital technology deployed to build the network of ocean observatories; and institutions, the scientific institutions central to the effort as well as the funding bodies and scientific expert committees involved.

The three main activities that these actors are occupied with are knowledge generation, knowledge sharing, and infrastructure maintenance. All of these three aspects have to be equally well supported in a balanced way in order for an infrastructure project such as the Ocean Observatories Initiative to deliver long-term success. We will see that all three activities, knowledge generation, sharing, and maintenance entail very specific challenges.

Expanding Paul Edward's concepts, we can arrange the main actors and activities involved in a scientific infrastructure in a three by three matrix. The intersections of actors and activities gives us an overview of the complexity of the multiple roles to be considered in assessing an infrastructure such as the Ocean Observatory Initiative. We will not be able to elaborate in detail on all of the actors and activities, yet, this 'infrastructure matrix' has informed the methodological review of the available sources and is going to be used in the following case study chapter as well.

Activities and Actors	Knowledge Generation	Knowledge Sharing	Infrastructure Maintenance
People	Scientists, Citizen Scientists	System Administrators, Educators, Lay Public	Engineers
Artifacts	Sensors, Observatories, Digital Network	Cyberinfrastructure, Databases, Public Data Portal	Observatory Maintenance
Institutions	Funding Bodies, Research Institutes, Universities	Cyberinfrastructure Providers, Data Standards	Long-Term Funders, Political Actors

Table 1: "The Infrastructure Matrix"

Knowledge sharing and public access were a central aspect in planning OOI and legitimizing its utility to the public beyond the scientific community of oceanographers. However, I will argue that the ideas of 'data products' and public access for educators and citizen scientists has not turned out in practice as had been envisioned conceptually. In the case of OOI, there were apparent problems with knowledge generation mainly due to issues in on-time delivery of the software backbone for data-processing, which led to an overhaul of the digital infrastructure provider close to the envisioned completion date of the infrastructure.

Furthermore, we will see how the dependence on political developments with respect to long-term funding of the OOI infrastructure, especially concerning the maintenance of the network, have rendered the infrastructure extremely vulnerable to abandonment, since infrastructural funding by the National Science Foundation was cut dramatically almost immediately upon completion of the OOI network. Maintaining a long-term knowledge infrastructure poses a specific challenge for a funding system that operates within timeframes of 5 to

10 years. This also distinguishes a knowledge infrastructure from infrastructures such as roads and bridges, whose dereliction is more obvious and can pose a threat to the users of the transport infrastructure. A derelict knowledge infrastructure, on the other hand, does not put its potential users in danger and the public is less likely to be outraged by its abandonment, especially if demand for the knowledge it generates is declining.

In the conclusion of this chapter, I am also going to reflect on whether or not the case of the Ocean Observatories Initiative exhibits the characteristics of Fourth Paradigm Science as postulated by Jim Gray (see chapter 1 and chapter 2). This will lead us to both criticize and expand Gray's ideas of Fourth Paradigm Science by confronting it with the tangible example of the OOI knowledge infrastructure. Since an infrastructure that figures as a network is, in essence, built upon components that have a history of previous use, I will now take a step back and consider some prerequisite projects in oceanography before taking a closer look at the planning and implementation of the Ocean Observatories Initiative.

3.2 Background - Predecessors of Ocean Observation Technologies

Commanding the Deep and Controlling the Seas

Oceanography's access to the open ocean and the deep sea as environments has always been closely intertwined with technological developments that enabled diving and remote transfer of information. Early oceanographic expeditions, such as the *Challenger* in 1872 to 1876 could merely explored the ocean's surface layers along the path of its course.[193] Only with the invention of mediating technology such as sonar, hydrophones, buoys, and satellite-based remote sensing was oceanography eventually able to assemble a 'deeper' image of the oceans. In comparison to terrestrial exploration and knowledge infrastructures, ocean environments entered the scope of science only well into the 20th century, in conjunction with satellite technology and space exploration.

Many technologies and infrastructures in science have some past connection to government infrastructures or military research and applications. When we worry about 'dual use' dilemmas such as in nuclear technology, space and satellite technology, and many applications of information and communication technologies, military and civil applications are frequently two sides of the same coin.[194] Underwater surveillance technologies, however, have a clear military origin. The exploration of the ocean, and especially the submarine realm has always been linked to commercial and military exploits, for example, when early oceanography served the needs of the navigators of the British Royal Navy.[195]

Likewise, the ability to construct vessels that enabled humans to be submerged underwater always had a military aspect. One only has to think of Jules Verne's

---

[193] R. von Willemoes-Suhm, *Die Challenger-Expedition: Zum Tiefsten Punkt Der Weltmeere : 1872-1876* (Edition Erdmann, in der Verlagshaus Römerweg GmbH, 2015).

[194] For extensive discussions of the dual use dilemma see e.g. S. Miller, *Dual Use Science and Technology, Ethics and Weapons of Mass Destruction* (Springer International Publishing, 2018). and J.B. Tucker et al., *Innovation, Dual Use, and Security: Managing the Risks of Emerging Biological and Chemical Technologies* (MIT Press, 2012).

[195] Helen M. Rozwadowski, *Fathoming the Ocean: The Discovery and Exploration of the Deep Sea* (Cambridge: Harvard University Press, 2005).

fantasy of the submarine *Nautilus*, which is used by the apparently pacifist Captain Nemo to sink the warships of imperial powers.[196] While submarines were first devastatingly deployed during the First World War, the technologies to dive beyond the range of several hundred meters took a lot longer to develop and were only worked on in a more scientific context after the Second World War.[197]

The surveillance of the submerged world had clearly become a question of survival during the all-out U-Boat war in the Atlantic during the Second World War. Attempting to protect convoys from submarine attacks, the Allies developed means such as depth charges and surveillance by destroyers and aircraft. Sonar was the major advance that gave surface-based submarine hunters an advantage in targeting submerged threats. Also, sonar enabled an entirely new look at the submerged world, since it delivers depth-profiles of the seafloor more accurately than manual sounding ever could.[198]

After the Second World War, even the atomic bomb was regarded by some as a "wonderful oceanographic tool".[199] Throughout the 1950s, oceanographers from the Scripps Institute in San Diego were involved in assessing the impacts of the fallout from nuclear weapons tests in the Pacific Ocean such as on the Bikini Atoll. "Relying on the studies of tracers, weapons tests and nuclear waste, oceanography readily took advantage of their opportunity to expand their claims of scientific expertise and cultural authority."[200] While physicists

---

[196] Jules Verne, *20,000 Leagues under the Seas*, trans. William Butcher (Oxford: Oxford University Press, 1998).

[197] See J.P. Delgado and C. Cussler, *Silent Killers: Submarines and Underwater Warfare* (Bloomsbury Publishing, 2011).

[198] Lino Camprubí, "The Sonic Construction of the Ocean as the Navy's Operating Environment," in *Navigating Noise*, ed. N. v. Dijk, et al. (Köln: Walther König, 2017).

[199] Ronald Rainger, "A Wonderful Oceanographic Tool'. The Atomic Bomb, Radioactivity and the Development of American Oceanography," in *The Machine in Neptune's Garden: Historical Perspectives on Technology and the Marine Environment*, ed. Helen M. Rozwadowski and David K. van Keuren (Sagamore Beach, MA: Science History Publications/USA, 2004).page 93 ff.

[200] Ibid.p. 94

and engineers building nuclear weaponry and rockets commanded a high status since they were working to fight of the ideological enemy, oceanography inherently carried no such national importance.[201]

Oceanographers would not only survey the geology of the nuclear test site before the detonation but also ascertain the geochemical impacts of the radiation on the oceanic environment. Later on, their work also included what was called "chemical ecology", measuring the impact and spread of radioactivity in marine organisms. It soon became clear that the oceanographers' knowledge about the impact of nuclear detonations on the ocean environment could also be useful in assessing means of nuclear waste disposal. The Atomic Energy Act was passed in 1954, and in 1955, Scripps' proposal to the AEC (Atomic Energy Commission) to study nuclear waste disposal in California's coastal waters.[202] The scholar Ronald Rainger has even suggested that "oceanographers were using nuclear waste disposal as a means to gain power, to become politically influential."[203]

With the development of nuclear-powered submarines by the United States (the *USS Nautilus* was launched in 1954), fitting nuclear cruise missiles to submarines in 1957, and of nuclear missiles on the nuclear-powered George Washington Class submarines, the situation changed again.[204] These submarines rarely needed to return to harbor to fuel up and did not just threaten other ships but were part of the nuclear deterrence. In fact, the International Geophysical Year from July 1957 to December 1958, a major

---

[201] Although the work of Rachel Carson, after all a trained marine biologist, should eventually carry a lot of cultural authority, and yet she may have been seen as a tidepool-wandering Cassandra, not as an asset to the national defense effort during the Cold War.

[202] Rainger, "A Wonderful Oceanographic Tool'. The Atomic Bomb, Radioactivity and the Development of American Oceanography." p. 110.

[203] Ibid.p. 111, In fact, what fallout monitoring was for oceanographers during the Cold War, environmental research and ocean observation may be for oceanography in the Command and Control Anthropocene, a legitimizing effort by a scientific discipline to make itself relevant in the fight with largest adversary of the era.

[204] C. Oldham, *Underway on Nuclear Power: 50th Anniversary of Uss Nautilus* (Faircount LLC, 2004).

scientific co-operation between the great powers of the Cold War, coincided with the first nuclear submarines carrying intercontinental ballistic missiles across the world's oceans.[205]

That same year, Sputnik entered the Earth's orbit on October 4th, 1957, an event dubbed the 'Sputnik Crisis' that led President Eisenhower to create the National Aeronautics and Space Administration (NASA) in 1958.[206] The following year, in 1958, the United States Navy purchased a submersible deep-diving craft called *Trieste* from the Swiss explorer Auguste Piccard. This bathyscaphe would be used by Auguste's son Jacques Piccard and the American Don Walsh to dive to the bottom of the Mariana Trench, the Challenger Deep first recorded by the *HMS Challenger* expedition of 1872 to 1876.[207] The bathyscaphe reached the bottom of Challenger Deep on January 23rd, 1960.

To counter the Soviet threat, the US Navy eventually launched the construction of underwater surveillance networks, based on moored sonar buoys to track the movements of Soviet nuclear submarines and ships in the Atlantic and Pacific Oceans. This Sound Surveillance system (SOSUS) has been described by Naomi Oreskes as a model instance of 'dual use' technology in her article "Changing the Mission: From Cold War to Climate Change".[208] What had once been used to track the Cold War adversary could also be used for civic and scientific purposes.

---

[205] M. Nicolet, *The International Geophysical Year Meetings: Annals of the International Geophysical Year* (Elsevier Science, 2013).

[206] P. Dickson, *Sputnik: The Shock of the Century* (Walker, 2001).

[207] See David M. Lawrence, *Upheaval from the Abyss: Ocean Floor Mapping and the Earth Science Revolution* (New Brunswick: Rutgers University Press, 2002).

[208] Oreskes and Krige, *Science and Technology in the Global Cold War.* p. 145

The two main centers of oceanographic research in the United States, the Scripps Institution of Underwater Listening and Location in San Diego, and the Woods Hole Oceanographic Institution were regularly involved in military projects concerning underwater sound detection supported by the Office of Naval Research.[209] Yet, once the Cold War fizzled out after the collapse of the Soviet Union, the oceanographic infrastructures were still in existence and offered opportunities for dual use by scientists. In fact, we are going to see that Scripps on the West Coast and Woods Hole on the East Coast were also the two major scientific institutions driving the development of the Ocean Observatories Initiative.[210]

From Cold War to Hot Climate

Another area of oceanographic research where precursors of the Ocean Observatories Initiative can be found is the study of the El Niño-Southern Oscillation.[211] "Beginning in 1984, a team based at the United States' National Oceanic and Atmospheric Administration (NOAA) Pacific Marine Environmental Laboratory in Seattle planned and implemented the Tropical Ocean Global Atmosphere/Tropical Atmosphere Ocean (TOGA/TAO) project."[212] Earlier, in 1970, "the United Nations inaugurated the International Decade of Ocean Exploration (IDOE). Scientific internationalists behind this multilateral program hoped a sustained, cooperative effort in the spirit of the International Geophysical Year and the on-going Global Atmospheric

---

[209] Ibid.

[210] For the history of Scripps and Woods Hole see Scripps Institution of Oceanography, *Celebrating 100 Years of Oceanography at Scripps Institution of Oceanography* (Oceanography Society, 2003). and V. Cullen, *Down to the Sea for Science: 75 Years of Ocean Research, Education, and Exploration at the Woods Hole Oceanographic Institution* (Woods Hole Oceanographic Institution, 2005).

[211] Gregory T. Cushman, "Choosing between Centers of Action," in *The Machine in Neptune's Garden: Historical Perspectives on Technology and the Marine Environment*, ed. Helen M. Rozwadowski and David K. van Keuren (Sagamor Beach: Science History Publications/USA, 2004).p. 133 ff.

[212] Ibid.p. 136

Research Program (GARP) would lead to the rapid 'conquest' of the oceans and atmosphere by science."[213]

In 1990, the US Congress established the Strategic Environmental Research and Development Program (SERDP) as part of the Department of Defense's DARPA. The formally military underwater infrastructure was now being re-appropriated for research applications, for example in the acoustic underwater mapping project called ATOC, standing first for Acoustic Tomography of Ocean Climate, which was later changed to Acoustic Thermometry of Ocean Climate. After plans to use SOSUS to measure ocean tomography, the project was reoriented toward climate research, since, at the time, climate research required more data on the temperature of the oceans.

Thus, oceanographers, in particular, were one of the groups of scientists who needed to reorient their work and research toward civil agendas after the end of the Cold War. However, as Naomi Oreskes claims, they were "naïve about the social, political, and cultural operating condition of American life at the end of the Cold War."[214]

In the context of a civil reorientation of United States oceanography, ideas for a digital observatory of the ocean were discussed as early as 1988.[215] Efforts to find a new role for oceanographic research agendas and existing infrastructures gradually began to link up with the emerging scientific discourse on climate change and related efforts to complete a global climate observation system. In 1993, the International Ocean Network (ION) was formed. That same year, the Intergovernmental Oceanographic Commission of the UNESCO submitted the "Report of the IOC Blue Ribbon Panel for a Global Ocean Observing System (GOOS)", in which experts argued for the installment of a worldwide ocean monitoring system that would complement the work of the

---

[213] Ibid.p. 153

[214] Oreskes and Krige, *Science and Technology in the Global Cold War*.p. 171

[215] See http://oceanobservatories.org/planning-history/

World Meteorological Organization and the United Nations Framework Convention on Climate Change.[216]

GOOS had been endorsed in the Agenda 21 of the Rio Earth Summit in 1992, when the 1st US workshop was held and the report "First Steps to a U.S. GOOS" was published.[217] The National Oceanic and Atmospheric Administration (NOAA) then began making more detailed plans for an American component of a Global Ocean Observing System in 1995 and 1996, calling the project Integrated Ocean Observing System (IOOS).

Thus, one might ask, if there already was an ongoing effort initiated in the 1990s, and well connected with the global climate monitoring efforts of the UN agencies, how was a national project such as the Ocean Observatories Initiative justified? IOOS was a NOAA-focused project within a broad international framework. The Ocean Observatories Initiative, however, was mostly planned and funded by the National Science Foundation (NSF) and the National Research Council (NRC). The fact that a networked scientific infrastructure was linked to other infrastructure programs on various levels was an expression of the existing institutional structures in science policy and funding. It was the perpetuated and yet contingent structure of funding institutions resulting in multiple and parallel efforts to build and maintain infrastructures, one by NOAA and the vastly more ambitious effort at OOI by the National Science Foundation. We will eventually see how, especially with respect to the maintenance of a long-term infrastructure, this dependence on specific institutional structures can put the very aims of the environmental research infrastructure at risk.

---

[216] "Report of the Ioc Blue Ribbon Panel for a Global Ocean Observing System (Goos). The Case for Goos," (Intergovernmental Oceanographic Commission (of UNESCO), 1993).

[217] See https://ioos.noaa.gov/about/ioos-history/

There have been other national efforts to construct ocean observation networks similar to the Ocean Observatories Initiative. In Canada, NEPTUNE (North-East Pacific Time series Undersea Networked Experiments) was implemented in 2008.[218] Japan constructed a submarine observatory network titled with the acronym ARENA (Advanced Real-time Earth Monitoring Network in the Area). The infrastructure most closely resembling the Ocean Observatories Initiative network, however, is the European ESONET (European Seafloor Observatory Network), which was funded by the European Union under the 6th European research framework from 2002 to 2005. ESONET's successor project, funded under the 7th European framework program, is called EMSO (European Multidisciplinary Seafloor & Water Column Observatory). Since 29th September 2016, EMSO is supported as a European Research Infrastructure Consortium (ERIC).[219] Similar to OOI, EMSO comprises both cabled observatories as well as moored buoys offshore. Its observatory sites span from the Arctic to the Black Sea, from the Azores in the Atlantic to the Mediterranean. Since the aim and the setup of the EMSO project are so similar to the Ocean Observatories Initiative in the United States a comparison and tracing of their interconnections would warrant further study, which is unfortunately beyond the scope of this inquiry.[220]

---

[218] The Economist, "Visiting Neptune's Kingdom. Outer Space Has Hogged the Limelight for Too Long. Welcome to Inner Space.," *The Economist*, 15 November 2007 2007.

[219] http://www.emso-eu.org/site/news-events/history.html

[220] For further concise information on EMSO see M. Best et al., "Emso: A Distributed Infrastructure for Addressing Geohazards and Global Ocean Change," *Oceanography* 27, no. 2 (2014).

Figure 3 ESONET/EMSO

After this brief historical excursion, I will now move on to an outline of the conception of the Ocean Observatories Initiative, its novel networks of sensors and virtual observatories that were deemed to change the way we perceive ocean space as an environment. Notions of the ocean as well as the underwater world of the deep sea are notoriously hard to 'fathom'. The case study also aims to more broadly problematize this kind of 'access' to a wild space, while at the same time pointing out some of the opportunities that such enhanced observations create for expanding the 'common consciousness' of oceans as an endangered habitat, as well as an essential resource in our global climate system.

3.3. OOI Knowledge Generation: People, Institutions, and Artifacts

The Hidden Planet Report: Envisioning the Command and Control Anthropocene

In the year 2000, the Committee on Seafloor Observatories of the Ocean Studies Board (OSB) at the National Research Council published a report titled „Illuminating the Hidden Planet: The Future of Seafloor Observatory Science."[221] The National Research Council is a private non-profit organization as part of the national academies aiming to promote science and evidence based policy-making. The committee was chaired by William Ryan, of the Lamont-Doherty Earth Observatory, and Robert Detrick from the Woods Hole Oceanographic Institution. James Bellingham from the Monterey Bay Aquarium Research Institute (MBARI) and John Lupton from the NOAA-Pacific Marine Environmental Laboratory in Newport, Oregon, also represented key oceanographic institutions.

The National Research Council's Ocean Studies Board had promoted interdisciplinary research in several reports during the late 1990s. The main argument of the "Illuminating the Hidden Planet" report was that existing seafloor observatories should be combined into a networked virtual observatory to create a larger system that could yield novel insights into ocean systems. The OSB defined ocean observatories as follows:

> For the purpose of this report, seafloor observatories are defined as unmanned, fixed systems of instruments, sensors, and command modules connected either acoustically or via a seafloor junction box to a surface buoy or a fiber optic cable to land. These observatories will have power and communication capabilities and will provide support for spatially distributed sensing systems and mobile platforms.[222]

[221] Ocean Studies Board National Research Council, "Illuminating the Hidden Planet: The Future of Seafloor Observatory Science," ed. National Academy Press (Washington, DC 2000).

[222] Ibid. p. xi

Seafloor observatories were to explicitly function without the need for human operators or pilots. Thus, a manned research submarine or even a remotely operated vehicle (ROV) is not a seafloor observatory. The seafloor observatories were then connected to "spatially distributed sensing systems" comprising in what I am going to a call a *virtual observatory.*

From January 10 to 12, 2000, the Symposium on Seafloor Observatories was held in Islamorada, Florida. The interdisciplinary discussions were supposed to assess the feasibility and scope of a network of ocean observatories and communicate its conclusions to the National Science Foundation (NSF), which was supposed to fund any such efforts. The task set by the NSF for the symposium was "to (1) assess the extent to which seafloor observatories will address future requirements for conducting multidisciplinary research in the oceans and (2) gauge the level of support for observatory science within the ocean sciences and the broader scientific community."[223]

Thus, the report was to both evaluate the level of support for such a seafloor observatories project within the scientific community as well as project future requirements of ocean related research agendas. Since a promising large investment into an infrastructure such as an ocean observatory network was bound to shape future research agendas, the report's assessment appears to have been faced with a 'hen-and-egg' problem. How was the community supposed to know what future oceanographic research needed to look like if it considered the needs of an already existing research environment with its entrenched institutions and individual research agendas? On the other hand, how would a large infrastructure project such as the Ocean Observatories Initiative win the support of the scientific community for a bold new vision of the future of oceanographic research? We will see later on how in the implementation of OOI, the discussion and valuation of OOI in the community's discourse appeared to oscillate between a rhetoric of new

[223] Ibid. p. 1f.

futuristic scientific methods and the interests of existing personal and institutional networks.

The keynote speaker on the first evening of the Symposium on Seafloor Observatories was John Delaney, from the University of Washington, who lectured on "Neptune: Oceanography at the scale of a tectonic plate". Delaney also contributed to the Microsoft Research Volume *The Fourth Paradigm* with an essay titled "A 2020 Vision for Ocean Science".[224] Delaney wrote that the "ocean has been chronically under-sampled for as long as humans have been trying to characterize its innate complexity."[225] Calling the ocean "under-sampled" already revealed the computer scientist's perspective on an environment simply as a source of information to be recorded and ordered in a database. Characterizing the ocean as innately complex, Delaney evoked the notion of the ocean as vast and unfathomable, especially since he did not say why the "innate complexity" of the ocean would be any more complex than that of other environments. However, Delaney elaborated, ocean observatories would transform the under-sampled ocean into a "data-intensive environment" entailing its own challenges: "For scientists operating in this data-intensive environment, there will be a need for development of a new suite of scientific workflow products that can facilitate archiving, assimilation, visualization, modeling, and interpretation of the information."[226]

The general conclusion of the "Hidden Planet" report eventually was that ocean observatories were especially promising in collecting long-time datasets of observations, which could be hard to assemble from ship-based measurements. Also, the committee concluded that the support for a network of ocean observatories within the community was "enthusiastic and supportive," hardly a

---

[224] Note the pun, 2020 is both the year as well as the term for perfect eyesight, often called "20-20 vision" in a US-American context.

[225] Hey, Tansley, and Tolle, *The Fourth Paradigm. Data-Intensive Scientific Discovery*.p. 30

[226] Ibid.p. 31

surprising conclusion, given that the report recommended extensive funding of oceanographic research by the NSF.[227]

Furthermore, a possible international integration of the US-based network into a larger international Global Ocean Observing System (GOOS), already mentioned above, was pointed out in favor of the project. A deeper understanding of the oceans' role in a changing global climate was cited as the main scientific merit of the establishment of ocean observatory networks. Especially the influx of anthropogenic pollutants and nutrients in coastal waters, which in turn influences processes in coastal ecosystem dynamics and biodiversity, was to be monitored. The main benefits cited by the report were "advances in societally relevant areas of oceanographic research, such as marine biotechnology, the ocean's role in climate change, the evaluation of mineral and fishery resources, and the assessment and mitigation of natural hazards, such as earthquakes, tsunamis, and harmful algal blooms."[228]

It is remarkable that the report explicitly talked about "assessment and mitigation" of the effects of climate change rather than exploration and prevention of climate change. The implication of this is that the authors of the Hidden Planet report had abandoned the goal of preventing harmful climate change effects and were proposing new kinds of knowledge infrastructures as a management tool for the United States to address the inevitable effects of global warming.

Around the same time, the prospect of marine biotechnology and its promises of new pharmaceutically valuable substances was a popular selling-point for oceanographic research, although the benefits eventually failed to materialize.[229] Thus, the major goal, was to be able to monitor changing ocean environments

---

[227] National Research Council, "Illuminating the Hidden Planet: The Future of Seafloor Observatory Science." p. 2

[228] Ibid. p. 7f.

[229] See Stefan Helmreich, *Alien Ocean: Anthropological Voyages in a Microbial Sea* (Berkeley: University of California Press, 2009).

and their anthropogenic causes. The idea of an ocean observatory network was inherently an idea of the Command and Control Anthropocene, when humans worry about the adverse impacts of an environment that is itself the result of global human activity and strive to manage them to their own benefit.[230]

Indeed, the Anthropocene confronts us with a novel kind of reflexivity that humans have become aware of especially by the use of large-scale knowledge infrastructures such as OOI. An environmental research infrastructure does not only assess the state of an environment, it also, in turn, assesses the human impact on that environment, and the impact of that Anthropocene environment on human life and habitats. The scale of this human impact on humans living in a human-made environment is normally only graspable in particular instances or locales. Only large-scale knowledge infrastructures and digital observatories can actually assemble a larger picture of the Anthropocene environment on a larger, albeit not entirely global, scale.

Questions of cyberinfrastructure were given particular attention early on, since the highly distributed network was supposed to result in *one* observatory infrastructure accessible via *one* online digital interface. The report's executive summary stated that "a challenge to any observatory data management structure will be the processing, distributing, and archiving of the very large datasets produced. A fully integrated plan for data handling should be developed early in the planning stages for any seafloor observatory program."[231]

Shortly before the Hidden Planet report was officially released in December 2000, the National Science Board of the National Science Foundation eventually approved funding of the Ocean Observatories Initiative as a Major Research Equipment and Facilities Construction (MREFC) project. In spring of 2001, the NSF Ocean Sciences division released a report titled "Ocean

[230] Nixon, "The Anthropocene. The Promise and Pitfalls of an Epochal Idea." p.12

[231] National Research Council, "Illuminating the Hidden Planet: The Future of Seafloor Observatory Science." p. 7

Sciences at the New Millennium".[232] This report assessed two pilot projects of the Ocean Observatories Initiative that were early implementations of parts of the observation network. First, in July 2001, the W.M. Kendrick Foundation had granted an award for the proto-NEPTUNE experiment, a seafloor observation infrastructure in co-operation with Canada. In March 2002, the National Oceanographic Partnership Program (NOPP) funded the NEPTUNE system engineering program office and the implementation proceeded. Later that year, in September 2002, the National Science Foundation funded the Monterey Accelerated Research Systems (MARS) cabled observatory test bed in Monterey Bay. This test bed was the first major component of the Ocean Observatories Initiative that included buoys as well as autonomous underwater vehicles (AUVs).

In the summer of 2003, the National Research Council's Committee on the Implementation of a Seafloor Observatory Network for Oceanographic Research released another major report titled "Enabling Ocean Research in the 21st Century". The committee had been tasked specifically with addressing issues on how to implement the conceived Ocean Observatories Initiative (OOI) and built on the earlier 'Hidden Planet' report. The chair, Robert Detrick, from Woods Hole, placed the OOI explicitly within a larger historical context of oceanographic science:

> In the ocean sciences, new technology inevitably leads to new discoveries and to fundamental advances in basic knowledge. In the years following World War II, for example, the first global-scale mapping and sampling of the seafloor by oceanographic research vessels led directly to the discovery of seafloor spreading and the development of the theory of plate tectonics which has since revolutionized ideas of earth structure and evolution.[233]

[232] Decadal Commission National Science Foundation, "Ocean Sciences at the New Millennium," ed. Inc. Geosciences Professional Services (2001).

[233] Ocean Studies Board National Research Council, "Enabling Ocean Research in the 21st Century: Implementation of a Network of Ocean Observatories," ed. National Academy Press (Washington, DC2003). p. vii

The idea that new research technologies inevitably lead to new fundamentally new discoveries was a major narrative in the rhetoric of promoting and justifying an expensive infrastructure project such as the Ocean Observatories Initiative. Detrick went on to write:

> The ocean sciences are now on the threshold of another major technological advance as the scientific community begins to establish a global, long-term presence in the oceans in order to understand the temporal variability of ocean systems on time scales ranging from seconds to decades or longer. This opportunity arises from the confluence of a number of emerging new technological capabilities.[234]

The new technologies listed were, among other communication technologies, new sensors, computational and modelling capabilities, "data archival systems that can store, manipulate, and retrieve huge volumes of data from arrays and sensors" as well as "computer networks that can bring real-time data to the desktop, which could potentially vastly increase participation of researchers, students, educators and the general public in ocean research and discovery."[235]

Further science policy reports accompanied the work on conceiving a national ocean observation network, e.g. in June 2003, the Pew Oceans Commission published a report called "America's Living Oceans". Almost a year later, in September 2004, the US Commission on Ocean Policy released a major assessment of American oceanographic research in the report "An Ocean Blueprint for the 21st Century".[236]

---

[234] Ibid. p. vii

[235] Ibid. p. viii

[236] United States. Commission on Ocean Policy, *An Ocean Blueprint for the 21st Century: Final Report* (U.S. Commission on Ocean Policy, 2004).

Implementing the Ocean Observatories Initiative

Eventually, the Ocean Observatories Initiative Project Office was established in March 2004 at the Joint Oceanographic Institutions, a consortium of academic institutions in oceanography based in Washington D.C..[237] With planning underway, the Ocean Observatories Initiative released a Request for Assistance for Conceptual Science Experiment, i.e. the tender for the major components of OOI was made public.

Shortly after, in May 2005, the Ocean Observatories Initiative published its "Science Plan". Building on the 'Science Plan', the Project Office proceeded to work out the details of the Ocean Observatories Initiative's conceptual design during the year 2006.[238] The OOI was listed in the President's fiscal year plan 2007 as a new Major Research Equipment and Facilities Construction (MREFC), funded by the National Science Foundation.

In March 2006, the oceanographic community congregated in Salt Lake City, Utah, for a Design and Implementation Workshop. The results were subsequently presented for peer review as the OOI Conceptual Network Design in June 2006 and published as the Revised Conceptual Network Design. A Conceptual Design Review of the infrastructure plans for the Ocean Observatories Initiative then took place in August 2006 in Moss Landing, California.[239]

After years of conceptual planning, the grants for actual construction of infrastructure components of the Ocean Observatories Initiative were finally awarded over the course of the year 2007 under the MREFC framework. In March 2007, the Conceptual Network Design Revised Infrastructure Plan

---

[237] As of today, the Joint Oceanographic Institutions have united with several other ocean-related interest groups to form the Consortium for Ocean Leadership, See https://oceanleadership.org/about-us/

[238] ORION Executive Steering Committee, "Ocean Observatories Initiative Science Plan," (Washington, D.C.2005).

[239] Kendra Daly, "The Ocean Observatories Initiative and Orion Organization. Ooi Conceptual Design Review," (Moss Landing, CA2006).

was released. At the same time, the grant for the construction of the Ocean Observatories Initiative Regional Cabled Nodes Design Support Services was awarded to the University of Washington in Seattle.

Shortly after, in May 2007, the Ocean Observatories Initiative Cyberinfrastructure grant was awarded to the University of California, San Diego, in close proximity to the Scripps Institute. The other predominant oceanographic institution, the Woods Hole Oceanographic Institution was awarded the ocean Observatories Initiative "Coastal/Global IO" grant.

With the main component grants allocated to the major oceanographic centers in the United States, the Scripps Institute, the University of Washington, and the Woods Hole Oceanographic Institution, delegates from all involved institutions met in Arlington for a Preliminary Design Review of the Ocean Observatories Initiatives' network components in December 2007. The central role of Scripps and Woods Hole in the planning, design, and implementation of the Ocean Observatories Initiative shows the great institutional continuity that the OOI knowledge infrastructure was built on as well as the power of existing institutional networks within the oceanographic community.

The OOI Final Design Review was discussed the following year at another meeting in Arlington.[240] Following the Final Design Review in 2008, the concepts for Cost and Schedule and Science Review for the construction of the Ocean Observatories Initiative were eventually completed in March 2009. Shortly after, in May 2009, the National Science Board authorized the overall funding of the OOI network projects. Actual funding for OOI construction began in September of 2009, nearly a full decade after the project had first been discussed by the US oceanographic community.

[240] OOI, "National Science Foundation Panel Report on the Final Design Review of the Ocean Observatories Initiative (Ooi) " (National Science Foundation, 2008).

I have already mentioned the predominance of existing centers of oceanography such as Scripps and Woods Hole in the design and implementation of the Ocean Observatories Initiative. The workshop reports and available sources do not show major controversies over the organization of the network, the site selection or institutional setup. We will see in the following case study on the National Ecological Observatory Network that there is much more potential for disaccord within a research community about how and where to set up an environmental research infrastructure. One reason for the lack of major conflict in the planning of OOI could have been that the major oceanographic institutions in the US divided responsibility among them fairly clearly: Woods hole was responsible for the East Coast, Scripps was responsible for cyberinfrastructures, MBARI focused on its experience with AVUs, and the University of Washington installed a cabled underwater observatory.

Thus, it appears that all major institutional actors were able to carry on with their own research priorities, while at the same time integrating into a larger environmental research network. After learning how closely oceanography had been entangled with large scale military surveillance efforts, one might also assume that the oceanographic community was comparatively well experienced in setting up a large knowledge infrastructure. One indication of this could also be the comparison with the issues the ecological research community had to face while planning, implementing, and managing the National Ecological Observatory Network, as we shall see in the following case study chapter 4.

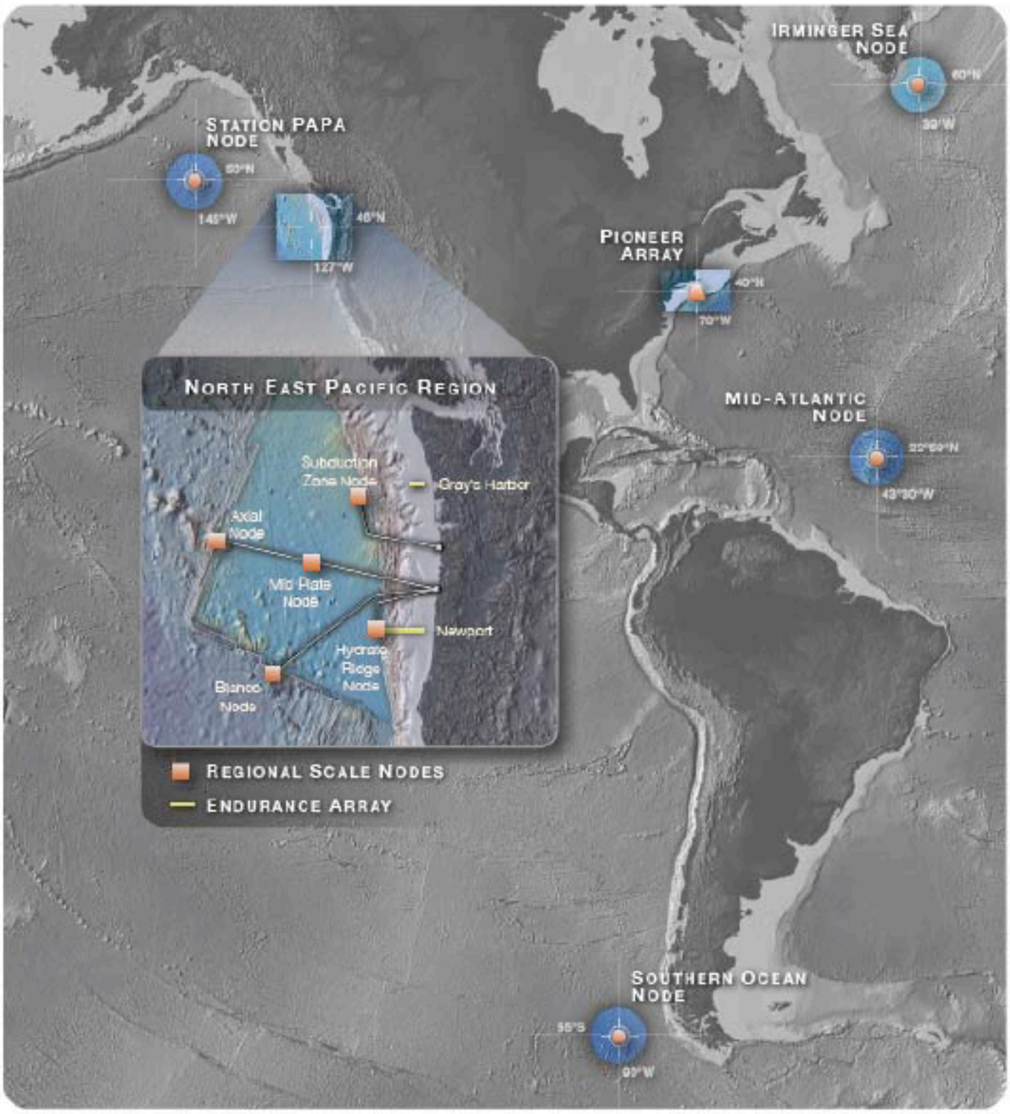


Figure 4 The OOI Network Design[241]

[241] "Final Network Design," ed. Consortium for Ocean Leadership (Washington, DC2011).

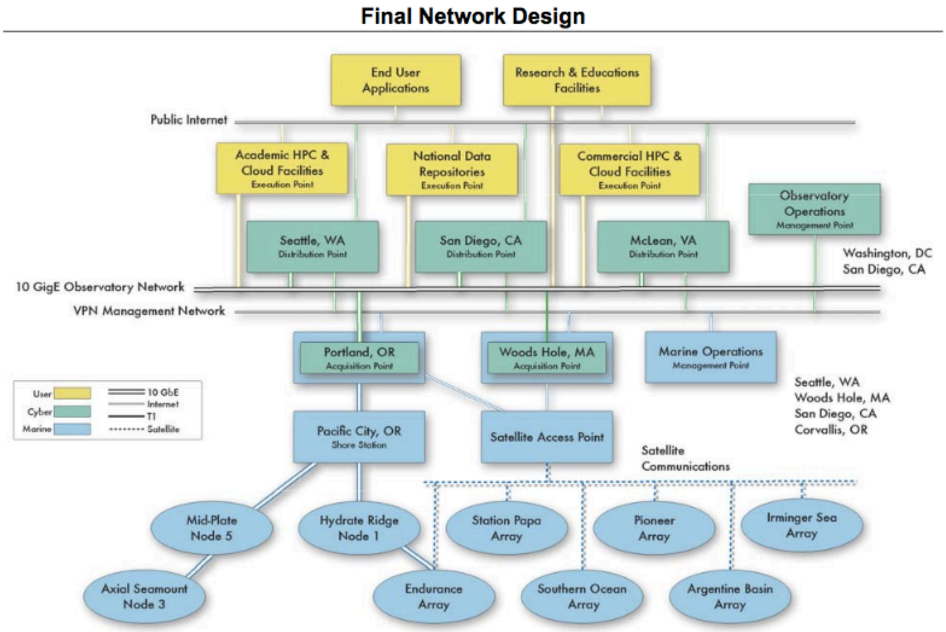


Figure 5 OOI Final Network Design[242]

The OOI Network Design

The figure above shows a schema of the 'Final Network Design' of the Ocean Observatories Initiative. It uses three colors to indicate the different kinds of components of the network: blue representing marine components, green denoting the cyber infrastructure of OOI and yellow representing the 'User' components of the network situated in the public internet domain.

The Ocean Observatories Initiative's marine components comprise six regional monitoring sites that are linked to a satellite access point. The measurements sites are connected by satellite to the Endurance Array, Station Papa Array, the Southern Ocean Array, the Pioneer Array, the Argentine Basin Array, and the Irminger Sea Array. Out of these, the Endurance Array is also connected to the submarine cable infrastructure installed by OOI. (See also Figure 4)

[242] Ibid.

The submarine cable infrastructure hits the shore at Pacific City on the coast of Oregon. The shore station at Pacific City is connected to both the Endurance Array via a cable node at Hydrate Ridge as well as to the measurement array at the Axial Seamount volcano via a cable node termed Mid-Plate Node 5.

Thus, the OOI network combines both cable and satellite transmission of information even among its marine components. Furthermore, the network does not create a tightly knitted grid, but links together very disparate sites that were deemed representative of their respective ocean areas by the oceanographers. The marine operations of the various network components are managed by the major institutions involved in the OOI project, i.e. the oceanographic institutions Scripps in San Diego and Woods Hole on the East Coast as well as in Oregon close to the submarine cable infrastructure and at the University of Washington in Seattle.

The two main acquisition points for all the data collected by the monitoring arrays and the submarine cable infrastructure are in Portland, Oregon, for the submarine cables, and at the Woods Hole Oceanographic Institution for all data transmitted via satellite. All data collected by OOI can then be distributed among the institutions involved via a closed VPN (Virtual Private Network). However, the actual 'backbone' of the infrastructure is the "10 GigE Observatory Network," i.e. a cabled optical fiber connection with a speed of 10 Gigabits per second, which is 1,25 Gigabyte or 1250 MB per second. This fast connection relates data from the data acquisition points in Portland and Woods Hole to the OOI data distribution facilities in Seattle, San Diego and McLean, Virginia.

The "End User Applications", meaning the OOI Data Portal (discussed below), access Ocean Observatories data only via the public internet and do not have direct access to the infrastructure backbone linking together the scientific institutions involved. Research and education, and other scientific institutions, however, are granted direct access to OOI data via the 'backbone'. Especially data repository providers and cloud services, both academic and commercial, are connected directly to OOI data. Furthermore, HPC (high performance computing) providers need a substantial connection to the observatories' data since doing the kind of data-intensive oceanography envisioned by OOI and Fourth Paradigm proponents would be inconceivable without substantial resources for data transfer from the site of storage to the site of analysis.

Considering the final network design, it has also become clear how crucial the cyberinfrastructure component of the infrastructure really was. Well beyond providing access to OOI observatories via a public internet platform, storage and sharing of data according to specific standards and protocols would be impossible and would render the massive efforts at constructing seafloor observatories offshore quickly obsolete.

3.4 Knowledge Sharing and Maintenance: Challenges of Ocean Observation

After discussing the conceptualization, implementation, and technology of the Ocean Observatory Initiative, we need to consider the way that knowledge generated by the infrastructure was actually maintained and shared among the scientific community and the public. As has been outlined above, the actual cyberinfrastructure formed the backbone of all data sharing within the observatory network and represented a crucial challenge to creating knowledge that could actually be maintained and eventually shared. From a lay perspective, the OOI data portal represents the main interface between the public and knowledge generated by the Ocean Observatories. The data portal's interface is going to be considered in more detail, especially since Jim Gray was involved in its conception as an expert.

Furthermore, the eruption of an underwater volcano proved to be the first larger challenge of OOI as an infrastructure that serves the public interest as well as that of the oceanographic community. Yet, I will discuss how, at the same time, the event of the eruption led to controversies around the purpose of a long-term ecological observatory in the service of 'broadcasting' individual, albeit newsworthy, events.

During planning of OOI, it had become increasingly clear how central a well-designed digital infrastructure was going to be if the project were to achieve its ambitious goals. The Ocean Observatories Initiative "Science Prospectus" outlined the main goal of the OOI cyberinfrastructure (CI) as follows:

> The CI will allow access to other (i.e., non-OOI) data streams to provide users with a coherent four-dimensional view of the ocean. Using the OOI CI, scientists will be able to, for example, combine OOI water column data with NOAA and NASA satellite ocean-surface imagery and NOAA Integrated Ocean Observing System (IOOS) subsurface data. These interactive and data-aggregation capabilities will complement parallel international efforts by Canada, Japan, and Europe and have the potential to change fundamentally how ocean science is conducted.[243]

The international efforts referred to have already been outlined above, Canada's NEPTUNE, the Japanese AREA as well as the European ESONET. The ultimate goal of the OOI cyberinfrastructure was to link up to other international efforts at ocean and climate observation to create an overarching 'system of systems'.

At the same time, the OOI cyberinfrastructure was supposed to make data available to the general American public to justify its use for educational efforts and citizen science. The OOI data policy stated that „calibrated and quality-controlled data must be made publicly available with minimal delay."[244] Yet, considering the amount of data, including streaming and real-time data, that OOI's sensors produce, not all data could be made equally accessible due to

---

[243] "Scientific Objectives and Network Design: A Closer Look," (2007). p. 21

[244] Ibid. p. 22

capacity constraints of the technology that was currently available. To address these limitations, the cyberinfrastructure envisioned so-called Cyberinfrastructure Points of Presence (CyberPoPs).

> These CyberPoPs include integrated, real-time data processing and archive sites located at a few central facilities and at marine observatory shore stations or control centers. CyberPoP capabilities include a secure, highly available, scalable computation design that can be deployed in environments ranging from moorings that may be extremely resource-constrained to the TeraGrid.[245]

The discussion of the OOI cyberinfrastructure makes clear how complex efforts were necessary to make the data produced by an environmental research infrastructure available. In fact, the data produced by neon comprises everything from real-time streaming via cabled nodes to readings from sensors on moorings transferred via satellite. To make all of this data available on one level via the Cyberinfrastructure Points of Presence required an effort by the entire infrastructure, its measurements and instruments as well as collaboration by the individuals and institutions involved. To make this data appear seamlessly integrated and make it available via one interface, OOI needed to construct a digital observatory, the Ocean Observatories Initiative Data Portal.

The Ocean Observatories Initiative Data Portal

After Jim Gray's disappearance at sea, the search for his missing ship had involved many from the oceanography community, especially in Monterey and San Diego. The journal SIGMOND Record published a tribute volume to Jim Gray in June 2008, to which James Bellingham contributed an essay titled "Exploring Ocean Data".[246] Ed Lazowska had introduced Bellingham to Jim Gray in the fall of 2004, when the two began discussing the needs for cyberinfrastructures in ocean sciences. After being provided with some data

---

[245] Ibid. p. 22

[246] James G. Bellingham, "Collaborative Oceanography and Virtual Experiments. Distribution Statement A.," (Monterey Bay Aquarium Research Institute, 2009).

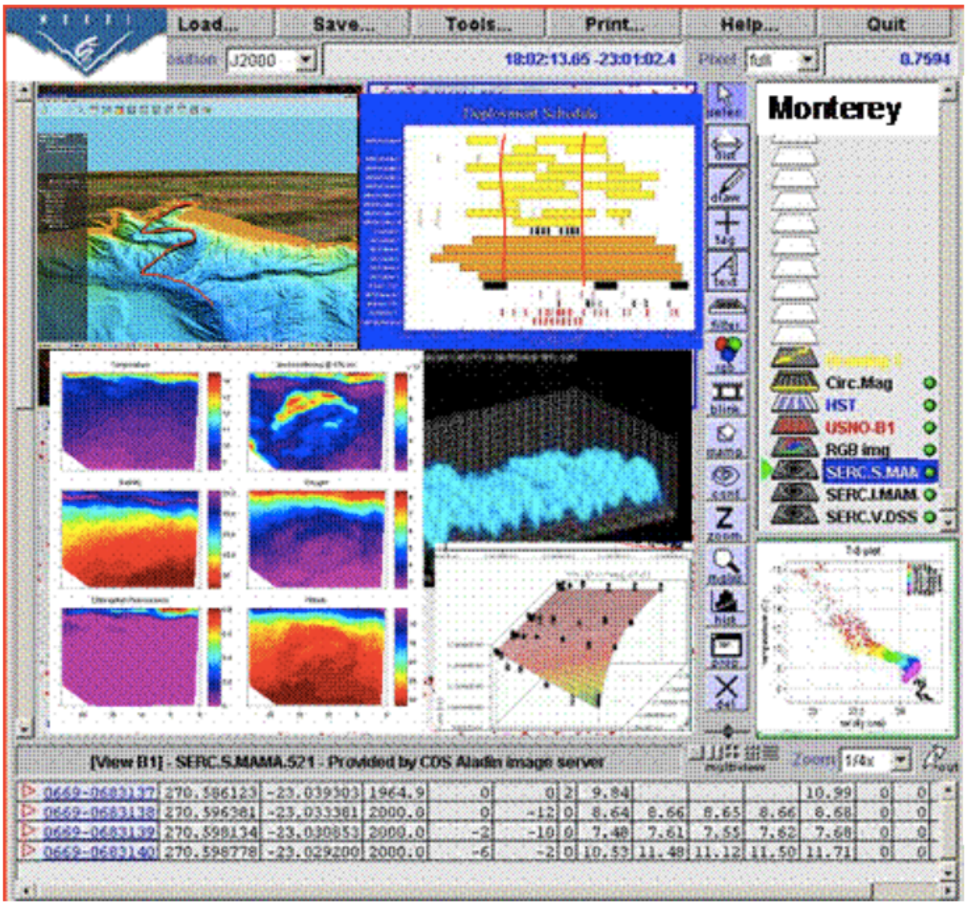


Figure 6 Jim Gray's concept for an OOI Data Portal

from the previous year, Gray sent Bellingham a mockup of a data exploration tool for oceanography in an email from April 28, 2005.

Jim Gray responded to his oceanographer colleague along with the mockup of a data exploration portal (see Figure 6):

> "1. I think the world wind viewer Keith is cooking up (or some derivative of it) will be the way to get oriented in space. One can ask for "tracks" or "platforms" or "footprints" (for satellite or survey data) be rendered as layers above the backdrop. These layers could be selected from a list and turned on/off (world wind has a prototype for that). Lets [sic] call that the LOCATION window. (upper left windows in screen shot below). Other windows can send the location window events and as you move around the location window it can send events to other windows.
>
> This "BRUSH" effect is a fairly intuitive way to explore multidimensional data.

As you scroll through the "DEPLOYMENT SCHEDULE" window, it would affect what is rendered in the LOCATION window. (I show a time limit on the deployment window affficting [sic] the other windows and a track on the LOCATION window affecting the others)

2. Other windows can have plots of x vs y and x vs y vs z for any x,y,z you care to define. When those dimensions are spatial or temporal there is an obvious backdrop but generally they are not. The pane at right controls which layers are visible in each window.

3. So now we are into defining x,y,z. They can be a database query but more likely they are the output of some analysis tool. DB queries are "easy" but require the scientist to think at a low level and speak a funny language. We will start with that and that will be there as an escape hatch in case the analysis tool does not do what is needed, but... The goal is for the DB to be hidden." (Email Jim Gray to James Bellingham, April 28, 2005)[247]

The "world wind viewer" referred to in Gray's email was a prototype ocean geobrowser developed by the computer scientist Keith Grochow at the University of Washington.[248] The tool comprised a number of windows serving different selection and output functions, but ultimately hides the underlying database below the graphic interface.

The Location window was supposed to orient the user with respect to the location of the selected input measurement sites, while the Deployment Schedule provides orientation about the timeframes of when the sensors actually and actively delivered data. A pane on the right also allows the user to select different levels of data, which are then visualized in the Location window as well as displayed as charts within several more customizable windows. Location, Deployment Schedule, and the selection of layers are simply an interactive way of querying the database for a certain set of data to be displayed. The final result of the query would, in any case, be a number of multi-dimensional graphs with customizable settings.

---

247 Ibid.

248 Keith Grochow wrote his PhD Thesis on virtual observatories for oceanography, see Keith Grochow, "The Design of Cove: A Collaborative Ocean Visualization Environment " (University of Washington, 2011).

Thus, this kind of data exploration tool is merely just that, an interface for exploring an underlying database. While its strength would have been to spare a user the need to query a database in a database management query language, such as SQL, the tool did not do any data analysis as such, and a user had to be both familiar with the available data as well as be very concise about the kind of information she wanted to display. However, the strengths of Jim Gray's concept are that it would have allowed a user to superimpose a structure over the database that is not dependent on the structure of the infrastructure as such. This would have freed the user from the issues arising due to the diversity of data circulating within the OOI cyberinfrastructure.

Considering the actual and current state of the Ocean Observatories Initiative Data Portal, the interface is much more closely bound to the structure of the underlying database and data sources than Jim Gray's idea had envisioned.

Since it is part of the Ocean Observatories Initiative's mission to make data available via the public internet as directly as possible, the OOI Cyberinfrastructure Consortium has created an OOI Data Portal that is accessible online at https://ooinet.oceanobservatories.org/. (See Figure 7).

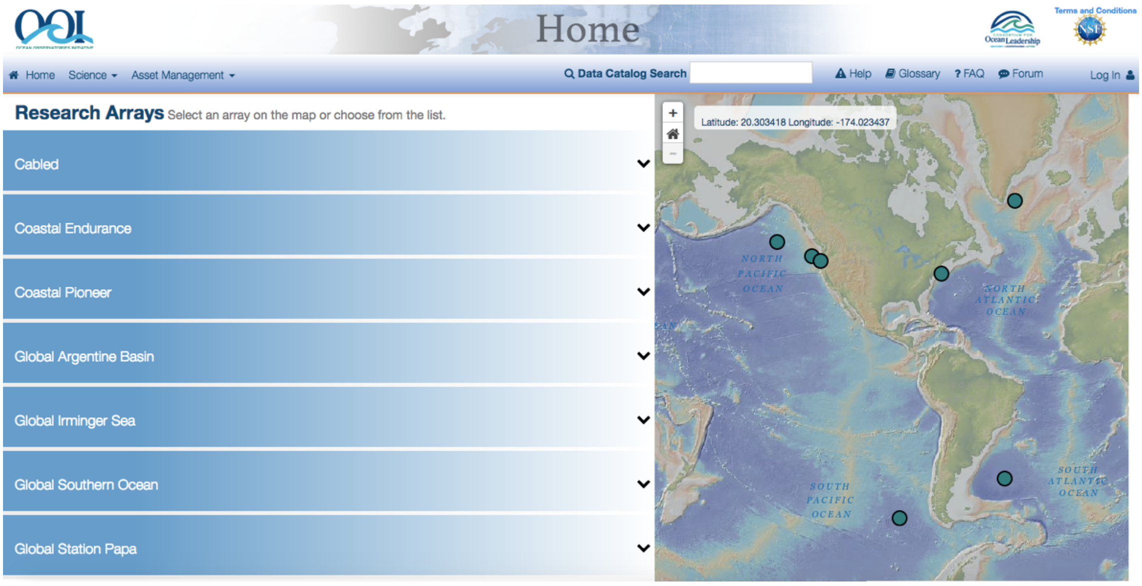


Figure 7 The OOI Data Portal

One can see at first glance that the OOI Data Portal user interface is organized according to the location and type of the various research array components of the Ocean Observatory Initiative infrastructure. The three types of array structures are the cabled observatories off the coast of Oregon, the coastal Endurance array in Oregon as well as the Pioneer array on the East Coast, and the four global nodes of moored buoys.

However, one feature of the interface logic proposed by Jim Gray remained in place. The first layer of selection is a location map, from which a user can select between the various data sources. Yet, the logic of selecting the data sources as the primary structure of the database is dominant in the current Data Portal. Upon selecting an individual array, the portal provides a list of all individual moorings that comprise the array as well as a detailed list of all individual measurement devices and sensors installed at this mooring.

At the next level, a user can choose between "Data Access & Plotting" and "Assets and Events". The latter gives an overview of when OOI's various moorings, seafloor observatories, and underwater vehicles were actively recording and observing. One can then move on to "Data Access & Visualization" and select inputs from all different sensors fitted to all of OOI's observatory assets. The data portal offers options of plotting the data in various ways. However, to even know what one is looking at, what the different sensors are measuring, and how one might apply this information to a scientific study requires at least some level of oceanographic expertise. Although it was an explicit goal of the Ocean Observatories Initiative to make environmental research data available to everyone, it did not at all make it usable or comprehensible to just anyone.

Any lay user of the data portal is likely to quickly give up or move on to the tab labeled 'camera' which promises more entertaining content of a submarine volcano. As we are going to see in the following section, OOI planners themselves appear to have quickly realized that data itself was a hard sell and more visual stories worked much better to try to justify the usefulness of the infrastructure to the public.

Racing an Underwater Volcano

The physical and digital infrastructure of the Ocean Observatories Initiative was set for completion in May 2015. However, a *Nature* article by Alexandra Witze from November 2014 revealed that the project had encountered issues with the implementation of its cyberinfrastructure.[249] Originally, the contract to develop cyberinfrastructure had been awarded to the University of California, San Diego, and at the time, the OOI had already spent 37 million US Dollars to develop data management software.

Yet, in the fall of 2014, OOI management decided to terminate the contract with UCSD and instead tasked a group at Rutgers University with the database development.[250] Tim Cowles, who oversaw the OOI at the Consortium for Ocean Leadership cited delays in the project development schedule as the reason for the contract termination. The San Diego team,

---

[249] Alexandra Witze, "Ocean Observatory Project Hits Rough Water," *Nature*, 27 November 2014 2014.

[250] Ibid.p. 474

headed by John Orcutt, stopped working on the OOI infrastructure by November 1st, 2014, and started to transfer the operation to their colleagues at Rutgers.[251]

The termination of the contract with the San Diego group appears curious at first glance since a delay of a few months is not particularly unusual in a project spanning nearly a decade of technological development. Also, transferring an entire project to a new provider just months away from a deadline for completion constituted a major risk in itself. The new personnel would have to learn the ropes and adjust to the requirements of a new project while facing a tight deadline from day one. A major factor that pushed the hurried implementation of the OOI network along appears to have been an event that unfolded on the ocean floor off the coast of Oregon. The underwater volcano "Axial" was approaching an eruption. The University of Washington team around John Delaney had been working frantically to complete the connection and installation of various measurement instruments on the site of the Axial volcano. Their goal was to enable a real-time observation of the underwater eruption and make the data available to the community via the envisioned OOI digital platform. "I want the data in the hands of the community. That's what it was all about to begin with," Delaney was quoted in *Nature*.[252]

Growing more and more anxious about the delay of the cyberinfrastructure components, Delaney's team began working on an alternative using a data center at the Washington D.C.-based Incorporated Research Institutions for Seismology.[253] Facing waning acceptance within the scientific community and high ongoing costs, OOI had yet to deliver any valuable observation data to scientists and the public. Thus, a newsworthy event such as the eruption of a volcano underwater appears to have appealed to Delaney as an opportunity to prove the infrastructure's merits in a spectacular and tangible way. Considering

---

[251] " Us Ocean Sciences Told to Steer a New Course," *Nature* 517, no. 7536 (2015).

[252] "Ocean Observatory Project Hits Rough Water."p. 475

[253] Ibid.

the broad media interest in the event, Delaney actually appears to have succeeded. In December 2016, the *Washington Post* reported on the real-time monitoring of the Axial Seamount's eruption captured by the OOI infrastructure.[254] The seafloor at the Axial Seamount split open on April 24, 2015, starting off the third largest underwater volcanic eruption on record. The Axial Seamount had been remarkably active in the preceding decades, erupting both in 1998 and 2011. The eruption in 2015 was observed by the OOI infrastructure and data flowed in real time, being stored in the cyberinfrastructure to be accessed by the public afterwards. The *Washington Post* quoted William Willcock from the University of Washington stating that "these are 'the most detailed observations ever made' of an undersea volcano."[255]

Although the Axial Seamount eruption had occurred the year before, the first successful use of real time monitoring by an OOI infrastructure created media attention only in December 2016. National Geographic also ran an extended article on the eruption, commenting that "fortunately, scientists had installed an elaborate volcano-monitoring network on Axial Seamount just a few months earlier, making the submarine mountain one of the world's most wired volcanoes."[256] It was an article in the December 2016 issue of *Science* that eventually brought the journalists' attention to the Axial Seamount eruption.[257] The article was picked up by several news sources and constitutes the first public impact that the Ocean Observatories Initiative was able to generate. Considering that the legitimacy of the OOI was already being questioned within the research community as well as by the National Science Foundation itself, this was a key success for the project.

---

[254] S. Kaplan, "A Massive Underwater Volcanic Eruption Is Captured in Real Time," *Washington Post*, 16 December 2016 2016.

[255] Ibid.

[256] N. Drake, "Underwater Volcano Offers Rare Look at Eruption in Real Time," *National Geographic*, 15 December 2016 2016.

[257] Scott L. Nooner and William W. Chadwick Jr., "Inflation-Predictable Behavior and Co-Eruption Deformation at Acial Seamount," *Science* 354, no. 6318 (2016).

The pressures to get the infrastructure up and running in time for the expected submarine eruption was evident. However, it is clear that the rationale in this particular case was both one of scientific value as well as of public relations interests. The particular way that OOI was conceived created the need for justification not just of scientific value but also of some value to the general public, which is supposed to be the audience for the network's 'data products'. We are going to discuss later, how this double bind entails problematic contradictions for a large scale digital scientific infrastructure project such as the Ocean Observatories Initiative.

Larger Issues: Current State and Outlook

In June 2016, *Nature* journalist Alexandra Witze summed up the outlook for the Ocean Observatories Initiative in an article on occasion of the launch of the network: "the OOI's future remains murky. A 2015 review of US ocean-science priorities suggested that the programme's operational budget should be slashed by 20%, to around $44 million a year."[258] Constructing the OOI network infrastructure cost no less than 386 million US Dollars. "Yet each of the arrays must be serviced every year or two to replace broken instruments and install new ones."[259] "The NSF has not yet decided how it will save that 20%," Witze went on to comment. Thus, immediately after its launch, the Ocean Observatories Initiative faced an uncertain future, with its overall value to the scientific community and the public still unproven.

The chart below (see Figure 8) shows the development of the National Science Foundation's spending on oceanic research. While the overall funding as of 2014 hovered around 350 million US-dollars, there has been a steady rise of infrastructure spending at the expense of funding for individual research

[258] Alexandra Witze, "Massive Ocean-Observing Project Launches — Despite Turmoil. Network of Deep-Water Observatories Streams Data in Real Time.," *Nature* 534, no. 7606 (2016).

[259] Ibid.

projects. The Ocean Observatory Initiative has been the biggest driver of this funding expansion since the year 2000. The implementation of OOI resulted in infrastructure spending actually surpassing research funding for the first time in 2012.

At the time of writing, the uncertain funding outlook for OOI remains a threat to the long-term maintenance of the infrastructure. However, this issue points to a larger general problem that environmental research infrastructures for long-term environmental monitoring are going to face: a disconnect between the 'needs' of the infrastructure and the workings of the political cycle and its influence on science policy and research funding institutions.

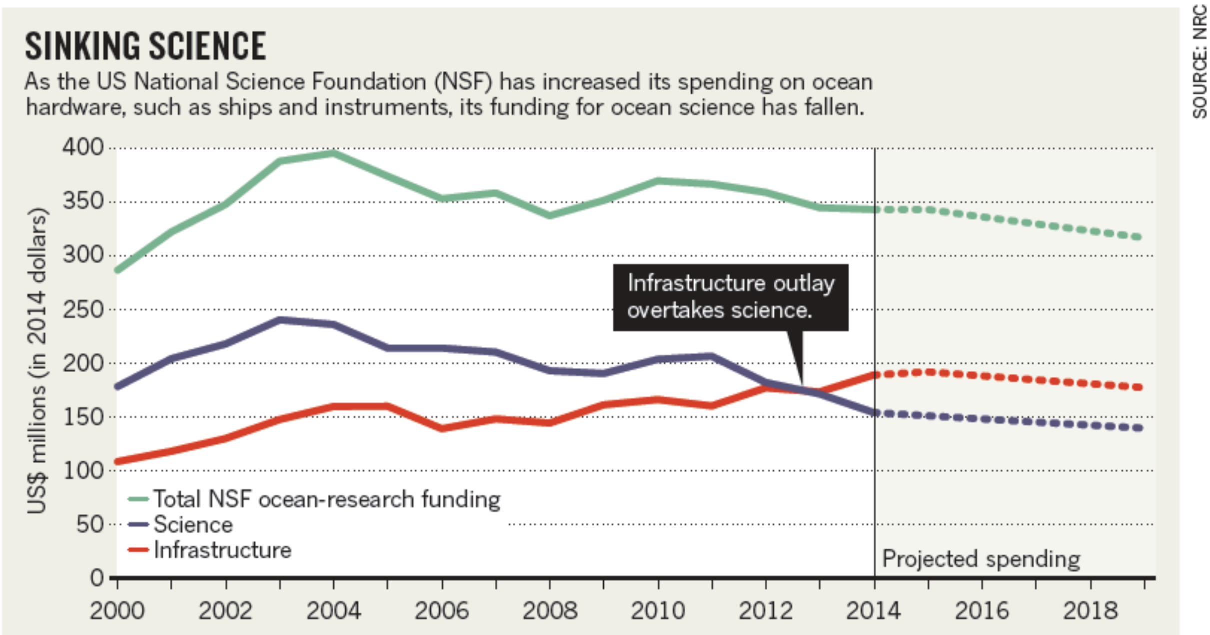


Figure 8 NSF Ocean Sciences Budget[260]

[260] Ibid.

3.5 Conclusion: A Digital Ocean

After having outlined the conception, implementation, and challenges of the Ocean Observatories Initiative I will attempt to draw some conclusions from this case study. First, I need to ask whether OOI can be regarded as an instance of what Jim Gray called Fourth Paradigm Science. And secondly, I will outline some challenges that the case study revealed that may be generalizable for data-intensive environmental observation in the future.

To ask whether the Ocean Observatories Initiative constitutes a case of Fourth Paradigm Science, we need to recapitulate Jim Gray's characteristics for Fourth Paradigm science:

1) Data captured by instruments or generated by simulator
2) Processed by software
3) Information/knowledge stored in computer
4) Scientist analyzes database / files using data management and statistics[261]

In the case of the OOI, data is not generated by a simulator, but all data generated by the infrastructure is captured by instruments in real-time and then transferred via submarine cables or even satellite links. Processing of the generated data by software also takes place. However, there is an inherent problem with postulating that in Fourth Paradigm Science, data is processed by software. Even without reference to Gitelman's book *Raw Data is an Oxymoron*, it is hard to conceive of any meaningful 'raw' data that is not somehow processed by software or had to be processed by some kind of algorithm to be explored and visualized. It is exactly this context-dependency of data that entails the risk of losing compatibility in the near future due to progress in hardware and software tools. Thus, while data processing by

[261] Hey, Tansley, and Tolle, *The Fourth Paradigm. Data-Intensive Scientific Discovery.*

software is clearly to be found in the OOI infrastructure, this is likewise not a sufficient condition to diagnose Fourth Paradigm Science.

What about "information/knowledge stored in a computer" as a condition with respect to the ocean observatories? In case of the OOI infrastructure, information is indeed being stored within a large and distributed database. This system is physically distributed and accessible to expert researchers as well as via the public internet. OOI information storage is thus a kind of cloud storage system that is centrally managed by the Rutgers University-based cyberinfrastructure lead institution. This kind of distributed and 'cloud'-based information storage system is also a feature that distinguishes the OOI infrastructure significantly from a typical 'Third Paradigm' science project. The accessibility to the public internet network is also a unique feature that at least add the potential for new applications in citizen science.

The final characteristic of the Fourth Paradigm is the use of data management software in accessing the collected data. I have also already discussed that the application of statistical methods as such is not a sufficient distinguishing feature of a Fourth Paradigm, since statistical methods have been part of scientific methods long before the invention of digital technology. However, it is the automated application of such methods that distinguishes the infrastructure of the Ocean Observatory Initiative. It is not the application of statistics but the features of a digital observatory interface that serves as an interface for the entire infrastructure and not just one experimental setup. We have seen how Jim Gray was involved in creating a prototype of such a virtual observatory interface for OOI.

Thus, we can conclude that out of Gray's four conditions for Fourth Paradigm Science, the use of a virtual observatory as an interface for an entire digital knowledge infrastructure network is the central necessary condition. In essence, Fourth Paradigm Science is marked by the deployment of a virtual observatory that mediates a knowledge infrastructure and its underlying database management systems.

This interface is only as valuable as what goes on in the black-box of the knowledge infrastructure. Thus, in conclusion, we have to take a look at how knowledge within OOI is not only generated but shared and maintained over time. These considerations prefigure our reflections on policy recommendations in the concluding chapter 5.

This case study has examined issues of knowledge generation, knowledge sharing, and infrastructure maintenance. In its current state, the OOI infrastructure generates knowledge in the form of various data products. Knowledge sharing has been implemented via the OOI online portal providing access to practically all of the data generated by the various nodes of the OOI network. However, the goal of making this information accessible to the lay public has only been reached in principle, not in practice. It is possible to access OOI data even as a private individual without any expertise in oceanography, and yet, the virtual observatory does not manage to actually encourage non-experts to engage with the data in a meaningful way. If the aim of public engagement via the OOI infrastructure is to be realized as more than lip-service to funding agencies, there are still major challenges to be overcome in actually engaging the lay public in the research data.

Most importantly, we have seen that infrastructure maintenance, particularly data maintenance and curation, is and will be an ongoing and crucial issue for the OOI infrastructure. The enormous recurring costs for annual maintenance can not reliably be supported by a National Science Foundation funding structure as it exists today. If large scale knowledge infrastructure such as OOI are supposed to be operational and viable over a longer time, funding agencies such as the NSF will have to rethink the way they structure their infrastructure funding. Currently, OOI represents a large amount of sunk costs and the scientific community is unlikely to simply abandon the project.

Yet, if funding is to remain competitive, some assessment of the returns and uses of an infrastructure has be to in place to prevent runaway costs and an efficient use of resources to reach maximum value for the scientific community. Thus, when it comes to long-term funding of maintenance costs, the future of OOI and similar infrastructures is likely to remain uncertain and will also depend on the whims of government administrations and their particular agendas. This is a key difference between knowledge infrastructures such as environmental research infrastructures and public infrastructures such as transportation infrastructure. While the utility of a transport infrastructure is immediately apparent upon use, a knowledge infrastructure has to constantly justify its utility through narratives linking to contemporary concerns of security, risk or public participation.

Planning and justification of the OOI was strongly driven by ideas of Fourth Paradigm Science that promised a new digital observation of oceans. However, it was also driven by institutions and individual actors that wanted to promote individual scientific projects such as the seafloor observation of the Axial seamount volcano. Echoing ideas of a Fourth Paradigm served both as inspiration and as a justification for the OOI network.

The infrastructure's research objectives were, at first, largely aligned with the aims of the Ocean Observatories Initiative, yet, the science policy objectives, crucially of the National Science Foundation, shifted over time. At the time of the completion and launch of the network, the use of the OOI was being questioned from within the oceanographic community. Eventually, a shift in priorities manifested itself in a slashing of funding for OOI infrastructure by 20% just before OOI's completion in 2015. The NSF appeared to have realized that its ocean-science budget had become dominated by infrastructure spending, which overtook the funding of individual research around 2014. The OOI was mostly responsible for this shift caused by a skyrocketing infrastructure budget.

Stewardship of the OOI network is thus also in question, since OOI constitutes an infrastructure that put more money into construction than it put into the maintenance of the network infrastructure. Both the physical components exposed to wear and tear as well as the database technology are likely to need substantial updates and sustained data curation over the next 25 years, the timeframe of the project. While the planning of OOI envisioned its use for long-term and real time observation, justifying its use via the collection of long-term datasets to enable researchers to identify trends in the impact of climate change on ocean systems, the implementation of OOI was largely driven by other aims. Firstly, by the individual research agendas and preferences of the major investigators involved, and secondly, by specific events such as the eruption of the undersea volcano Axial off the coast of Oregon.

Thus, all Fourth Paradigm Science infrastructures and projects are likely to face the tension between the individual scientists comprising a research community and the ideals of the Fourth Paradigm deployed as a legitimizing narrative. In addition, individual and unusual events also represent a paradigmatic contradiction for Fourth Paradigm Science. A research infrastructure that is built for long-term monitoring and detecting subtle and gradual shifts in an environment is not inherently well suited to provide research data on spectacular events such as a volcanic eruption. We have seen in the case study this opposition between long-term monitoring and real-time event 'broadcasting' led to conflicts in implementing the infrastructure. This contradiction is likely to occur in similar environmental research infrastructures, as we are going to see in the case study of the National Ecological Observatory Network.

Finally, the use of the OOI project and its enormous construction and maintenance costs have been questioned frequently by members of the oceanographic research community. The reasons for these critiques are doubts about the efficiency of the investment into an infrastructure such as OOI, on the one hand. On the other hand, the OOI has been proposed, run, and implemented by a group of oceanographers, who were successful in securing funding for their own institutions and research groups with a strong computational focus. This may well be the cause for resentment among colleagues who have a different idea about how to conduct oceanographic research and are skeptical about the many statements of 'revolutions' and 'future' shape of ocean science.

It remains to be seen how a project such as the Ocean Observatories Initiative is going to change the role and image of oceanographers in the years to come. It seems that the discipline is in the midst of a transformation similar to the one oceanography underwent after the end of the Cold War. However, not all oceanographers will be equally eager to transform themselves from explorers of the seas to desk-bound explorers of oceans of data.

Chapter 4: Case Study National Ecological Observatory Network (NEON)

4.1 Introduction – Context and Methods

The following chapter examines the second case study in the inquiry into data-intensive environmental monitoring infrastructures: the National Ecological Observatory Network, NEON. In many ways, NEON represents the continental counterpart to the Ocean Observatories Initiative, although it is even more expansive in scope. NEON and OOI were the most ambitious scientific infrastructure projects conceived and launched since the 1990s in the US. At the time of writing, they were the largest budget items in the Major Research Equipment and Facilities Construction (MREFC) funding of the National Science Foundation.

Table 1. MREFC Account Funding, by Project

(dollars in millions)

	FY2011 Actual	FY2012 Estimate[a]	FY2013 Request	FY2014 Estimate	FY2015 Estimate	FY2016 Estimate	FY2017 Estimate	FY2018 Estimate
AdvLIGO	$23.58	$20.96	$15.17	$14.92	—	—	—	—
ALMA	13.92	3.00	—	—	—	—	—	—
ATST	5.00	10.00	25.00	42.00	20.00	20.00	9.93	—
IceCube[b]	5.29	—	—	—	—	—	—	—
NEON	12.58	60.30	91.00	98.20	91.00	80.66	—	—
OOI	65.00	102.80	65.00	27.50	—	—	—	—
MREFC Total	**$125.37**	**$197.06**	**$196.17**	**$182.62**	**$111.00**	**$100.66**	**$9.93**	—

Source: U.S. National Science Foundation, *FY2013 Budget Request to Congress*, p. MREFC-1.

Notes: Totals may not add due to rounding.

a. In FY2012, $30.0 million was transferred from the R&RA to the MREFC, as provided by the Science Appropriations Act, 2012, P.L. 112-55.

b. IceCube and South Pole Station Modernization are expected to report FY2012 actual funding from FY2011 carryover.

Figure 9 The National Science Foundation's Major Research Equipment and Facilities Construction budget from 2013.[262]

[262] Christine M. Matthews, "U.S. National Science Foundation: Major Research Equipment and Facility Construction," in *Report for Congress*, ed. Congressional Research Service (2012).

The method of this NEON case study is going to be the same as in the preceding case study. I had assembled major concepts by Paul Edwards into an Infrastructure Matrix (see chapter 3.1) to show the main actors and activities central to the case study analysis. To recapitulate, the main actors were the people, artifacts, and institutions involved in knowledge generation, knowledge sharing, and infrastructure maintenance of the case study project.

I will not have the space to discuss all of these actors and activities in equal depth. However, we will see that in the case of NEON, the debate between a networked versus a systemic character of the planned infrastructure was very controversial among the scientific community and administrators involved. In addition, NEON is an infrastructure project with an envisioned annual maintenance cost of 65 Million US-dollars, and thus, I will also focus on the challenges this poses for such a long-term environmental monitoring infrastructure. The parallels to the Ocean Observatories Initiative's struggles with organization, oversight, and long-term funding are going to become apparent in the case study as well.

In 2015, an article in the 24 September 2015 issue of *Science* announced "Ecology's megaproblem."[263] The article described how the continental scale science infrastructure project National Ecological Observatory Network (NEON), funded with 434 Million US-dollars by the National Science Foundation, began to run into substantial troubles during the fall of 2015.

Scott Ollinger, an ecologist from the University of New Hampshire, had been named Observatory Director in 2013, yet soon learned that the NEON project was headed for financial troubles. Since 2007, the top scientific post within NEON, the Observatory Director, had been held by no less than 5 different people. Jeffrey Mervis wrote in *Science* that there had been "management

[263] See http://specialprojects.sciencemag.org/neon/ J. Mervis, "Ecology's Megaproblem. Fledgling National Observing Network Faces Harsh Realities," *Science*, 24 September 2015 2015.

problems that have dogged NEON since its birth" and that there had been a "tense relationship with the community of scientists who will ultimately use its data."[264]

In December 2014, a hearing of the Committee on Science, Space and Technology had discussed the results of an audit of the National Ecological Observatory Network, in which Allison Lerner, Inspector General of the National Science Foundation, had reported that NEON had used "management fees for questionable expenditures."[265]

In April 2015, these first warning signs of mismanagement of the NEON project were exacerbated by a disclosure by the Federal Auditor J. Kirk McGill of the Defense Contract Audit Agency, part of the US Department of Defense.[266] In his disclosure, McGill accused the National Ecological Observatory Network managing entity NEON, Inc. of gross mismanagement, abuse of authority as well as gross waste of funds. Specifically, McGill wrote, "my team discovered that the grantee [NEON] was illegally funneling millions of dollars of taxpayer money into unlawful activities including lobbying, extravagant Christmas parties, employee gifts, and the like."[267]

Thus, in 2015, the most ambitious environmental research infrastructure project that the United States ecology community had ever undertaken was in serious trouble. While costs were exploding, hundreds of millions had already been invested in technologies and observatories all over the country. Eventually, NEON pulled through and is currently starting operations. However, this case study will show how the National Ecological Observatory Network ended up

264 Ibid.

265 Allison C. Lerner, "Statement of Allison C. Lerner, Inspector General National Sciece Foundation before a Hearing of the Committee on Science, Space, and Technology," ed. United States House of Representatives (Washington, D.C.2014).

266 Joshua Kirk McGill, "Disclosure Pursuant to Allegations Related to Suspected: Fraud, Gross Mismanagement, Gross Waste of Funds, Abuse of Authority Etc," ed. United States Congress (Washington D.C.2015).

267 Ibid. p. 3

on the brink of failure and what issues could have contributed to and exacerbated the problem that almost terminated the project. Yet, before I delve into science policy issues, let us take a step back at the forerunners of large-scale ecological observation infrastructures in the United States.

4.2 Historical Background – Networks in Ecology

The historians of science Elena Aronova, Karen Baker, and Naomi Oreskes have extensively discussed the development of 'Big Science' projects in ecology in the United States after the Second World War in their illuminating paper "Big Science and Big Data in Biology".[268] Aronova, Baker, and Oreskes suggested that biology and ecology lacked a model for Big Science after large scale scientific efforts had been dominated by the physical sciences, especially by nuclear physics and engineering as in the case of the Manhattan Project and the Apollo space program during the postwar period.[269]

The International Geophysical Year (IGY) in 1957 and 1958 provided a more viable model for Big Science efforts in the natural sciences, since it emphasized large scale data collection and data sharing. The first large effort in ecological sciences, inspired in part by the scientists' experiences during the IGY, was the International Biological Program (IBP), which ran from 1964 to 1974. We have already seen in the previous case study that the IGY was an important point of reference for infrastructure projects and co-operations in oceanography as well.

---

[268] Elena Aronova, Karen Baker, and Naomi Oreskes, "Big Science and Big Data in Biology: From the International Geophysical Year through the International Biological Program to the Long-Term Ecological Research Program, 1957-Present." *Historical Studies in the Natural Sciences* 40, no. 2 (2010).

[269] For more on Big Science in biological sciences see Vermeulen, "Big Biology: Supersizing Science During the Emergence of the 21st Century." and Sabina Leonelli, *Data-Centric Biology. A Philosophical Study* (Chicago: University of Chicago Press, 2016).
Two forthcoming books are also relevant:
Bruno J. Strasser, *Collecting Experiments. Making Big Data Biology* (Chicago: University of Chicago Press, 2019). Sabina Leonelli, *La Ricerca Scientifica Nell'era Dei Big Data* (Rome: Meltemi, 2018).

“In the United States in particular, the IBP [International Biological Program] was seen by its planners as a means to promote a Big Science model of research in biology and to transform ecology [...] into a modern Big Science.”[270] Yet, while the IGY had an immediate geopolitical appeal, and thus a unifying theme of global co-operation to gather knowledge about the entire globe, the International Biological Program struggled to establish such a unifying theme and motivation. When biological productivity and human welfare were chosen as a focus, things became ever more complicated as the IBP began to be regarded as infused with various political concerns, which made researchers uneasy. As Aronova et. al. point out, “ecology and environmental politics evolved to the point where they became inseparable in the public imagination, the topic of biological productivity in its relation to environment and overpopulation problems had come to denote an overtly political concern.”[271]

Yet, it was not simply the concern over a perceived politicization of their research that fueled discontent among the American ecology community. Many researchers were also unhappy with the large amount of bureaucracy, paperwork, and budgeting a complex Big Science project entailed. The IBP had also aimed to set up large data collection centers to make its results available to other researchers and public administrators, however, after internal disagreements among the institutions involved the data storage centers were never realized before the IBP terminated in 1974.

The International Biological Program had failed to deliver on its promises mainly because it had failed to set up an adequate method for data collection. Modelling based on the large amounts of data gathered was also considered to be a failure, and “by the end of the program the entire approach of all-encompassing models – whether deterministic or probabilistic – was declared

[270] Aronova, Baker, and Oreskes, "Big Science and Big Data in Biology: From the International Geophysical Year through the International Biological Program to the Long-Term Ecological Research Program, 1957-Present." p. 186

[271] Ibid. p. 200

'dead or near a dead end.'"[272] Despite the failure, the National Science Foundation pressed on and decided to continue funding several projects that were part of IBP under a new program called Long-Term Ecological Research (LTER), initiated in 1980.[273] LTER was organized in a much more decentralized way, granting independence to the various research sites. This decentralization was also mirrored in its setup for data storage, "data were now stored in a mix of site-based, network-based, and theme-based digital data repositories, accessible online."[274]

Geoffrey Bowker and Karen Baker point out that LTER had succeeded to reconcile the ecological science community with projects for long-term data collections and collaboration, stating that LTER "provides a sheltered forum in which to explore information management grounded within a scientific program and to consider the meanings and impacts of interdisciplinarity, data sharing, and technology use on the work of long-term research."[275] The Long Term Ecological Research program's success also seemed to prove the merits of a distributed and decentralized data-gathering program versus a more systemic infrastructure. Thus, LTER's success was a direct prerequisite for the initiation of the National Ecological Observatory Network. The ecological science community's experience with the decentralized LTER organization should also inform a major controversy over the design of NEON, as we are going to see in the following sections.

---

[272] Ibid. p. 217

[273] Ibid. p. 218

[274] Ibid. p. 223

[275] Geoffrey C. Bowker and Karen Baker, "Information Ecology: Open System Environment for Data, Memories, and Knowing," *Intelligent information systems* 29 (2007).

4.3. Knowledge Generation: People, Institutions, and Artifacts

4.3.1 The Biodiversity Observation Network (BON)

Hosted by the National Science Foundation, a group of ecologists met at the University of Virginia's Blandy Experimental Farm in early September 1998 to discuss plans for a new research infrastructure in ecological sciences.[276] The group's final report referred to the success of the LTER (Long Term Ecological Research) network and proposed a network of up to 50 monitoring sites to observe "crucial questions regarding biodiversity that are not tractable without a [sic] better understanding its temporal and spatial patterns."[277] The network of observatory sites was supposed to be supported by a so-called Biodiversity Technology and Analysis Support Center (BTASC), linking the various components of the research infrastructure. At this point, the core area of research was envisioned to be biodiversity, yet, the idea of opening up the infrastructure to the entire scientific community was already formulated, stating that the "observatories and associated data will be available to all researchers."[278]

In 1998, the technology to implement this sort of environmental research infrastructure was not yet readily available, considering, e.g., that Jim Gray had just launched TerraServer at Microsoft Research. The meeting report acknowledged this, stating that it "is expected that many of the new technologies and analyses required for the complex study of biodiversity will require original research," especially in the fields of "metadata" and "distributed access to data."[279]

---

[276] BON, "Final Report: Biodiversity Monitoring Workshop. An National Biodiversity Observation Network," (University of Virginia' Blandy Experimental Farm: sponsored by the National Science Foundation, 1998).

[277] Ibid. p. 2

[278] Ibid. p. 4

[279] Ibid. p. 6

However, a core question that the meeting appears to have posed was left open; the researchers were asking whether the Biodiversity Technology and Analysis Support Center should reside in "a single location or should it be dispersed at different locations according to functional considerations?"[280] This question already showed the tension between constructing a distributed research infrastructure and the established structures and institutions within the field of ecological research as well as the funding bodies. We will later see how, similar to the Ocean Observatories Initiative, the tensions between institutionalized structures, implicit personal networks, and the structure of a distributed knowledge infrastructure could never be fully resolved.

During the following year, in 1999, the National Science Foundation hosted three further workshops to develop plans for what was then called Biodiversity Observation Network (BON), the first one taking place in January 1999 in Santa Barbara, California. The workshop report placed the project in a larger societal context, claiming that knowledge about biodiversity "is critical to science and society – for maintaining the nation's natural resources, for growing its economy, for sustaining human health, and for improving the quality of human life."[281] The report claimed that since biodiversity was threatened by "the daily conversion of natural systems to human-managed systems,"[282] an observatory network was needed to better understand, and eventually manage the environment. In particular, the network was supposed to capture that "human dimensions of biodiversity," and "the interactions between biodiversity and human social, cultural and economic dynamics."[283]

---

[280] Ibid. p. 6

[281] "Report of the Second Workshop on the Biodiversity Observation Network," (National Center for Ecological Analysis and Synthesis, Santa Barbara, California: sponsored by the National Science Foundation, 1999). p. 2

[282] Ibid. p. 2

[283] Ibid. p. 3

These two aspects, environments threatened by being incorporated into 'human-managed systems', on the one hand, and, on the other hand, research infrastructures as attempts to provide knowledge for the management of human impacts on the environment are proposed as solutions to the threat of climate change. Indeed, this apparent contradiction lies at the heart of our relationship to nature in the Command and Control Anthropocene. Humans are deploying technology in order to gain an expanded control and knowledge of an environment that has been shaped by anthropogenic forces of various forms in the first place.[284]

The workshop report also stressed the network character of the planned Biodiversity Observation Network. The infrastructure was supposed to "develop in a manner that promotes the 'network' attribute from the beginning, rather than waiting for an array of isolated installations to mature into a network."[285] Furthermore, the importance of technological development to support the digital infrastructure of the network was again emphasized. Five main goals for a digital strategy were laid out: data description, acquisition, archive and retrieval, data access, and communication.[286]

The California Academy of Sciences in San Francisco hosted the next BON workshop in May 1999, chaired by Patrick Kociolek from the Academy and Michael Donoghue from Harvard University. Also in attendance were a large number of scientists from the University of California, Berkeley, just across the San Francisco Bay. The Biosphere 2 Center in Arizona, an Ecology research project that has by now become infamous, also took part in the workshop.[287] The working group envisioned a network of 50 to 75 observatories, mainly

---

[284] See Nils C. Hanwahr, "Marine Animal Satellite Tags," in *Future Remains. A Cabinet of Curiosities for the Anthropocene*, ed. Gregg Mitman, Marco Armiero, and Robert S. Emmett (Chicago: University of Chicago Press, 2018).

[285] BON, "Report of the Second Workshop on the Biodiversity Observation Network." p. 9

[286] Ibid. p. 12

[287] For a fictional description of the Biosphere project see T.C. Boyle, *The Terranauts* (Bloomsbury Publishing, 2016).

focused on biodiversity monitoring, to be financed over several years or even decades.[288] At this point, the focus of the project was still clearly on biodiversity and would only later shift to the impacts of climate change on the national environment.

The notion of the BON infrastructure as a network was once again articulated. The report from the California Academy of Science meeting recommended that each observatory should be a collaboration between a Biological Collections Institution (BCI), such as a museum or university, and a Biological Field Station (BFS), run by various other institutions, to establish "a new level of interaction across biodiversity disciplines."[289] On the other hand, the report also recommended the establishment of a National Center for the Analysis of Biodiversity (NCAB) to coordinate the network efforts as a hub. Thus, the juxtaposition, and possible conflict, between the network and systems character of the Biological Observatory infrastructure was present even at this early stage of planning.

The last workshop of the year 1999 was hosted by the National Science Foundation in Santa Barbara, CA, and brought scientists together with representatives from the NSF as well as experts form the United States Geological Survey (USGS). The focus on biodiversity was expanded to "environmental consequences" and "biocomplexity". During the summer of 1999, the National Science Foundation eventually took a more top-down approach to their various scientific infrastructure projects. The NSF proposed the National Ecological Observatory Network (NEON) "initially as a program to enhance infrastructure for all of field biology."[290] The plan to develop NEON

---

[288] BON, "Report of the Third Workshop on the Biodiversity Observation Network," (California Academy of Sciences, San Francisco, California: sponsored by the National Science Foundation, 1999). p. 25

[289] Ibid. p. 1

[290] "Report of the Fourth Workshop on the Biodiversity Observation Network," (National Center for Ecological Analysis and Synthesis, Santa Barbara, California: sponsored by the National Science Foundation, 1999). p. 2

consortia prior to the establishment of BON as an infrastructure for biodiversity in particular was rejected by the participating scientists, who recommended that "the BON sites be established prior to the development of NEON." However, the NSF's drive for a NEON infrastructure would have enabled plans for a biodiversity-focused BON to build on the "centralized technology" of the overarching ecology infrastructure. The workshop envisioned that BON research sites were to be led by individual Principal Investigators (PI) and would constitute the biodiversity expertise of the National Ecological Observatory Network. [291]

Nevertheless, a big item of discussion in Santa Barbara was an issue of networks versus systems, individuals versus community. It was debated who was to decide about the details of the infrastructure's design: the funder and system builder NSF, or the individually networked and local expert researchers? There was also a potential conflict between individual research agendas, the larger research community, and the networking required to breathe life into a scientific infrastructure, as we have seen in the OOI case study. "During these discussions, a 'top-down' (= NSF-imposed methods) versus 'bottom-up' (= investigator-driven) design was a central issue."[292]

The "NSF-imposed" design was rejected by the scientists for three main reasons: "the specialized knowledge required to devise the best sampling protocols [...]; the enormous heterogeneity of environmental conditions [...]; and the modest budget envisioned," the meeting report concluded.[293] The danger of little top-down control, however, was that the divergent networked efforts would prove to be incompatible, the resulting data would be incompatible with any overarching effort to assemble a nation-wide virtual observatory.

---

[291] Ibid. p. 2

[292] Ibid. p. 6

[293] Ibid. p. 6f.

This issue was to be addressed, the group recommended, "primarily by giving incentives to Principal Investigators."[294] The PIs were to exchange best practices in terms of research and sampling protocols as well as data formats, while the NSF was supposed to review the standards and practices set by the researchers in the field. At this point, most of the participants were either biologists or government administrators, with the only digital infrastructure expert being David Stockwell from the San Diego Super Computing Center.[295] We will see how the early ideas about simple best practices and a networked bottom-up investigator-driven approach to data standards should be transformed much more systematically once more computer scientists became involved in the planning of NEON as a digital infrastructure.

4.3.2 National Ecological Observatory Network Planning Workshops

With the involvement of the National Science Foundation and the installment of NEON as an overarching national infrastructure project, the scope of the infrastructure also broadened to include "global change and anthropogenic influences."[296] NEON was now supposed to "constitute a distributed network of replicated geographical habitats to serve as a platform for many areas of environmental biology from evolution and systematics to landscape- and continental-scape ecology."[297] However, nearly all of the terms in this statement remained contested: How was a distributed network supposed to be designed and controlled? How were geographical habitats to be "replicated", and who was going to select them? Who was going to provide the platform? And who was going to manage the resources of the infrastructure and choose actual

[294] Ibid. p. 7

[295] Ibid. p. 9

[296] NEON, "Report on First Workshop on the National Ecological Observatory Network (Neon)," (Archbold Biological Station, Lake Placid, Florida: Sponsored by the National Science Foundation, 2000). p. 4

[297] Ibid. p. 4

research themes? All of these questions were going to occupy the NEON community for the years to come and were never fully resolved.

The NSF workshop in January 2000 discussed, for the first time, many of the challenges that were going to haunt NEON to this day. The executive summary of the workshop report stated that the NEON infrastructure was going to face "many 'sociological' and institutional challenges [...], including accommodating and encouraging participation of researchers outside the NEON system, links among academic and non-academic institutions, data availability and professional reward systems."[298] Furthermore, the workshop also stated the initially envisioned timeframe of the project, saying that "NEON sites should be designed to last for at least 30 years."[299] The new infrastructure was also differentiated from predecessors such as the Long Term Ecological Research Network (LTER) since NEON was conceived to serve "a much broader range of disciplines in environmental biology, rather than just the ecological community as is the case for LTER."[300]

The figure below was contained in the first NEON workshop report from January 2000 and exemplifies the ongoing discussion about the balance between the network, or web, and the systems character of the envisioned NEON infrastructure. Model A in the figure represents "autonomous NEON observatories with coordination through a central administrative center (original plan)", while model B denotes a "fully integrated, distributed network of NEON observatories."[301]

[298] Ibid. p. 5

[299] Ibid. p. 6

[300] Ibid. p. 8

[301] Ibid. p. 10

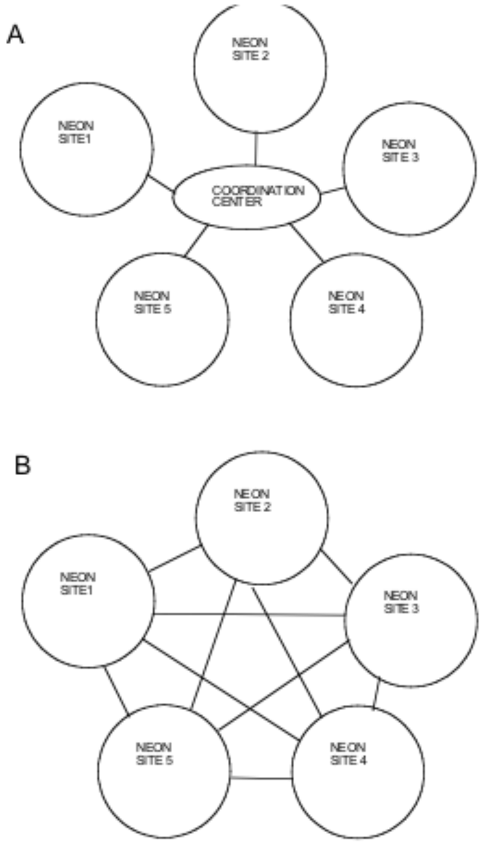


Figure 10 Centralized versus Decentralized Models of Observatory Organization[302]

The participants of the workshop felt that while Model A promised to ensure efficiency and a broadly prescribed spectrum of research themes it could stall new ideas emerging from the scientific community and might exacerbate "resource conflicts between NEON and external researchers."[303] There appears to have been a lot of disagreement on the general question of network versus system character of the envisioned infrastructure and the report concluded starkly: "It was clear that consensus was not reached on this overall issue and it will need to be addressed in future discussions."[304] In fact, this issue has never been fully resolved by NEON and could be one of the reasons for its ongoing troubles.

[302] Ibid. p. 10

[303] Ibid. p. 10

[304] Ibid. p. 11

The question of centralization was also a contested topic of discussion with respect to the technologies deployed by NEON. It seemed clear that a certain level of standardization and centralization was necessary, yet the long-term infrastructure project also had to allow for technological innovations to be integrated into its system, which posed a major challenge in planning: "given the pace of technology development, there is need to plan for infrastructure with sufficient flexibility and opportunity for innovation in the 30-year time span of NEON."[305]

The NEON1 report was also the first to specifically raise "sociological and institutional challenges associated with NEON."[306] Issues were thought to arise in the area of inclusiveness, when "grants may go exclusively to groups that are already organized into networks, rather than groups that might be deserving but not historically linked."[307] Over the course of the rollout of the infrastructure we are going to see that the prevalence of existing networks of individuals and institutions is a double-edged sword. While it can provide a solid established foundation of the research infrastructure, it can also potentially exclude outsiders and tilt the infrastructures network character toward a systems character.

Another issue was the "standardization and availability of data" within a network infrastructure, which will "present a major administrative burden that cannot be expected to fall on individual P.I.s."[308] Since standards and procedures for large-scale data gathering and sharing required some level of centralized effort as well as substantial resources, it presented a dilemma for an infrastructure that individual researchers would prefer to be as decentralized and network-like as possible.

305 Ibid. p. 12f.

306 Ibid. p. 13

307 Ibid. p. 14

308 Ibid. p. 15

In March of the year 2000, computer scientists and data experts finally got involved in the NEON project during the following NEON planning workshop at the San Diego Supercomputer Center in La Jolla, California. The workshop report stressed the innovative differences between NEON and its predecessor LTER, focusing on the level of research coordination in particular: "LTER sites are evaluated based on the productivity of their investigators in publishing research papers, [...] funding for a NEON observatory should depend on the excellence of its service to the community and the quality of science enabled by that service."[309]

The planning workshop also laid out a list of technological components to be included in a standard NEON observatory site, including network components, laboratory equipment, and field equipment e.g. for hydrological and meteorological measurements.[310] The expected annual budget of operating a NEON observatory was also envisioned to be at least 1,25 million US$.[311] Considering the planned run-time of the infrastructure of at least 30 years and the aim of hundreds of observation sites, one can begin to grasp the huge financial scale of this major environmental research infrastructure project. However, the working group could not yet agree on so-called "core measurements" that "should address the needs common to the broadest possible user community."[312] Thus, while the scientists identified a number of technologies and research fields to be included, there was no consensus on what data should actually be captured, recorded, and distributed.

---

309 "Report to the National Science Foundation from the Second Workshop on the Development of a National Ecological Observatory Network (Neon)," (San Diego Supercomputer Center, La Jolla, California2000). p. 6

310 Ibid. p. 7f.

311 Ibid. p. 11

312 Ibid. p. 10

Discussing the central importance of informatics, the NEON planners concluded that "balancing the autonomy of the individual observatories with network-level standardization will be a challenging task, but must be addressed at the inception of the NEON network so that individual observatories implement compatible infrastructures."[313] As specific challenges the report listed data acquisition, quality management, storage and archiving, dissemination and access, integration and aggregation as well as analysis, synthesis, and modeling. It appears to be only in early 2000 that the enormous scale of the data infrastructure challenges became clear, as the success of the infrastructure as a whole and its ambition to provide a novel resource of insight for the entire scientific community depended on a solid digital infrastructure.

Since societal relevance was also one of the National Science Foundation's funding criteria, the NEON planners began to think about how NEON data and outreach could benefit laypeople and citizens with its novel approach to a nation-wide observatory infrastructure. However, at this point, concepts remained vague as the report stressed "an unprecedented opportunity to include underrepresented communities and citizen scientists via outreach activities."[314] This appeared to be a concession to the NSF's funding requirements by the researchers as it is left entirely open how the envisioned outreach would take place and who exactly the "underrepresented" communities were. Nevertheless, at this point of the planning, informatics and governance were identified as the largest challenges in conceiving an infrastructure such as NEON, and thus, the following planning workshop was supped to deal with further details of the informatics necessary to realize such a nation-wide project.

---

[313] Ibid. p. 13

[314] Ibid. p. 17

In June 2000, NASA's Goddard Space Flight Center saw the next gathering of biologists and computer scientists, this time hosted by the United States Geological Survey and the National Science Foundation. The workshop report coined the term for an allegedly novel "interdisciplinary field of study as Biodiversity and Ecosystem Informatics (BDEI)."[315] The working group acknowledged the enormous challenges awaiting them in building a national infrastructure for an observatory such as NEON, while at the same time stressing the need for significant investment in Biodiversity and Ecosystem Informatics by the public.

Parallels to investments in climate science and climate research infrastructures were drawn explicitly, stating that "just as we are developing a capacity to predict long-term climate events, we would now like to predict public health and ecological outcomes far into the future. Unfortunately, we currently lack the technologies to do this."[316] The report was thus drawing an analogy between climate science at a global scale and a National Ecological Observatory Network providing crucial knowledge on public matters such as environmental health and more local concerns such as ecosystem change and its impact on communities in the United States.

Notably, the scholar Geoffrey Bowker, who conceived the crucial term "infrastructure inversion", was also part of the workshop group. Infrastructure inversion was going to be an important part of the work of BDEI, just as it had been for climate scientists, as "historical information serves prominently in the work of biodiversity and ecosystem scientists. [...] These historical sources are often as pertinent as contemporary data."[317] Thus, if NEON was supposed to be a long-term observatory, it would not only have to look to the future but unlock access to historical environmental data as well. This historical data

[315] Dave Maier et al., "Research Directions in Biodiversity and Ecosystem Informatics. Report of an Nsf, Usgs, Nasa Workshop on Biodiversity and Ecosystem Informatics Held at Nasa Goddard Space Fight Center, June 22-23, 2000," (2000). p.1

[316] Ibid. p. ii

[317] Ibid. p. iii

would have to be digitized in a form that could be combined with the kind of data that NEON observatories were going to generate in real-time.

Nearly two years later, in June 2002, NEON planners met again to discuss standards for the digital infrastructure to be constructed. The report pointedly posed the question: "What will we be able to do with NEON that we cannot do now?"[318] Again, the main societal concerns of public health and the impact of climate change on national and local ecosystems were stressed. Surveillance of the environment and mitigation of the impact of climate and environmental change were explicitly named as major goals of the NEON infrastructure. The aims of NEON were, for the first time in the sources, explicitly linked to addressing the and mitigating the risks created by the impacts of climate change on the United States environment:

> Abrupt and catastrophic changes in ecological systems induce economic and cultural and dislocations of national significance that may on the one hand arise from natural process or on the other hand from intentional actions designed to damage regional and national ecosystems and human well-being. NEON will provide a base of understanding and to contribute to the mitigation of these damages.[319]

It is, however, surprising that a central capability of an environmental monitoring observatory that is explicitly long-term oriented was stated to be a better understanding of "abrupt and catastrophic changes." We have already seen in the case study on the Ocean Observatories Initiative that the use of a long-term observatory can be hard to communicate and justify politically, while an enhanced surveillance of more tangible risk events such as the abrupt eruption of a volcano or a catastrophic tidal wave are more fathomable than slow long-term transformations. Narratives of sudden and spectacular impacts

---

[318] NEON, "Report to the National Science Foundation from the Fourth Workshop on the Development of a National Ecological Observatory Network (Neon): Standard Measurements and Infrastructure Needs," (The Millennium Hotel, Boulder, Colorado2002). p. 1, non-paginated

[319] Ibid. p. 2

representing environmental change tend to render much more striking narratives than gradual change or nearly imperceptible slow violence.[320]

The June 2002 workshop in Boulder, Colorado, was the first major NEON project meeting after the terror attacks of September 11th, 2001. Shortly afterward, between September 18th and October 9th 2001, several US Senators received letters containing spores of anthrax, killing 5 people.[321] What could the mentioned intentional disruptions be if not 'environmental terrorism'? And what was meant by the economic and cultural dislocations? Rather than talking about ecosystems and climate change, the NEON working group had rapidly adopted the vocabulary of national security and terrorism after 9/11 in the United States. That is remarkable, given that we are talking about biologists and computer scientists discussing the planning and justification of an environmental research infrastructure.

Later in the same workshop report, the justification for NEON's use sounded more familiar again, focusing on biodiversity and climate change impacts in the long run: "Monitoring distributions and abundances across the Tree of Life, including microbes, plants and metazoans has never been accomplished, yet understanding how natural and anthropogenic environmental change affects organisms is a key goal in understanding ecosystem function through time."[322]

The following NEON planning workshop addressed the role of non-academic research institutions such as museums of natural history within the ecology and environmental monitoring endeavor. Participants congregated in mid-June 2002 at the Field Museum of Natural History in Chicago to discuss how existing biodiversity collections could be fruitfully combined with the systematic ecological data to be produced by the NEON observatories. Once

---

[320] See Rob Nixon, *Slow Violence and the Environmentalism of the Poor* (Harvard University Press, 2011).

[321] "Responding to Anthrax Attacks," *The New York Times*, 16 October 2001 2001.

[322] NEON, "Report to the National Science Foundation from the Fourth Workshop on the Development of a National Ecological Observatory Network (Neon): Standard Measurements and Infrastructure Needs." p. 3

again, the planners aimed to learn from the LTER program, discussing a presentation by Darlene Judd on the co-operation between the Oregon State Anthropod Collection and a nearby LTER observatory site.[323]

The group reiterated the state of funding plans, which actually comprised three different funding lines from the National Science Foundation to cover costs for research observatory instrumentation, maintenance and operations of the observatory sites, and funding for research as such.[324] However, major planning issues still remained unresolved such as how to set up a central coordination center for the infrastructure to focus bioinformatics efforts in particular. In addition, representatives from museums and natural history collections pointed out that, as of yet, no funding plan for the "retrospective collection data capture" existed and urged the NSF to take the need for this kind of historical data gathering through infrastructure inversion into account.[325]

In September 2002, the NEON community met once again at the National Center for Ecological Analysis and Synthesis at UC Santa Barbara to focus on the planning of NEON's information management. Yet again, the unresolved conflict between the need for standardized data formats and central coordination versus the independence and local knowledge of observatory site research teams was controversially discussed. The central importance of information management was stressed and its part of the annual funding budget estimated at up to 40% of the total budget of each observatory site as well as up to 75% of the budget of the infrastructure central coordination unit as such. The report summary stated that "the program will not be a network without comprehensive information management," proposing a "NEON

[323] "Final Report of the Neon-V: Cripton Workshop. Collections, Research, Inventories, and People for Taxonomic Opportunities in Neon," (Field Museum of Natural History, Chicago2002). p. 2f., not paginated

[324] Ibid. p. 5

[325] Ibid. p. 12

Coordinating Unit" as the focal of NEON information gathering and management.[326]

In particular, the participants stressed "that the NEON Coordinating Unit should be created prior to, or at least no later than, the first NEON observatories, and that carefully crafting responsibilities, authority, and accountability of the coordinating unit was among the most critical tasks."[327] The major tasks of the Coordinating Unit were outlined as administration, technology acquisition as well as advisory and methods training for observatory site personnel, public relations and outreach, and general network coordination of the NEON community's collaborative research efforts. With respect to the NEON data, the Coordinating Unit was to be responsible for data archiving and data standards and metadata curation. In addition, the Unit was supposed to run and provide a national "portal for accessing NEON resources."[328]

Considering such centralized coordinating power, several workshop participants questioned how the project could "ensure that the NEON Coordinating Unit is accountable to the observatories and effectively uses its resources for the benefit of the whole network."[329] We will see that this accountability was precisely what caused a scandal for NEON after a federal auditor of the infrastructure project turned whistleblower pointed out the large-scale misuse of funds by members of the NEON planning office. To achieve the massive coordination effort in information management that NEON planners undertook, carefully worked out metadata standards for

[326] "Report to the National Science Foundation from the Sixth Workshop on the Development of a National Ecological Observatory Network (Neon): Information Management," (National Center for Ecological Analysis and Synthesis, University of California Santa Barbara, Santa Barbara, CA2002). p. i

[327] Ibid. p. i

[328] Ibid. p. 2

[329] Ibid. p. 3

highly heterogeneous datasets were absolutely crucial to turn the infrastructure into a functioning network.[330]

4.3.3 NEON Coordination and Implementation Planning

After more than two years of planning workshops and research community consultations, the implementation planning for NEON eventually gained traction in late 2003 during two implementation conferences, at the National Museum of Natural History and at the American Institute of Biological Sciences (AIBS), taking strides toward realizing the vision outlined during the numerous workshops and meetings described above.

In August 2002, the American Institute of Biological Sciences (AIBS) had actually launched its own resource program supposed to support an infrastructure project such as NEON under the name Infrastructure for Biology at Regional to Continental Scales (IBRICS).[331] The implementation conferences were chiefly concerned with creating sensible governance structures for the NEON infrastructure, both during its project rollout as well as during routine operations. The whitepaper from 2004 was also the first time the idea of creating "NEON Inc." as a central coordination and governance body was spelled out in detail. The chart shown in the figure below is actually a fleshed-out version of earlier diagrams we have discussed. One can see immediately that the planners opted for a clearly centralized and top-down governance structure with the National Science Foundation, and an Inter-Agency Working Group, providing oversight for the cluster of institutions that was going to become NEON Inc.

---

330 Ibid. p. 7

331 American Institute of Biological Sciences, "Plan for Developing and Governing the National Ecological Observatory Network (Neon). Report from the Neon Coordination and Implementation Conference," in *IBRCS White Paper*, ed. AIBS (Washington, DC2004).

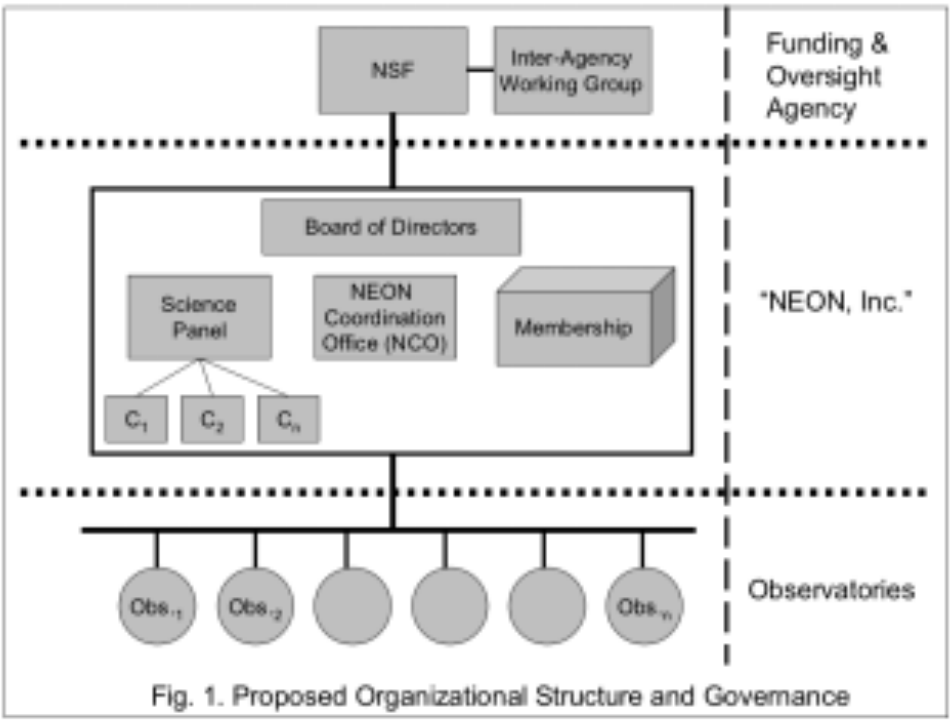


Figure 11 NEON Proposed Organizational Structure and Governance [332]

NEON Inc. was supposed to achieve two main goals: first, oversee and coordinate the rollout of the technological infrastructure needed to realize NEON, and second, provide a framework of research questions to be addressed. However, the funding for individual research projects was not to be handed out by NEON Inc., which would only be funded by the NSF-based Major Research Equipment and Facilities Construction (MREFC) (the same funding program as the Ocean Observatories Initiative) during infrastructure construction and from some other, as yet to be clarified, source during operational routine.[333]

The NEON planners nevertheless wanted to address the premonitions of individual scientists worried about too much central control by including a box labeled "Membership" within the NEON Inc. diagram. The whitepaper defined this kind of NEON membership very broadly, stating that

[332] Ibid. p. 5

[333] Ibid. p. 8

“membership should be inclusive, drawing in organizations and institutions that may or may not be directly affiliated with one or more observatories.”[334]

Such a broad definition of how NEON ‘members’ were supposed to be involved was risky, since it left entirely undefined who was actually going to be regarded as a stakeholder of the infrastructure and how this ‘membership’ was going to be represented within NEON Inc. Thus, the question of system versus network characteristics of the NEON infrastructure remained open while it started to tilt strongly toward a systematic top-down approach that was also mirrored in the conceptualization of funding structures for both infrastructure and individual ecological observatories.

Figure 12 below shows another flowchart from the 2004 whitepaper on NEON coordination and implementation depicting the ‘NEON Funding Flow’. It clearly outlines the separation of funding flows for infrastructure building versus research projects. In practice, however, the separation was unlikely to be as neat as the diagram of funding flows suggests. Generally, the rollout of the NEON ‘core’ infrastructure was going to be funded under the NSF’s Major Research Equipment and Facilities Construction (MREFC) program. On the other hand, NEON Inc. was also going to be eligible for research funding, as were the individual observatories.

It is important to note here that NEON sites were not identical to NEON observatories since one observatory may actually comprise several regional sites and involve several regional institutions involved in the observatory. Thus, research projects were supposed to be funded at the level of individual observatories, individual sites, and regional project co-operations. Furthermore, NEON project carried out by NEON member institutions that were not themselves involved in running an observatory site were envisioned to be able to apply for NSF funding for NEON-related research.

[334] Ibid. p. 7

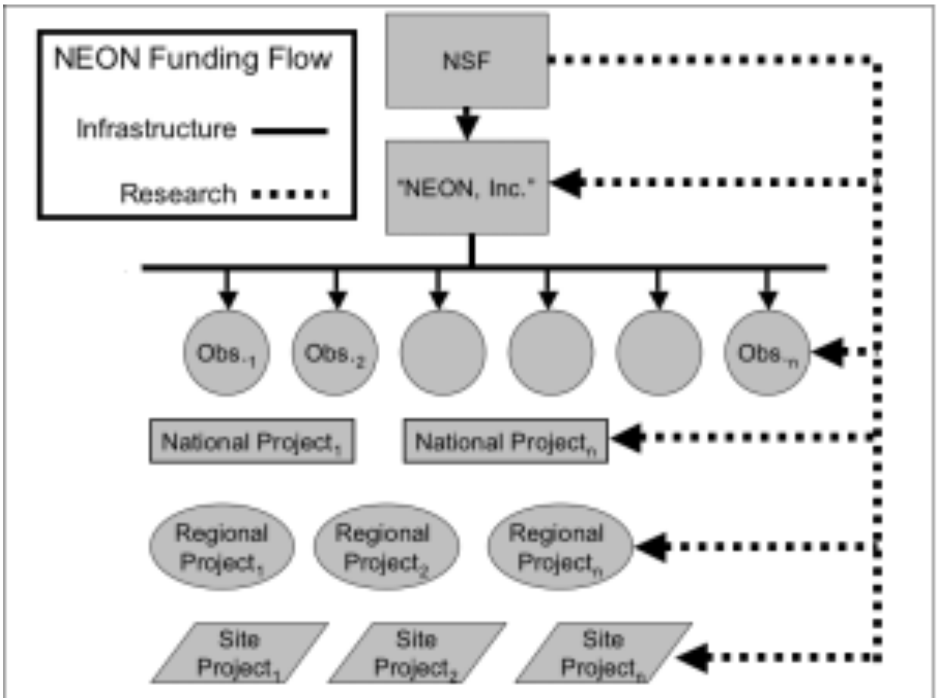


Fig. 2. Proposed flows of infrastructure and research funds within NEON. Infrastructure funding refers to both major research equipment and annual maintenance and operations.

Figure 12 NEON Funding Flows[335]

Eventually, the NEON implementation workshops identified three key next steps to be taken in 2004 to begin realizing the ambitious new infrastructure project. First, task forces were to be established to flesh out requirements for infrastructure governance, administration, education and outreach, and, most importantly "IT, Data Management, and Standardization."[336] Second, the planners needed to "identify and prioritize important national and regional scientific questions for NEON foci that are regionally specific and relate to broad scientific themes."[337]

[335] Ibid. p. 8

[336] Ibid. p. 15

[337] Ibid. p. 15

Once the overarching research questions were identified by an envisioned task force, these research goals were to be coordinated with the task force on digital infrastructure. Yet, NEON was faced with a hen-and-egg problem at this point, since it was unclear who was supposed to define the broad research questions of the network, while digital infrastructure experts who needed to specify requirements for a fundamental IT and data management framework had to construct an infrastructure that would accommodate the broadest possible range of biological and ecological research projects. To come up with those broad-ranging and yet specifically defined research questions for NEON, the National Research Council (NRC) had formed a Committee on the National Ecological Observatory Network. The group chaired by the University of Minnesota's G. David Tilman published its extensive report in 2003 under the title "Neon: Addressing the Nation's Environmental Challenges."[338]

The committee contained only two non-university members: Carol Fialkowski from the Field Museum of Natural History in Chicago, and Dorothy Gibb from Horne Engineering Services in Fairfax, Virginia. Nowhere in the entire report was the term "Anthropocene" mentioned or any work by Paul Crutzen cited, yet NEON's statement of purpose invoked the idea of the research infrastructure as a tool for measuring the impacts on humans caused by human impacts on the environment and its ecosystems:

> Human technology, land use, and resource acquisition have accelerated the pace of regional and global environmental change to the extent that human actions are now a major force in the stability and functioning of most terrestrial, aquatic, and marine ecosystems. [...] The nation needs and deserves a scientific understanding of its natural and managed ecosystems that is sufficient to assess how alternative human actions might impact the functioning of ecosystems and the services that they provide the nation and to identify science-based solutions to ecological problems.[339]

---

[338] National Research Council, *Neon: Addressing the Nation's Environmental Challenges* (Washington D.C.: The National Academies Press, 2004).

[339] Ibid. p. 1

While this passage acknowledged the major human influence on environments that, in turn, impact human society, the NEON Committee did not appear to share a full understanding of the Anthropocene. Distinguishing "natural and managed ecosystems" would be incoherent with the Anthropocene concept since the dichotomy of 'natural' versus 'managed' has collapsed on the Anthropocene planet. Nevertheless, the "major environmental challenges" proposed by the Committee report were all distinctly Anthropocene issues:

1. "biodiversity, species composition, and ecosystem functioning"
2. "ecological aspects of biogeochemical cycle"
3. "ecological implications of climate change"
4. "ecology and evolution of infectious diseases"
5. "invasive species"
6. "land use and habitat alteration"[340]

All of these key challenges had a national focus and a pronounced emphasis on security and public health. Yet, the main focus was on security and environmental management since the NEON infrastructure was supposed to enable the United States to "mitigat[e] large-scale adverse impacts before they become severe threats to society."[341] Besides the strong focus on environmental national security, the Committee regarded NEON as a tool for policy-making by "helping society to choose policies that provide the greatest long-term net benefits."[342]

Furthermore, the Committee pointed out that previous planning reports on NEON "reflect the different foci of the different types of scientists at the workshops," and that a synthesis of various NEON concepts was still needed.[343] Specifically, with respect to the infrastructures network structure, an alternative was suggested, namely that the "network should be designed in such

[340] Ibid. p. 5

[341] Ibid. p. 7

[342] Ibid.

[343] Ibid. p. 78f.

a way that even the first elements are built to fulfill the national networking role."[344] Concerning the controversial question of centralized versus distributed governance of the infrastructure, the Committee clearly sided with the NSF's idea of a top-down NEON coordination unit.

"Both centralized and decentralized approaches were proposed in the workshops, possibly because some workshop groups were not aware of the management requirements for MREFC. The proposed operational governance structure may result in a large bureaucracy that would hinder NEON operations, rather than help."[345] Thus, the decision to organize the NEON network around a centralized coordination unit crystalized due to the requirements of the NSF's own funding programs that would not allow for a distributed governance structure with community-shared responsibility.

Looking ahead to the next steps in planning NEON, the Committee concluded its report by recommending major co-operations with institutions in bioinformatics as well as other federal agencies. In particular, informatics centers such as the National Center for Ecological Analysis and Synthesis, the National Biological Information Infrastructure, and GenBank were recommended as models to emulate in constructing the National Ecological Observatory Network.[346]

For the first time ever for a project in biology and ecology, in the fiscal year of 2003, the "NSF requested Congress allocate $12 million to the Major Research Equipment and Facilities Construction (MREFC) account to initiate construction of the first two NEON observatories."[347] The following year, another $12 million dollars for the observatories as well as $6 million for NEON-related research activities were requested by the NSF. Yet, as of 2004, the United States Congress had denied the National Science Foundation's

---

[344] Ibid. p. 82

[345] Ibid. p. 83f.

[346] Ibid. p. 90

[347] Ibid. p. 19

infrastructure budget proposals. The next step in the planning process would be to put some meat on the bones and get experts together to flesh out the six core research questions and start looking for observatory sites for NEON.

4.3.4 NEON Specialist Workshops on Scientific Questions

Starting in 2004, NEON began to convene several specialist workshops to elaborate the six core research questions the infrastructure was supposed to address. The first specialist workshop titled "Biodiversity, Species Composition, and Ecosystem Functioning" took place in late July 2004 in Carmel Valley, close to the University of California Riverside. Conveners were Brent Mishler and Craig Moritz from UC Berkeley and the American Institute of Biological Sciences. The workshop group actually came up with another list of research questions, differentiating between "initial conditions" and "predicting the direction and rate of change."[348]

Figure 13 below, titled "Interacting Drivers of Biological Change" from the biodiversity workshop report, shows the conditions of an ecosystem at the center (biodiversity, species composition, ecosystem functioning) and the dynamic drivers of transformation as various forces on the initial conditions (land use and habitat alteration, biogeochemical cycles, invasive species, climate change, and evolution of infectious diseases). The core idea was to ultimately be able to assess the impacts of drivers of change not just on ecosystems and biodiversity but, first and foremost, on the ecosystem services provided to humans in the United States. The underlying goal was, thus, to predict ecological dynamics in order to manage ecosystems as providers of ecosystem services. The working group concluded: "NEON will provide the resources to make realistic predictions about directions and rates of change, and determine responses to management decisions. [...] From there,

348 NEON, "Biodiversity, Species Composition, and Ecosystem Functioning," (Carmel Valley, CA2004). p. 4

management decisions can be made or, when the process is beyond human control, understood and accepted."[349]

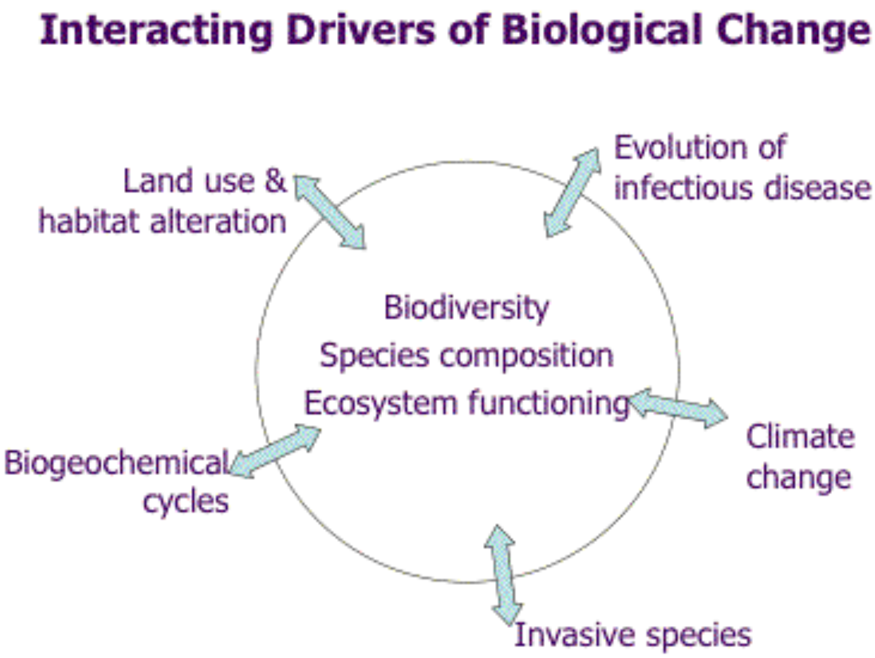


Figure 13 Interacting Drivers of Biological Change[350]

This was a striking conclusion to a specialist workshop on biodiversity, since the complex and intertwined impact of human activity appeared to figure in either as a rational management decision, or as something that was ultimately beyond human control and should be accepted once it was understood. Nonetheless, the specialists conceived of NEON as a tool to make decisions on environmental management and ecosystem services, since it will "provide the resources to make realistic predictions about directions and rates of change, and determine responses to management decisions."[351]

---

349 Ibid. p. 19

350 Ibid. p. 2

351 Ibid. p. 19

Another NEON science workshop took place in August 2004 in Tucson, Arizona, this time focusing on "Ecological Impacts of Climate Change."[352] Main conveners at this workshop were Julio Betancourt from the United States Geological Survey (USGS) and Pat Mulholland from Oak Ridge National Laboratory. The workshop group discussing the potential role of NEON in monitoring the impacts of climate change on ecosystems in the United States was well aware of the major global research and infrastructures already in place and operated by their colleagues in meteorology and climate sciences. The report explicitly acknowledged the ambition to match the scale of climate science: "If ecologists are to be successful in distinguishing competing and interacting causes of large-scale ecological changes and associated feedbacks to the atmosphere and hydrosphere, they will need to match the spatial and temporal scales of analysis employed routinely by climatologists."[353]

The participants of the climate change workshop once again took up the familiar discussion of a centralized versus a distributed infrastructure network. They also debated whether the infrastructure should be used to conduct experiments within the network and gather data in a targeted way, or whether the network should be more long-term oriented and observational. Participating scientists saw "tradeoffs between (a) experimental approaches and purely observational and synoptic approaches, and (b) investing in a few, highly instrumented sites and investing in a more distributed network."[354]

The question of impacts of climate change clearly demanded a more distributed infrastructure network than the study of biodiversity. The scale of climate change, even when observations were limited to the continental United States, was beyond the scale of a small number of specialized ecological observatories. The effort would have to be broader and draw on more data as

[352] American Institute of Biological Sciences, "Ecological Impacts of Climate Change: Report from a Neon Science Workshop," (Washington D.C.: AIBS, 2004).

[353] Ibid. not paginated, Executive Summary

[354] Ibid. p. 2

well as historical data, like climate scientists have done. The workshop report thus emphasized that "many of the workshop participants were skeptical about the efficiency of the hub-and-spoke design of early NEON discussions, with the activities focused at a few central observatories that control the distributed networks."[355] Clearly, climate change impacts required a scaling up of ecology to a continental level to study its impacts.

Especially the issue of timescales proved to be a challenge for the scientists discussing climate change impacts. While NEON was envisioned to be constructed and funded for a timescale of 30 years, climate change impacts were going to take effect on a timescale of at least, but likely longer than, 30 years. The NEON timescale of "observations will not be enough time to capture many of the large-scale ecological responses to climate change," the report concluded.[356] Thus, like climate science according to Paul Edwards, climate change impact researchers were going to have to rely on historical ecology to construct timescales of data long enough for their purposes – data would have to be 'made' to address this particular challenge. That also meant NEON was going to have to incorporate and cooperate with a large number of institutions that could provide the historical records required, an effort the workshop report compared to the Historical Climatology Network.[357]

Besides biodiversity and the impact of climate change on national ecosystems and environments, NEON planners also convened a special workshop on modelling as a component of the observatory infrastructure. In July 2005, the community met at the Marine Biological Laboratory in Woods, Hole, MA, the famous oceanographic institution on the United States East Coast. Although NEON would span the entire United States continental territory, its results would nevertheless have to be extrapolated from the observation sites by data modelling. The workshop report concluded that "models will be required by

---

[355] Ibid. p. 2

[356] Ibid. p. 5

[357] Ibid. p. 7

NEON for spatial projections within regions and across the nation as a whole, and for high-level synthesis within and across the major questions that organize NEON research."[358]

Thus, modelling efforts on top of data collection and curation infrastructure were regarded as essential for the entire project to realize its ambitious goals. The workshop participants at Woods Hole recommended the implementation of a specialized NEON Forecasting Center "to make systematic predictions at the content-scale" and "integrate forecasting capability across the Observatories."[359] Yet, once again, the planning workshop struggled to reconcile the needs of a systematically controlled and yet distributed network structure and ended up suggesting "a flexible design that can facilitate the evolution corresponding to network needs," on the one hand, and, on the other hand, "a separate high-level component for planning and oversight of models."[360]

Generally, it was clear that the 'making' of continental data was going to crucially depend on modelling, since modelling was necessary to actually place "observations into the spatio-temporal context of the NEON data archive."[361] Furthermore, without modelling, the presentation of all the data generated by the observatory network in a publicly accessible virtual observatory would be impossible. This is why modelling and virtual observatories were also crucial to NEON's outreach and public education goals.

---

358 NEON, "A Workshop on Modelling in Neon," (Woods Hole, MA: The Ecosystems Center, Marine Biological Laboratory, 2005). p. 1

359 Ibid. p. 2

360 Ibid. p. 2

361 Ibid. p. 5

In particular, the participating scientists took issue with the way NEON planning documents used the term 'forecast', since this term "carries a lot of baggage and a number of people at the workshop thought it should be replaced by 'prediction'."[362] One participant explained the uneasiness with the term in a longer statement:

> I believe that its use [the term "forecast", NH] in the ecological context is not only inaccurate but also misleading. Even weather/climate scientists use different terms for long term estimates (e.g. scenarios), and those studies are typically linked to or use as input to ecological studies. Perhaps most significantly, the term forecast implies an inevitability, whereas long-term future conditions depend strongly on highly uncertain future human actions – a message that should be kept clear. Using this term for political/strategic reasons I think will end up being negative.[363]

The concerns expressed by this scientist include many of the dilemmas that NEON as an infrastructure and a source of public information was facing and is still faced with today. 'Forecast' will, for most people, be closely associated with 'weather forecast', and, although we all know that weather forecasts are not always reliable, this association may nevertheless conjure up an expectation of inevitability, especially intangibility by any human action. As the scientist emphasized, NEON's modelling of data is more closely similar to climate modelling and predictions made in the climate sciences. However, since NEON's scope reaches all the way down to small ecosystems impacted by large-scale change, the margin of error and susceptibility to various courses of human interference and action are even larger than is the case for climate prediction models.

[362] Ibid. p. 8

[363] Ibid. p. 8

4.4 Knowledge Sharing and Maintenance

4.4.1 Integrated Science and Education Plan

After funding had finally been approved by Congress and the National Science Foundation in 2006, the stage was set for the construction of NEON observatories to begin. A group of scientists and stakeholders named the NEON Design Consortium had assembled the insights form the numerous previous NEON workshops and science community consultations into an Integrated Science and Education Plan (ISEP) published in October 2006.[364]

The Integrated Science and Education Plan defined the way forward for the National Ecological Observatory Network for the first time in a comprehensive way. Most importantly, the plan defined the locations of envisioned NEON observatories as well as the specific observation technologies to be deployed. For the sake of observatory site placement, the United States territory was divided up into 20 ecologically distinct regions ranging from the Atlantic Northeast to the Pacific Tropical region of the Hawaiian islands (see Figure 14).[365]

The 20 regions were going to be monitored by "core sites" representative of the ecology of the individual region. The dynamic of ecosystem change in relation to the various regions was going to be monitored by so-called "gradient sites", observation systems that span the transitory zones of larger ecosystem areas. This overall network would be supplemented by "sites of opportunity" and "experimental sites" to be set up in locations of special interest for ecological change.[366]

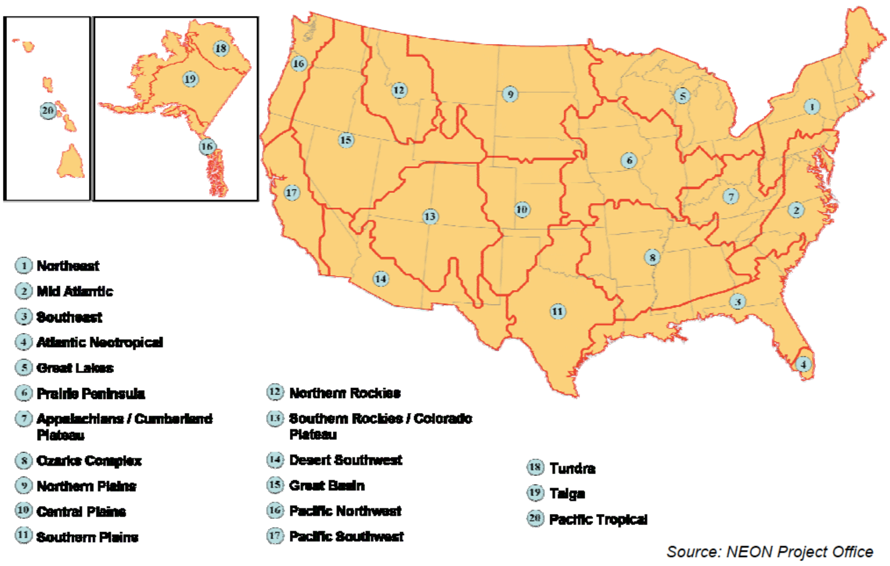
---

364 "Integrated Science and Education Plan for the National Ecological Observatory Network," (2006).

365 Ibid. p. 25

366 Ibid. p. 31

NEON core sites were going to be outfitted with on-site facility headquarters as well as a standard package of observation technology comprising five instrumentation packages. A Fundamental Instrumentation Unit (FIU) provided the core of the observatory in the form of tower carrying various instruments to measure local climate and atmospheric fluxes. A Fundamental Sentinel Unit (FSU) would include assessment sites of hydrology, biodiversity, and soil composition. This would be complemented by instruments comprising a Land Use Package (LUP). A core site would be completed by two mobile assets, a Mobile Relocatable Platform (MRP), an instrumentation tower that could be moved around on trailer as needed by the local investigators, and an Airborne Observation Platform (AOP) providing remote sensing capabilities based on a light aircraft.[367]



[368]

Figure 14 NEON Observation Regions

The Integrated Science and Education Plan also defined the organization and committees comprising the complex NEON operation. The NEON Senior Management Team, in October 2006, was headed by Chaitan Baru from the UC San Diego Supercomputer Center and Bruce P. Hayden from the

---

[367] Ibid.

[368] Ibid. p. 25

American Institute of Biological Sciences. Hayden also headed the crucial National Network Design Committee. The Facilities & Infrastructure Committee's subcommittees were headed by, again, Chaitan Baru and Deborah Estrin from the UCLA Center for Embedded Networked Sensing. The task of the Science & Human Dimensions Committees was mainly to flesh out the scientific core questions to be addressed by the NEON project.

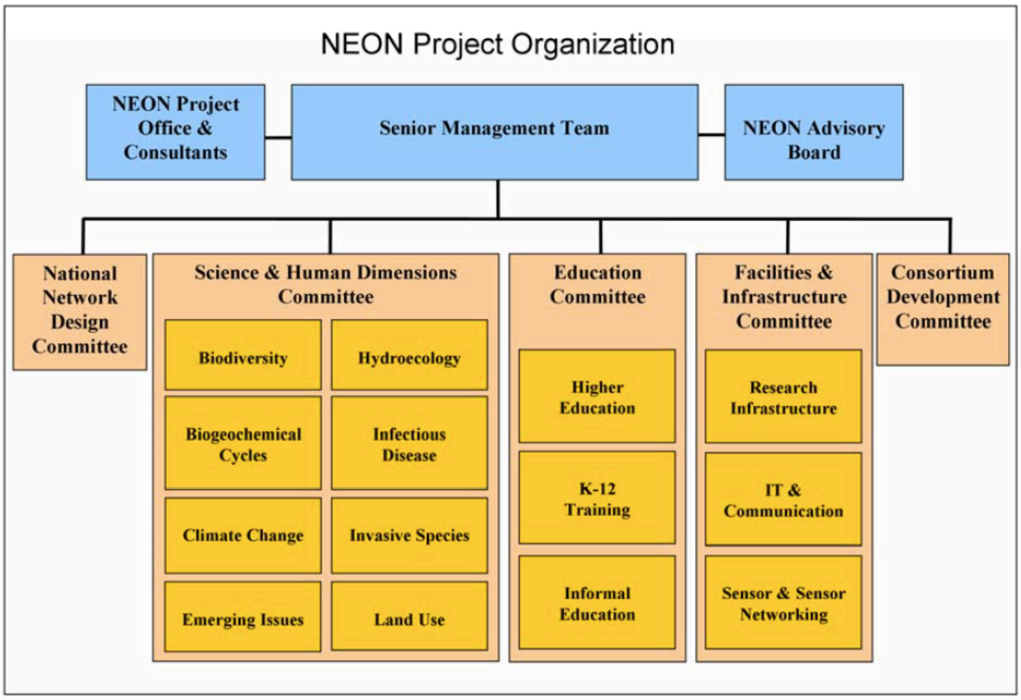


Figure 15 NEON Project Organization[369]

With the organization finally in place, funding secured, and the observation technologies of observatory sites defined, the ecology community was ready to take the next step and decide where exactly the core sites of the observatory were going to be located. NEON had published a Request For Information to the scientific community and gathered a large number of suggestions for suitable sites within the defined 20 ecological regions within the United States. An NSF workshop in Sioux Falls, South Dakota then brought together

[369] Ibid. p. 92

NEON planners once again to distill the gathered suggestions into a definitive list of core sites in March 2007.[370]

After core sites were selected, NEON Inc. started to put out tenders for contractors to set up the observatories and also began to look for principal investigators and observatory directors, positions bound to be coveted by many ecologists in the country. At a Neon Members Meeting in October 2008, NEON Inc. CEO David Schimel presented the state of the project to the scientific community.[371] Central to the public engagement vision was going to be the STEAC, Science, Technology and Education Advisory Committee, chaired by Dr. Chris Field. The former Assistant Director of the National Radio Astronomy Observatory (NRAO) Anthony Beasley, serving as Chief Operating Officer and Project Manager for NEON, also presented his project outline for the years to follow.

4.4.2 NEON Final Design Review and Implementation

Tony Beasley, as Chief Operating Officer (COO), was in charge of overseeing the implementation and construction of NEON and its potential financial risks. A final design review was presented in November 2009, focusing in particular on technology deployment and financial risk management.

Plans for NEON public engagement were also presented, outlining prototypes for a public web portal as well as a special online portal for educators. Strikingly, the major risk was "Target decision-support audience is not defined", stating that "target users have not been defined specifically enough to accurately complete this process."[372]

---

[370] "Neon, Inc. Announces the Selection of the National Core Sites and Research Question Designs," (2007).

[371] Powerpoint presentation given by David Schimel at NEON Inc. Members Meeting in October 2010.

[372] W. Gram, "National Ecological Observatory Netowork. Education and Public Engagement," (NEON, 2009).

The head of the cyberinfrastructure team Robert Tawa presented plans for the digital backbone of the observatory network. Development and construction of cyberinfrastructure and software integration presented major risks to the NEON implementation schedule. In particular, it proved to be difficult to recruit adequately qualified personnel in time to begin construction. Tawa raised staffing as a major project risk, stating that it could happen that "NEON will not be able to hire the requisite personnel in a short time, leading to schedule slippage, excessively high workload for existing personnel, and possible low morale and burnout."[373]

Project manager Beasley identified a number of minor and major risks within all areas of the project and assessed that the majority of risks in NEON construction fell within engineering and cyberinfrastructure challenges, representing a financial risk of more than 50 million US-dollars. Thus, with NEON construction envisioned to be finished by 2015 and due to start the following year, 2010, it was already clear that NEON was not only a research project on a yet unprecedented scale, but also a huge financial risk and a major management challenge. After the successful final design review process, the construction of NEON was approved formally by the National Science Board in May 2010, and the National Science Foundation resumed infrastructure construction funding in August 2011.[374]

[373] R. Tawa, "National Ecological Observatory Network Cyber Infrastructure (Ci)," (NEON, 2009).

[374] Anthony Beasley, "Written Testimony of Dr. Anthony Beasley, Chief Operating Officer and Project Manager, National Ecological Observatory Network (Neon), Inc. Before the United States House of Representatives Subcommittee on Research and Science Education Hearing Entitled "Nsf Major Research Equipment and Facilities Management: Ensuring Fiscal Responsibility and Accountability"," (Washington D.C.: NEON, 2012).

4.4.3 Assessment by Congress and a Scandal

Tony Beasley, NEON project manager and Chief Operating Officer appeared before the United States Congress Subcommittee on Research and Science Education in a hearing titled "NSF Major Research Equipment and Facilities Management: Ensuring Fiscal Responsibility and Accountability."[375] At this point, NEON's chief project manager cited the challenge "to help our scientists understand the formal project management techniques needed to produce a facility design and operations model on the scale being considered" as the major problem encountered during the initial ecological research infrastructure construction. The management around David Schimel as CEO of the non-profit management entity NEON Inc. based in Boulder, Colorado, was still painting a picture of a smooth rollout of the large infrastructure project. Yet, things were not going as well as they appeared, as budget overruns in engineering and cyberinfrastructure had started to inflate NEON's need for cash by millions of dollars every year.

With construction well underway across the 20 ecosystem regions covered by NEON, the NSF Inspector General Allison C. Lerner testified in front of the US House of Representative's Committee on Science, Space, and Technology in March 2014. The NSF had outsourced the financial auditing process to the contractor Defense Contract Audit Agency (DCAA). In 2011, DCAA had attempted to audit NEON's proposed 433,7 million US$ budget three times and deemed the budget inadequate for audit and approval each time.[376]

NEON Inc. had submitted a revised budget in February 2012, which was again audited by DCAA who found more than a third of the proposed costs to be "unsupported" and questionable.[377] However, the NSF itself dismissed the concerns of the Inspector General and in April 2014 "asserted that NEON had

---

375 Ibid.

376 Lerner, "Statement of Allison C. Lerner, Inspector General National Sciece Foundation before a Hearing of the Committee on Science, Space, and Technology." p. 1

377 Ibid. p. 2

supplied supporting documentation for the proposed costs."[378] Particularly so-called "management fees" were contested as well as other contingency costs. For example, NEON Inc. had filed 25.000 US-dollars for a holiday party as well as 11.000 US-dollars for coffee services as management fees, a clear violation of the NSF accounting standards according to Inspector General Lerner.[379]

However, according to Allison Lerner, budget irregularities were haunting most of the National Science Foundation's major infrastructure projects. An audit of the Ocean Observatories Initiative found 88 million US-dollars of questionable contingency costs, around a quarter of the total infrastructure budget. Yet, NEON was headed for more serious budget and management troubles. On April 27, 2015, the DCAA senior auditor Joshua Kirk McGill submitted 'whistleblower' evidence to Congress claiming that senior officials had ordered the removal of evidence for questionable use of contingency costs by NEON from the audit reports. McGill stated that his team "discovered that the grantee [NEON] was illegally funneling millions of dollars of taxpayer money into unlawful activities including lobbying, extravagant Christmas parties, employee gifts, and the like," and he further went on to claim that "upper-level Defense Contract Audit Agency management has undertaken a systematic campaign to cover-up the audit results."[380]

Unfortunately, McGill's evidence was substantial and not only revealed that NEON Inc. had mishandled funds, but by 2015, NEON was also running seriously over budget. The National Science Foundation was forced to let go of the NEON CEO Russell Lea and bring in Eugene Kelly as a temporary replacement. NEON Inc.'s James Collins and James Olds, Assistant Director of the NSF Directorate for Biological Sciences, had to report to the US

---

[378] Ibid. p. 2

[379] Ibid. p. 3

[380] McGill, "Disclosure Pursuant to Allegations Related to Suspected: Fraud, Gross Mismanagement, Gross Waste of Funds, Abuse of Authority Etc." p. 3

Congress Committee on Science, Space, and Technology that NEON Inc. had accumulated a budget overrun of more than 80 million US-dollars and was already 18 months behind schedule.[381]

NEON Inc. was aware of the budget issues and had assembled recommendations for a reduction of the infrastructure's scope in August 2015.[382] The report proposed to eliminate 15 monitoring sites and scrap several types of expensive sensor equipment to lower costs.[383] However, James Olds of the NSF was not satisfied when NEON Inc. submitted its revised plan, which also included "additional costs and a further delay of 2 years."[384] By 11 December 2015, the NSF terminated its contract with NEON Inc. and started to look for a new way to manage the troubled infrastructure project. After several months of negotiations with several contenders, the NSF signed a contract on 11 March 2016 with the private nonprofit research and development organization Battelle Memorial Institute to take over management of NEON.[385] Despite being a nonprofit enterprise, Battelle has deep ties to the United States military and national security community.[386] Battelle manages several national laboratories for the United States Department of Energy, e.g. the Lawrence Livermore National Laboratory in Berkeley, CA, a major center of nuclear weapons research. Battelle's Chief Executive Officer Lewis Von Thaer was a formerly an executive at such defense contracting companies such as DynCorp and General Dynamics.[387]

---

[381] "Neon Warning Signs: Examining the Management of the National Ecological Observatory Network.," (Washington D.C.: U.S. House of Representatives Committee on Science, Space, and Technology, 2015).

[382] Todd Dawson et al., "Neon Scope Management Recommendations," (NEON, 2015).

[383] J. Mervis, "Neon Contractor Hanging by a Thread, Nsf Tells Congress.," *Science*, 21 September 2015 2015.

[384] "Nsf Fires Managers of Troubled Neon Ecology Project.," *Science*, 11 December 2015 2015.

[385] "Nsf Picks Battelle to Run Neon," *Science*, 14 March 2016 2016.

[386] See www.battelle.org

[387] https://www.battelle.org/about-us/management

The National Ecological Observatory Network was set up as an infrastructure to Command and Control the United States national environment in the face of the impacts of a changing climate. We have seen how the community of ecological scientists was unable to manage a knowledge infrastructure according to the allocated budget and timeframe. Thus, it appears as no surprise that a defense contractor such as Battelle has more expertise in setting up such an observation infrastructure. While the transition of management to Battelle appears to have turned around the NEON project, a final assessment is hard to formulate since the infrastructure is still in the process of being rolled out. Nevertheless, I will next attempt to formulate some conclusions about the major themes in the planning and construction of the National Ecological Observatory Network and its relation to notions of Fourth Paradigm Science and the Command and Control Anthropocene.

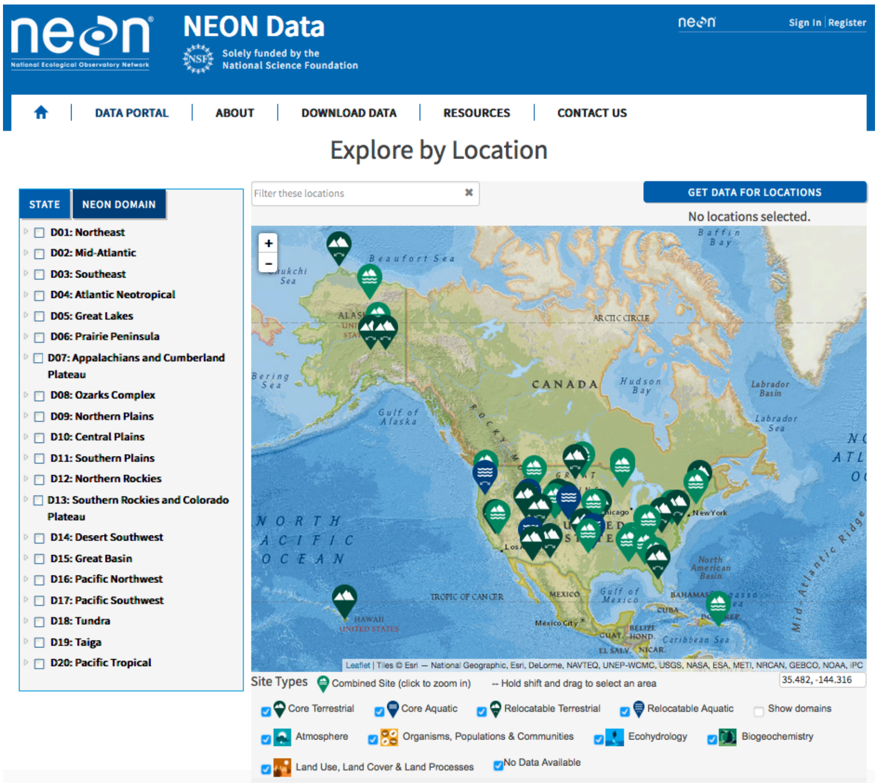


Figure 16 The NEON Data Portal[388]

[388] http://data.neonscience.org/home

4.5 Conclusion: A Digital Ecological Network

The U.C. Berkeley ecologist James Hunt contributed an essay to the Microsoft Research publication *The Fourth Paradigm*, titled "Redefining Ecological Science Using Data", in which he discussed the application of Jim Gray's ideas to the field of ecology in general and the National Ecological Observatory Network in particular. He diagnosed both an expanding ubiquity of available ecological data as well as specific challenges to data-intensive research in the field:

> Sensor deployments by research groups are shifting from short campaigns to long-term monitoring with finer-scale and more diverse instruments. Satellites give global coverage particularly to remote or harsh regions where field research is hampered by physical and political logistics. Internet connectivity is enabling data sharing across organizations and disciplines. The result of these first three factors is a data flood. [...] Unlike sciences such as physics or astronomy, in which detectors are shared, in ecological science data are generated by a wide variety of groups using a wide variety of sampling or simulation methodologies and data standards.[389]

According to Scott Collins, the first NEON program director at the National Science Foundation, "the idea for a large ecological observatory sprang from NSF staff who were seeking ways for biologists to get a slice of the agency's big-science money: the Major Research Equipment and Facilities Construction budget."[390] Thus, the initial initiative for the National Ecological Observatory Network did not arise among the scientific community of biologists and ecologists, but was hatched by science administrators. At the outset in 2000, the best model for a large-scale ecology project would have been the Long Term Ecological Research (LTER) network. However, as was discussed above, despite its success, the LTER was not designed as a unified

[389] James R. Hunt "Redefining Ecological Science Using Data", in Hey, Tansley, and Tolle, *The Fourth Paradigm. Data-Intensive Scientific Discovery*. p. 23

[390] Chris Cesare, "Ecology Aims High," *Nature* 529, no. 21 January 2016 (2016).

infrastructure, but as a network of individual and site-specific projects proposed by individual groups of investigators. NEON was, from the very beginning, more driven by science policy than by individuals from among the ecology research community – this is one of the seeds of trouble contained in the design of NEON as such.

As James Hunt had pointed out, the issue with constructing a continental scale ecological network is that there is, on the one hand, a large variety of sampling methods and data standards that are specific to the local conditions, and that, on the other hand, a large network cannot function without a common standard for measurements and data formats. This dilemma affected the way that measurement sites and participating institutions were selected.

> The diversity of ecological dataset size, dataset semantics, and dataset publisher concerns poses a cyberinfrastructure challenge [...] Synthesis science drives not only direct conversations but also virtual ones between scientists of different backgrounds. Advances in metadata representation can break down the semantic and syntactic barriers to those conversations.[391]

Biology and ecology differ from other natural sciences with respect to the types of data generated by their infrastructures. The Earth Sciences community had been quite successful in 'making data global' during the International Geophysical Year, in 1957, since they could build on existing standards with respect to data formats. Notably, the types of shared data were also not nearly as diverse as the plethora of data encountered in ecological research.

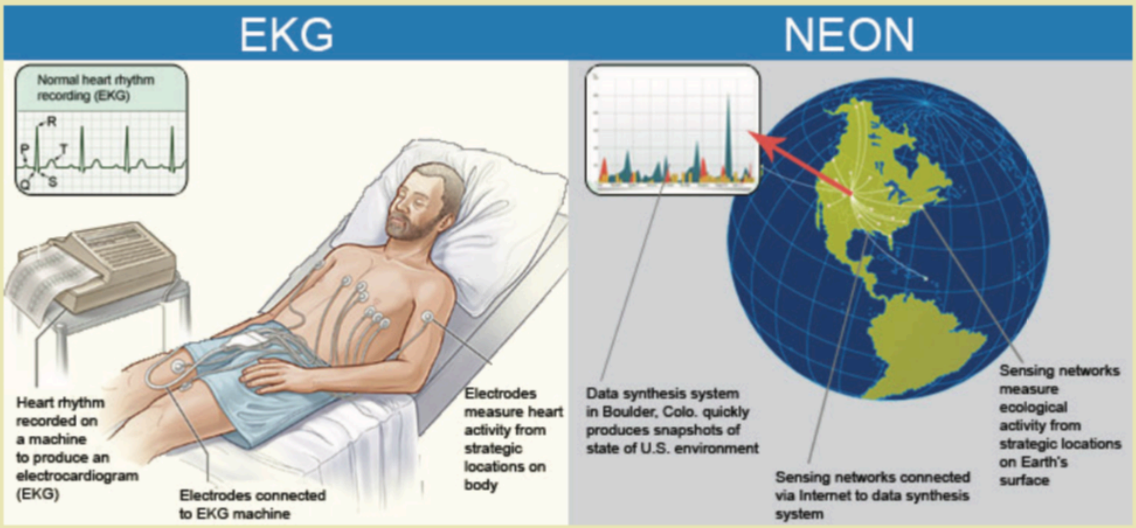
In the construction of NEON, the issues concerning a broad range of data types and challenges with varying or lacking standardization eventually played out at the level of database hardware and software. James Hunt's hope expressed above was that an interdisciplinary synthesis on a virtual level, on the level of a database and its interfaces, would foster new kinds of interdisciplinary knowledge. The "semantic and syntactic barriers" Hunt spoke of refer both to

[391] James R. Hunt "Redefining Ecological Science Using Data", in Hey, Tansley, and Tolle, *The Fourth Paradigm. Data-Intensive Scientific Discovery*. p. 26

the semantics and syntax of a database as well as those of the language of communication between individual human scientists and funding institutions.

The narratives deployed to legitimize and justify the investment in NEON were fueled by hopes of public participation as well as promises of national security and control in the face of environments changing due to global warming. The analogy between the 'health of the nation' and the 'health of the human body' was often used to illustrate the use and benefits of NEON. In particular, the topos of public health served as a site of translation between the politics of healthy bodies, a healthy public, and a healthy ecosystems that are construed to be part of the overall national body. In this way, the national ecological environment becomes part of the United States' national 'body' to be kept secure by Command and Control of the Anthropocene forces.

A figure from a NEON information brochure exemplifies this by depicting NEON as "an EKG for the Earth," just like Al Gore had envisioned Digital Earth to be a monitoring system for an intensive ward patient. The metaphor of intensive care perfectly captures all aspects of an environmental research infrastructure we have elaborated: narratives of care, security in the face of a threatened body, and command and control to keep the threatened "nation's ecological health" in check (see Figure 17 below).



Credit: National Science Foundation

NEON: AN EKG FOR THE EARTH

An electrocardiogram (EKG) is a single instrument comprised of geographically distributed components; it is used to measure heart activity to detect problems in heart health. Similarly, NEON will be a single instrument comprised of geographically distributed components; it will be used to measure ecological activity to detect problems in the nation's ecological health.

Just as each EKG sensor is located to provide a unique 'view' of heart activity, each NEON sensor and field crew will be strategically located to provide a unique 'view' of ecosystem activity. And similar to EKG sensors that continually pulse standardized data to a centralized, synthesizing computer that displays resulting analyses, NEON sensors and field crews will continually pulse standardized data to a synthesizing cyberinfrastructure headquartered at NEON Inc. that will publish resulting information.

As a single observing scientific instrument, NEON is commonly referred to as an 'observatory' by scientists—even though NEON differs from the stereotypical conception of an observatory as a large telescope occupying one location.

Figure 17 NEON: An EKG for the Earth[392]

Yet, a closer look at the planning and construction of NEON has also revealed the challenges and inconsistencies inherent to such a narrative of a Command and Control Anthropocene approach to managing and adapting to a changing climate. Controversies within the community of ecological sciences concerning centralization versus decentralization, network versus systems character of the infrastructure, mirror the fact that narratives of control and security inherently contradict the aims of public participation, openness, and transparency. Just as it is a specialized cardiologist making decisions based on an EKG, the output of NEON data is, at least implicitly, tailored for centralized decision-making.

---

[392] From National Science Foundation, "National Ecological Observatory Network: Revolutionizing Ecological Research " (National Science Foundation, 2013).

The conclusion to the case study of the Ocean Observatories Initiative outlined several issues faced by environmental research infrastructures in general. Data access and cyberinfrastructure management was one such challenge, while the security of long-term funding for a long-term knowledge infrastructure posed the largest issue. Similarly to OOI, and particularly after the NEON mismanagement scandal, long-term funding for the infrastructure was in doubt, especially since NEON still had to prove its merits. It remains to be seen if the National Science Foundation will continue its push for large environmental research infrastructures such as OOI and NEON, or if the ecological science community turns their focus back to more local and decentralized ventures modelled after the Long Term Ecological Research (LTER) network.

Thus, it is still an open question whether environmental research infrastructures will endure and how decision-makers and the public are going to use them. Are narratives of security and participation actually going to be realized? Or could there be other ways of regarding and deploying the data provided by NEON and OOI that could actually empower those affected by climate change impacts beyond notions of the Command and Control Anthropocene? These questions are going to be addressed in the following chapter.

Chapter 5: New Narratives and Policy for Environmental Research Infrastructures

5.1 Introduction: Reflecting on Environmental Research Infrastructures beyond the Command and Control Anthropocene

The preceding chapters have outlined how environmental research infrastructures have come to construct a frame of narrative concerning what an environment is, how it changes dynamically, and how it can be perceived and managed by scientists, experts, and government institutions through public access via online virtual observatories. In particular, we have seen how an environmental research infrastructure such as the National Ecological Observatory Network was invested with narratives of public health and security by Command and Control that contrasted with visions of public participation and openness.

Chapter 1 outlined notions of 'Big Data' and ideas for environmental research infrastructures by presenting both current and former discourses about the epistemology of massive data collection, data storage, and environmental management. It introduced us to Jim Gray's idea of a Fourth Paradigm Science, in which one perceives environments through digital interfaces based on massive digital knowledge infrastructures. I also outlined Al Gore's idea of a Digital Earth infrastructure program that was supposed to deliver both economic growth and freedom of information as a solution to a global environmental crisis.

Chapter 2 took a closer look at Jim Gray as the proponent of Fourth Paradigm Science and showed how his work and career were entangled with several leaps in digital database technologies since the 1970s that have laid the foundation for the large knowledge infrastructures being constructed in environmental research in the United States since the 1990s.

Chapters 3 and 4 then traced the genesis, structure, and challenges surrounding two data-intensive environmental research knowledge infrastructures, the Ocean Observatories Initiative, and the National Ecological Observatory Network, that were drawing on ideas of Fourth Paradigm Science and exemplify the Command and Control Anthropocene.

This concluding chapter will now attempt to reflect on the investigated case studies, OOI and NEON, to draw conclusions with regard to the question of how an environmental research infrastructure could be construed beyond the narratives of security and surveillance in the Command and Control Anthropocene. It will also formulate tentative recommendations for scientific experts, the lay public, and decision makers in science policy regarding how more community-oriented, rather than 'paranoid' narratives could be fostered and some of the challenges faced by long-term environmental research infrastructures could be addressed

To generate such recommendations, this chapter is going to consider what we have learned from taking a critical look at the applicability of the Fourth Paradigm science to environmental research as well as the challenging implementation of large environmental research infrastructures such as the Ocean Observatories Initiative and the National Ecological Observatory Network.

Yet first, I will take a step back and think about the vantage point of seeing an environment through a virtual observatory. How can the information generated by environmental research infrastructures such as the ocean Observatories Initiative and the National Ecological Observatory Network generate knowledge that is 'local' as well as 'global' knowledge?

Then, I will question Al Gore's vision of a connected global environmental consciousness by asking if a digital network is capable of creating the experience necessary for meaningful political debate and environmental engagement, or if the digital inherently produces a disempowered 'user' and political 'idiocy', as Alexander Pschera suggests.

I have argued that notions of Al Gore's Digital Earth or Jim Gray's Fourth Paradigm Science have informed the environmental research infrastructure projects examined in the case study chapters. We also need to remind ourselves of how closely related the technologies deployed in these infrastructures were to military surveillance technologies originating from the Command and Control systems of the Cold War United States. Thus, I follow the scholar and former network engineer Tang Hui-Hu in stating that digital infrastructures such as 'the cloud' are inherently paranoid networks (just as Command and Control systems) that align with and promote a neoliberal capitalist economic order, just as Al Gore had imagined in his Mission to Planet Earth.

After discussing these fundamental critiques of the narratives of political participation and security informing environmental research infrastructure projects in the United States, I will move on to discuss the theoretical and political question of how to maintain environmental research infrastructures over the long-term and how to use and disseminate the knowledge produced by them in policy-making.

To oppose the narratives around environmental research infrastructures with possible alternatives, I will refer to Deborah Stone's two contrasting policy models, the *market model* and the *polis model*, developed in her book *Policy Paradox*.[393] Based on the polis model I will suggest policies for managing environmental research infrastructures in the Command and Control Anthropocene in a more inclusive way. I will then discuss how Roger Pielke's concept of the "honest broker" could apply to research results generated from large environmental research infrastructures. Based on the polis model and the 'honest broker', we will point out some major challenges for environmental

---

[393] Deborah Stone, *Policy Paradox. The Art of Political Decision Making. Third Edition.* (New York: W.W. Norton & Company, 2011).

The concepts of market model versus polis model is used as an analytic descriptive method of policy analysis by Stone. We are going to use this approach in a normative way to generate policy options beyond what is currently proposed.

research infrastructures in the fields of knowledge generation, knowledge sharing, and knowledge maintenance.

Finally, I will close by formulating an appeal to not just make 'global data' but to 'make data terrestrial', speaking with Bruno Latour, in order not just to Command and Control environments in the Anthropocene but to inhabit them with both feet planted firmly on the ground.

5.2 'Seeing the Global' in a Virtual Observatory

The Vantage Point of a Virtual Observatory

We have seen in the case study chapters on the Ocean Observatories Initiative and the National Ecological Observatory Network how a vast knowledge infrastructure 'draws' a multi-layered map forming a virtual observatory. The 'environment' seen through the lens of an environmental research infrastructure appears to the user of a virtual observatory similar to the way a landscape appears to the pilot of a modern airplane as mediated and 'drawn' by their sensors and instruments.[394]

The vantage point provided by a virtual observatory, as is the case for the OOI and NEON data portals, is usually that of a 'view from above', it is a cartographic representation of various database layers.[395] And although the multiplicity of layers in the virtual observatory implies that one cannot view all layers of the observatory at once, the format of the observatory as such nevertheless suggests a totalizing view from above, although not quite a 'view from nowhere'.[396]

---

394 Compare "Ils appartiennent au territoire dessiné par leurs instruments." In Bruno Latour, *Face À Gaïa. Huit Conférences Sur Le Nouveau Régime Climatique* (Paris: Éditions La Découverte, 2015).p. 326

395 Compare Sebastian Vincent Grevsmühl, *La Terre Vue D'en Haut. L'invention De L'environnement Global* (Paris: Éditions du Seuil, 2014).

396 Thomas Nagel, *The View from Nowhere* (New York ; Oxford: Oxford University Press, 1986).

Infrastructure inversion has shown that the knowledge infrastructures underlying a virtual observatory, however, are everything but located 'above'. Thus, a view from above, a 'global view', a view of 'the oceans' or the 'national ecology' are in fact situated in a network of artefacts, people, and institutions, as we have seen in the case study chapters. This situatedness is what links the globalizing view of the virtual observatory to the localities of its infrastructure components. To a historian of science, or Science and Technology Studies scholar, any claim of 'objectivity' or a 'global view' is suspect, as Bruno Latour states in his *Face à Gaïa*, "one is never more provincial than when one pretends to have a 'global vision.'"[397]

Let us now further interrogate the question of situatedness of an environmental monitoring infrastructure from three angles. First, I will question the challenge of the situatedness of the knowledge infrastructure using Donna Haraway's notion of "situated knowledges". Second, I will ask what an experience or 'exposure' to environments via data products or virtual observatories could mean. And, lastly, how such an exposure relates to ideas of Command and Control, and what Tang-Hui Hu has called 'paranoid knowledge'.

Situated Knowledge Infrastructures

In her 1988 essay "Situated Knowledges", Donna Haraway develops the dilemmas of opposing 'normal' science from a feminist perspective while wanting to maintain some form of grip on 'truth' and 'objectivity'.[398] Haraway attempts to go beyond a mere deconstruction of 'objectivity' and states that the "issue in politically engaged attacks on various empiricisms, reductionisms, or other versions of scientific authority should not be relativism – but location."[399]

---

[397] Latour, *Face À Gaïa. Huit Conférences Sur Le Nouveau Régime Climatique*."On n'est jamais aussi provincial que lorsqu'on pretend avoir un 'vision globale'." p. 179

[398] Donna Haraway, "Situated Knowledges: The Science Question in Feminism and the Privilege of Partial Perspective," *Feminist Studies* 14, no. 3 (1988).

[399] Ibid.p. 588

She then goes on to oppose “universal rationality” with various “ethnophilosophies”, “common language” with “heteroglossia”, “world system” with “local knowledges”, and “master theory” to “webbed accounts”. As Haraway herself acknowledges, however, these dichotomies quickly fall apart when applied in practical considerations of specific examples.

“Webs *can* have the property of being systematic, even of being centrally structured global systems with deep filaments and tenacious tendrils into time, space, and consciousness, which are the dimensions of world history,” Haraway points out.[400] The environmental research infrastructures I have considered are situated between the statuses of “earth systems” and “local knowledge” as well as “master theory” and “webbed accounts”. The oscillation between the web or network characteristics and systems-building was a central notion in our analysis of environmental research infrastructure projects.

To criticize the kind of Fourth Paradigm science being implemented by OOI or NEON by proclaiming that they subsume ‘local knowledge’ simply as part of a ‘world system’ of an environmental database is not entirely convincing. No prior account of the state of the continental environment or the ocean assembled as much ‘localized knowledge’ by means of the various technologies and nodes of the monitoring network, as these knowledge infrastructures have done.

The assembly of myriad local data should not be regarded a disregard for local knowledge. In addition, local knowledge can only gain recognition in relation, and comparison, to other locales as well as general assumptions about the state of the world. Situated knowledge is only relevant as a contrast and a complement to contrasting established accounts. Even though the interface of a digital observatory presents the image of an ‘objective’ representation of the state of the environment, it is at the same time a highly ‘webbed account’ of the environment.

[400] Ibid.p. 588

Where exactly is this knowledge situated? The 'sensory organs' of knowledge infrastructures are not situated bodies, as in Haraway's account, but situated sensors and situated 'cyborgs'. How could a cyborg sensor ever be situated?[401] 'Situated', according to Oxford English Dictionary simply means 'placed', 'in a place'; in this sense, a cyborg or a sensor is just as situated as any human. Thus, could an infrastructure such as OOI or NEON also be regarded as 'situated knowledge' in Haraway's sense?

There is a dialectic of situated knowledge that also applies to knowledge infrastructures. Knowledge infrastructures are at once situated and local as well as distributed and universal. This double nature of environmental research infrastructures can be further described using the concept of "distributed cognition".[402] Just as navigation on a ship depends on the interplay of various cognitive agents, tools, and knowledge bases, an infrastructure comprising a virtual observatory can be described as a case of 'distributed cognition' that is at once local and universal. Edwin Hutchins described in his book *Cognition in the Wild* the ways in which humans have organized ways of cooperative processes of cognition such as navigation at sea. This kind of cognition is called 'distributed' since no single individual would be able to carry out the entire navigational procedure by herself. Also, 'distributed cognition' represents a kind of assemblage of various human and non-human entities and technologies that are necessary to navigate a ship or surveil an environment.

Since Hutchins' book was first published, navigation has largely left behind the methods he described (such as triangulation, multiple bearing, although these techniques are still taught to sailors in training), and is now almost entirely reliant on the Global Positioning System (GPS). Navigation via the satellite-based GPS network is much more similar to the distributed knowledge infrastructures we have described in the case studies of OOI and NEON. An important addition is, however, that such a knowledge infrastructure can

---

401 See Hanwahr, "Marine Animal Satellite Tags."

402 See Edwin Hutchins, *Cognition in the Wild* (Cambridge, MA: The MIT Press, 1995).

function via a network of distributed technological and non-human entities, and only relies on humans for maintenance and as the target audience of the generated information.

How do environmental research infrastructures transform the role of experts in evidence-based environmental policy? What narratives can one tell with a database and a knowledge infrastructure that one could not tell before? How does it transform the role of the 'user' attempting to engage with and understand our changing environments. Some critics of the digital have severely questioned whether such a mode of engagement could ever produce a meaningful experience and political action at all, as I will discuss next.

Presence, Experience, and political Idiocy

How does 'experience' of nature and environment take place when the experience of an environment by a 'user' is mediated by knowledge infrastructures and digital observatories? The German critic Alexander Pschera postulates, in his short tome *Vom Schweben*, that the "voyage across the ocean of knowledge is not a voyage of character."[403] Thus, a crucial question is how an experience of the environment that is integrated with an individual's biographic experience, relating to a subject in a way that could alter her narration of self could even take place in digitally mediated environments. Pschera calls this the 'idiocy of the digital': "Alone with oneself and dissolving in a digitally enclosed Other: that is the idiocy of the digital."[404]

Referring to the old Greek meaning of *idiot* as a private individual who does not participate in the political life of the polis, Pschera suggests that while the user of a digital observatory may be connected to troves of information, she does not participate in the public sphere of political discourse, and thus is not

[403] See Alexander Pschera, *Vom Schweben. Romantik Im Digitalen* (Berlin: Matthes & Seitz, 2013)."Reise über den Ozean des Wissens ist keine Charakterreise." p. 13.

[404] "Allein zu sein mit sich selbst und sich aufzulösen in einem auf digitalem Wege nahegebrachten Fremden: Das ist die Idiotie des Digitalen." Ibid.p. 16.

actually empowered by the mass of information. Rather, the illusion of knowledge, when one is merely drowning in disconnected information, turns the user into an *idiot* rather than an environmentally informed citizen.

The paradox of the user being simultaneously hyperconnected as well as 'idiotically' isolated from any meaningful political empowerment merely creates a 'phantom of proximity' to an environment, or as Pschera puts it, "being part of a technological space is not participation, but dissolution in phantom-like proximity."[405]

Thus, while the user of a digital observatory can conjure up any kind of information on the spot, from home, or even on a mobile phone, the experience of the locality of an environment dissolves into the mere 'event' of information. This paradox also applies to the way one might 'visit' foreign locales via tools such as Google Earth or Street View. Thus, more information about an environment, even mediated in an immersive digital observatory, is not necessarily going to induce an actual 'experience' of a natural location.

Furthermore, a knowledge infrastructure cannot constitute an empowered community capable of political action and meaningful decisions. A cybernetic view of the relation between the environment and individuals in society is fundamentally anti-discursive and void of meaningful narratives that could motivate consequences and actions. Alexander Pschera shares this view in his skepticism of Social Media as a tool for political communities: "This is a cybernetic process at work [...]. That is why there can be no grassroots democracy based on the Internet. Life in such communities does not have to engage discursively, but do just one think: to click or not to click."[406]

---

[405] "Teilnahme am technischen Raum ist nicht Partizipation, sondern Entgrenzung in phantomhafter Nähe." Ibid.p. 25.

[406] "Hier ist ein kybernetischer Prozess am Werk [...]. Deshalb kann es keine Basisdemokratie geben, die auf dem Internet beruht. Das Leben in einer solchen Gesellschaft muss sich nicht mehr diskursiv einlassen, sondern nur noch eines tun: Klicken oder nicht Klicken." Ibid.p. 35.

Thus, Pschera sees a 'non-discursive society', one that "does not pay heed to arguments any longer, only to quantities."[407] However, aren't virtual observatories exactly attempting to deliver more than just quantities of data, but to assemble data in a way that renders a new quality of experience possible? At least, that is their goal, although we may question whether such a quality is actually achieved in practice.

Another effect of the dissolution of discourse in Fourth Paradigm Science is the loss of historical context, Pschera goes on to claim: "Numerization dissolves discourse as well as the consciousness of one's own historicity."[408] From the perspective of Paul Edwards, however, climate science in particular, has been shown to have an acute awareness of its historical contexts, a necessary condition of creating global data with long timelines. Yet, Pschera's point concerns the political epistemology of scientific results more so than scientific practices.

According to Alexander Pschera, a digital infrastructure reproduces reality in a specific framework that transforms actions within this infrastructure into a game-like agency, also called 'gamification'. Pschera thus speaks of a "fundamentally ludic structure of the digital."[409] We have seen in the case studies how strongly the user interfaces of the Ocean Observatories Initiative and the National Ecological Observatory Network databases resemble forms of infotainment. Anyone using apps to track sharks or air quality sees the obvious family resemblances to gameplay application.[410]

---

[407] "nicht mehr auf Argumente hört, sondern nur noch auf Quantitäten" ibid.

[408] "In der Numerisierung löst sich der Diskurs ebenso auf wie das Bewusstsein der eigenen Geschichtlichkeit." Ibid.p. 42.

[409] Ibid.p. 71.

[410] Compare e.g. the Apps Haze Today: https://www.haze.gov.sg/

and Shark Tracker from Ocearch: https://www.ocearch.org/tracker/?list

Again, we need to stress the relation between databases, narration, environments, and their histories. Pschera, echoing Ursula Heise's notions of 'databases as the new epics', states that narration has been displaced by "chronicles" in the digital realm, but also in political discourse.[411] Knowledge about environments is assembled in a database, just like animals are assembled in a zoo, lacking context, yet framed by digital bars. Pschera uses the metaphor of the "aquatic petting zoo" to describe how all things are always at hand, tangible and yet, floating out of context in a deep blue ocean.[412]

How does this relate to how we deploy environmental research infrastructures to shape the course of society and, in particular, environmental policy? Pschera, in the conclusion of his short book *Vom Schweben*, offers a suggestion of how to deal with the way digital technology is mediating and de-corporealizing our environmental experience in many ways. His strategy is to "refuse the technological mode of understanding and to regard the digital reality as a piece of theater that can help us better understand the technologically transformed world."[413] Yet, we have seen in the case studies that knowledge infrastructures are much more than theater, they were built with the purpose to produce valid knowledge based on data that is invested with its own sovereignty by the individuals, institutions, and technologies involved. It may be possible to regard the visions of political participation and environmental consciousness that research infrastructures are invested with as a kind of theater. However, considering the actual capabilities of the environmental infrastructures to surveil and manage environments, the narrative of Command and Control of the changing Anthropocene environment is much harder to deconstruct as mere theater.

---

[411] Pschera, *Vom Schweben. Romantik Im Digitalen*.p. 77.

[412] "aquatischer Streichelzoo" ibid.p. 86.

[413] "Die Strategie gegen die Körperlosigkeit des Wirklichen besteht darin, die Form des technischen Verstehens zu verweigern und in einem ästhetischen Akt die digitale Wirklichkeit als ein Theater zu begreifen, das uns dabei helfen kann, die technisch transformierte Welt besser zu verstehen." Ibid.p. 92.

5.3 Sovereign Data and Paranoid Knowledge

Challenging the Premise of Infrastructure Inversion?

In his book on the history of cloud storage technology *The Prehistory of the Cloud*, the scholar and former telecommunications engineer Tung-Hui Hu questions the narratives that data is invested with in projects such as environmental research infrastructures. He calls the construal of data as a kind of factual essence "the sovereignty of data" that "comes out of the way we invest the cloud's technology with cultural fantasies about security and participation."[414] We have seen in the case studies how environmental research data generated by the infrastructures of the Ocean Observatories Initiative or the National Ecological Observatory Network were, likewise, loaded with narratives of national and public security as well as political and individual participation.

Critics such as Alexander Pschera have suggested that this kind of data sovereignty and its accompanying narratives of security and participation can be made more transparent by regarding them as a kind of digital theater. Yet, Hu believes that such a critique cannot be applied to a technology such as the cloud since it does not constitute one unified medium. One cannot analyze large environmental research infrastructures as a medium in the same way a media critic would analyze television or radio. Hu writes:

> Scholars claim that an awareness of the medium's materiality will lead to a more effective understanding of its ideological content. Yet the cloud, I am arguing, inevitably frustrates this approach, because by design, it is not based on any single medium or technology; it is medium-agnostic, rather than medium-specific.[415]

---

[414] Tung-Hui Hu, *A Prehistory of the Cloud* (Cambridge, MA: The MIT Press, 2015).p. xvi. He also goes on to claim that "The sovereignty of data is ultimately a politics of death, a "necropolitics," p. xviii.

[415] Ibid.p. xix.

If Tung-Hui Hu is correct in claiming that the cloud is medium-independent, his argument could likewise be made for the kind of knowledge generated by Fourth Paradigm Science or environmental research infrastructures. We have seen in the case studies that data-intensive environmental science also does not depend on just one medium or technology, which makes a comprehensive description of its infrastructures challenging. Does this, in turn, imply that data-intensive science is oblivious to a method such as infrastructure inversion? My claim has been precisely what Hu questions, namely that "an awareness of the medium's materiality will lead to a more effective understanding of its ideological content."[416]

However, why would 'the cloud' be such an oblivious technology that cannot be reduced or at least related to its material infrastructure via infrastructure inversion? I want to suggest that Hu's argument shows that, while historicizing the materiality of a knowledge infrastructure, we should address the question whether or not scientific infrastructures such as OOI or NEON constitute a medium as such.

Since I have not claimed that an entire knowledge infrastructure functions as a medium, in a way that e.g. television would, but is an assemblage of technologies, institutions, policies, and practices, the chosen methodology of infrastructure inversion can still serve the purpose of unveiling ideological and political scaffolding. In fact, one could say that Fourth Paradigm Science, which used cloud technology as part of its infrastructures, is a 'medium' similar to 'the cloud' in the way that it is "medium-agnostic". Yet, the same is not true for individual infrastructures in environmental research.

---

[416] Ibid.p. xix.

Paranoid Knowledge and the Information Market

Nevertheless, what may be true for 'the cloud' could still apply to the virtual observatories built on top of the environmental research infrastructures examined in the case studies. Hu calls the cloud "an enabler of supposedly distributed publics," and reaching and engaging distributed heterogenous publics is exactly what the virtual observatories are supposed to do.[417] The transparency the virtual observatories suggest may be just as misleading as most narratives bolstering the sovereignty of data may be: "As transparent and useful as these cheerful, rainbow-colored graphs may be in a world of big data, these interfaces are nevertheless visual fictions, ways of simplifying a hopelessly complex totality."[418]

However, more interesting than emphasizing that 'the map is not the territory' may be the kind of relation between the public and the private that a virtual observatory presenting 'sovereign data' can establish. It has been a central premise of this study that knowledge infrastructures are shaped according to many actors, institutions, individuals, and technologies, which is why there is always a potential "physical manifestations of a resurgence of sovereign power within the realm of data."[419] The data does not speak for itself, rather it can form the basis of a narration. I had already mentioned above how ideas of Fourth Paradigm Science relate to the realm of environmental information in a managerial way, turning environmental research infrastructures into tools to manage, command and control a changing climate. This kind of management is then carried out according to principles of risk management, inherent to economic markets more so than ecosystems as such. In the cloud as well as in environmental research infrastructures, Hu would see the forces of

[417] Ibid. p. 33

[418] Ibid. p. 124

[419] Ibid. p. 110

neoliberalism at work, which aims "to subordinate the public sphere to the logic of the marketplace."[420]

In Al Gore's idea of Digital Earth the information network and the ecological network were connected to promote both environmental protection and economic growth. Even though environmental research infrastructures are invested with narratives of political participation, the sovereignty of data inherently usurps this democratic narrative of networked environmental consciousness as well. Virtual observatories do not constitute political movements, they constitute individual users consuming information via a fixed infrastructure delivering sovereign data. Thus, the way that virtual observatory interfaces used by environmental research infrastructures such as OOI and NEON are set up, they do not produce public discourse but individualized users. Likewise, Hu concludes that "the cloud produces users rather than publics, and therefore individual rather than collective action."[421]

Tung-Hui Hu, in his assessment of cloud storage technology, draws a sweeping conclusion for the relation between knowledge infrastructures and the state of our society: "A system of knowledge in which everything seems to be connected [...] is a paranoid epistemology that offers to reveal meaning buried beneath the surface, but also serves to lubricate the market mechanisms by which that meaning was created."[422]

How could this statement apply to environmental research infrastructures? Firstly, the common notion of ecology suggests a system of knowledge, in which everything is connected. Likewise, the dream of cyberculture prophets such as Stewart Brand was a world, in which all people, even all things, are connected in a seamless digital web of information. In addition, the origins of digital technology and the Internet in relation to the underlying paranoid mood

[420] Ibid. p. 63

[421] Ibid. p. 147

[422] Ibid. p. 122

of the Cold War in the United States are well established.[423] Similarly, the promises of Big Data or Fourth Paradigm Science are those of universal connectedness rendering new insights and meaning that were previously beyond the grasp of human cognition. Thus, one could call the epistemology informing virtual observatories in Fourth Paradigm Science "paranoid" in Hu's sense of the term.

Whereas the narratives of Command and Control, individual users and a paranoid network propping up a neoliberal economic order, may be an accurate characterization of the reality of environmental research infrastructures, there are other possible narratives. I had started out assuming that environmental research infrastructures could function as tools to broaden the scope of policy options. Building on that assumption, how could we ensure that the knowledge from environmental research infrastructures does not merely reside in the cloud but can take effect in the space of public life? Could there be a way to tell the story that opens up options beyond the neoliberal model of sovereign data?

[423] Edwards, *The Closed World. Computers and the Politics of Discourse in Cold War America.*

5.4 The Market Model versus the Polis Model

The claim that a paranoid epistemology informs environmental surveillance and data-intensive science in a way that promotes the ideologies of neoliberal markets appears reasonable. And yet, it is not the only narrative one can tell about environmental research infrastructures in the Command and Control Anthropocene.

To come up with new narratives and policy alternatives of using the knowledge generated by environmental research infrastructures I will oppose the *market model* view, promoted explicitly by Al Gore and criticized by Tung-Hui Hu, and contrast it with the so-called *polis model* to develop an alternative approach to policies dealing with environmental research infrastructures.[424]

The political scientist Deborah Stone has developed the contrasting models for concepts of society, *market* versus *polis*, as a descriptive method of policy analysis. The market model and the polis model differ chiefly in how they construe decision making, information, and drivers of change in a society. Under the market model, decision making aims to "maximize personal gain" and efficiency, information is believed to be "accurate, complete, fully available," while the "individual quest to maximize [...] welfare" is seen as the main source of change.[425] The polis model, in contrast, regards loyalties and "promotion of public interest" as central to decision making, treats information as "ambiguous, interpretive, incomplete," while the main sources of political change are construed to be "ideas, persuasion, and alliances."[426] Information in the market model is the critical resource in stimulating policy change so that acquiring knowledge to analyze and interpret information become essential to

[424] Modelled after Stone, *Policy Paradox. The Art of Political Decision Making. Third Edition.* (thanks to Sophia Kalantzakos for recommending this approach)

For an interesting case study that applies the market and polis models to changing US educational policy see Patricia Burch, "The Professionalization of Instructional Leadership in the United States: Competing Values and Current Tensions," *Journal of Education Policy* 22, no. 2 (2007).

[425] Stone, *Policy Paradox. The Art of Political Decision Making. Third Edition.* p. 35

[426] Ibid.

maximizing welfare. Unlike in the market model where data and information play an elevated role, information in the polis model is not necessarily acted upon but is policy relevant only in conjunction with localized experiences, community practices, and relationships.

	Market model	**Polis model**
Information and Data	Information market is worth as much as the potential savings from adapting to climate change. Data generated by private knowledge infrastructures is private, even if transparency is in the public interest.	Information is a source of empowerment for communities to adapt to environmental change. Data generated by private efforts is open to the public if it has value for science or is derived from public data.
Decision-making and Risk	Decisions and risk are a matter of assessment and calculation, thus knowledge infrastructures provide objective assessments.	Decisions are a matter of public discourse and perceived priorities, thus knowledge infrastructures can broaden the scope of political discussion.
Drivers of Change	Individual users drive change by deploying knowledge infrastructures to maximize potential of the information market.	Knowledge infrastructures foster novel alliances, relationships, and loyalties, creating new policy options.

(Derived from D. Stone, *Policy Paradox*)

Under the market model, data and information are part of a market and can thus be traded and exchanged for profit. This market for the information generated by environmental research infrastructure is valued as highly as the potential savings from adapting to climate change. Data generated by private knowledge infrastructures would be regarded as private, even if transparency is in the public interest. In contrast, under the polis model, information is regarded as a public source of empowerment for communities to adapt to environmental change. Data generated by private efforts that is derived from public data (e.g. from an environmental research infrastructure) has to be open to the public as well. Nevertheless, not all information is free, since contributions to knowledge production under the polis model are valued in accordance to their contribution to the discourse.

Decisions and risk are a matter of assessment and calculation under the market model. Thus, knowledge infrastructures provide objective assessments and evidence that would support one particular choice in decision making. In the polis model, decisions are a matter of public discourse and perceived priorities. Knowledge infrastructures can broaden the scope of political discussion and help suggest policy options and compromises between various political groups. Finally, in the market model, political change is driven by individuals. In the case of environmental research infrastructures individual users drive change by deploying knowledge infrastructures to maximize the potential of the information market and, in turn, change the way the environment is managed. The polis model, however, would assume that knowledge infrastructures can foster novel alliances, relationships, and loyalties, and thus can help create new policy options and broaden perspectives on aims beyond environmental management.

Privatization of Knowledge: Predominance of the Market Model

Since a main goal of monitoring is prediction, and thus the assessment of the risks of particular paths of development, environmental monitoring will play a role in informing evidence-based environmental policy. In particular, since the effects of anthropogenic climate change are unlikely to be avoided entirely, adaptation to the local effects of climate change is going to be a major focus of the knowledge generated by large environmental research infrastructures.

A recent report on environmental observation by the Royal Society Science Policy Centre implicitly embraced the market model concluding that "sustained observations are essential not just for scientific detection and attribution of climate change but also for the development of climate related services – e.g. quantitative measures of changing risks in various sectors of society and the economy associated with changing climate."[427] Government programs for mitigation and adaptation of climate risks and climate change effects can build on the information generated by environmental monitoring infrastructures, however, the potential of services in the information market could also be substantial.

Thus, the private sector is going to be another factor in the interplay of knowledge infrastructures, adaptation to climate change, and climate risk assessments. Energy companies have been at the forefront of generating knowledge about environmental change, even if they did not usually use this knowledge to prevent environmental impacts, but fortify their market position, like the fossil fuel giant Exxon did as early as the 1970s.[428]

---

[427] Royal Society, "Observing the Earth. Expert Views on Environmental Observation for the Uk," (London2015). p. 29

[428] David Kaiser and Lee Wasserman, "The Rockefeller Family Fund Vs. Exxon," *The New York Review of Books*, 8 December 2016 2016.

In contrast to multinational corporations like Exxon that extract and monetarize natural resources, insurance companies do not have the choice of ignoring the effects of climate change. Insurances have to pursue a strategy of discounting the externalities of the extraction of fossil fuels. The appropriate pricing of insurance policies in the face of risk and uncertainty is a core task for all insurance companies, and yet, in the face of the effects of climate change, grave miscalculations can cause financial losses for the insurers and leave the victims of climate risks helpless.

The narrative of security in environmental research infrastructures relates to notions from accounting and insurance, both in principle and in the methods used. Health insurance companies are eager to learn as much as they can about the health and lifestyle habits of the people insured by them. In particular, the data generated by health apps from smartphones is coveted by insurers, who want to offer individual plans based on individual risk factors and create incentives to lower individual risk by rewarding behavioral change with lower premiums. Such knowledge and monitoring could also be used to penalize those with unhealthy lifestyles unwilling to adjust to more healthy habits, which lets some people fear financial disadvantages and even a new regime of healthy lifestyle surveillance.[429]

Similarly, insurers could play a large role in assessing and pricing the risks created by a changing climate, after all, they are the ones who have to reckon with the costs of exacerbating weather events or damage from flooding by rising sea levels. In this case, the planet Earth really takes on the role of an insured sick patient, a simile deployed by Al Gore and the National Ecological Observatory Network as well.

---

[429] S. Hoffman and A. Podgurski, "Big Bad Data: Law, Public Health, and Biomedical Databases," *J Law Med Ethics* 41 Suppl 1 (2013).

For example, the German insurance company Allianz, among other large insurance companies, has created a climate risk assessment group to attempt to quantify the financial risks created by climate change for the insurance business and find policies that allow Allianz to insure for climate change related damages.[430] This suggests three possible scenarios: either insurers are going to be increasingly overwhelmed by the cost damages from climate change related natural disasters, causing them to go bankrupt and leaving victims alone with the damages. Or, secondly, insurers will assess the risk of insuring for certain damages as too high, or are only willing to insure those companies and individuals who can afford to pay very high premiums, which would exacerbate the inequalities in the effects of climate change on different socioeconomic groups. Thirdly, on a more optimistic note, more activist insurance companies could utilize the knowledge generated by environmental monitoring to reflect the risks and externalities of climate change on a local level more closely than environmental policy and energy markets have been able to do. This could be a way of correcting the colossal market failure that is climate change.

As we have seen in the critiques of Open Science, more transparency and access do not necessarily lead to a better understanding of how scientific knowledge is generated or enhance value generated from publicly accessible data. Thus, there are a number of ways in which public Open Science is transformed into private data as part of a business-oriented data services value chain. Extra value from Open Science can only be generated if it is monetized at some point of the process. In fact, Open Science proponents such as the Royal Society Science Policy Centre are actually very aware of the process of how Open Science is supposed to generate value. Privatization and monetization by turning information into a service is assumed as part of the Open Science concept.

---

[430] Ralph Oliver and Olaf Storbeck, "Allianz to Stop Selling Insurance to Coal Companies," *Financial Times*, 4 May 2018 2018.

Private corporations, such as Exxon, were among the first institutions to recognize and address anthropogenic climate change although they later decided to take a denialist stance toward global warming and poured billions of dollars into lobbying efforts to disavow any kind of responsibility by the fossil fuel industry. Early efforts by Exxon included monitoring and data gathering on CO2 concentrations by attaching sensors to tanker ships travelling the globe. Gathering such global data was crucial for Exxon's business despite the fact that the company became a large 'climate denial' actor upon realizing that a global concerted policy action would threaten the core of their fossil fuel extraction business.[431] Environmental data being available globally would enable private actors without the means of deploying their own monitoring systems to use data to create commercial services based on environmental monitoring infrastructures. Without being able to conclusively assess the ramifications, the creation of a market for climate change impact information services could have a substantial impact on how private and public actors adapt to the impacts of climate change.

Weather reports and forecasts already constitute a vital and daily source of information for economic sectors such as financial investments, mobility services such as aviation and shipping as well as public sector providers such as energy utilities providers. Data from large-scale and long-term environmental monitoring infrastructures could create a market for climate change impact forecasts and trends reports similar to the existing markets for meteorological data. I want to briefly outline some possible impacts of the open availability of environmental data on the sectors of finance, mobility services, and public utility providers.

---

[431] Kaiser and Wasserman, "The Rockefeller Family Fund Vs. Exxon."

One example of a sector affected by the privatization of environmental knowledge are mobility services. A major impact on airline operations in a warming world can be a change in fuel burn due to the changing composition of the atmosphere.[432] Also, high temperatures reduce the lift force of airplane wings, which could lead to additional fuel burn, the need for airplane modifications, or even the impossibility for airlines to take off from certain destinations. Rising temperatures could affect regions that already experience temperature extremes such as the states of the Arab peninsula, where some of larges airline carriers reside. A reduction of capacity, an increase in cost, or even the loss of certain airports as viable aviation hubs could have a major economic impact on the region and the airline industry. In this case, technical adaptation to climate change based on reliable projections of global warming and its impacts on the efficiency of air travel will become an issue of survival for airlines.

Similarly, cargo shipping will be in need for projections of the impacts of climate change. While increases in the number and force of storms may still fall into the realm of meteorologists, projections of ocean currents or local temperature changes could be derived from environmental monitoring data. Higher temperatures could e.g. lead to higher energy costs for refrigerated goods, while changing storm patterns could force cargo ships to divert to longer routes. In any case, those economic actors best able to adjust to the impacts of climate change on a local and global level, based on the data accessible via environmental research infrastructures, will be in the best position to minimize the costs accrued from those impacts. The potential savings from such improved climate change adaptation represent the size of market for environmental monitoring information.

[432] Frédérique Rigal and Herbert Pümpel, "A Brief Overview of Climate Change Impacts on Aviation, Industry Needs, and the Resulting „Home Work" for Scientists.," in *WMO Aeronautical Meteorology Scientific Conference 2017* (Toulouse, France2017).

In sum, how the quest for Open Science is going to play out in practice is still hard to fathom. The ramifications of making data from large-scale environmental research infrastructures available to the public, both laypeople, professional researchers, and businesses, could turn out to look different from what had been envisioned in the grant proposals submitted to the NSF by projects such as OOI and NEON. In the market model, adaptation to climate change by private actors based on risk assessments would dominate – with mitigation at best being downgraded to become a marketable byproduct. The private sector indeed has an existential self-interest in the prevention of loss in revenue due challenging environmental conditions driven by climate change. However, beyond the use of more environmental knowledge to Command and Control changing Anthropocene environments based on a neoliberal agenda, we need to ask how we could use knowledge from environmental research infrastructures to actually broaden the scope of policy actions in the face of climate change and its implications for future generations.

5.5 Can a knowledge infrastructure be an 'honest broker'?

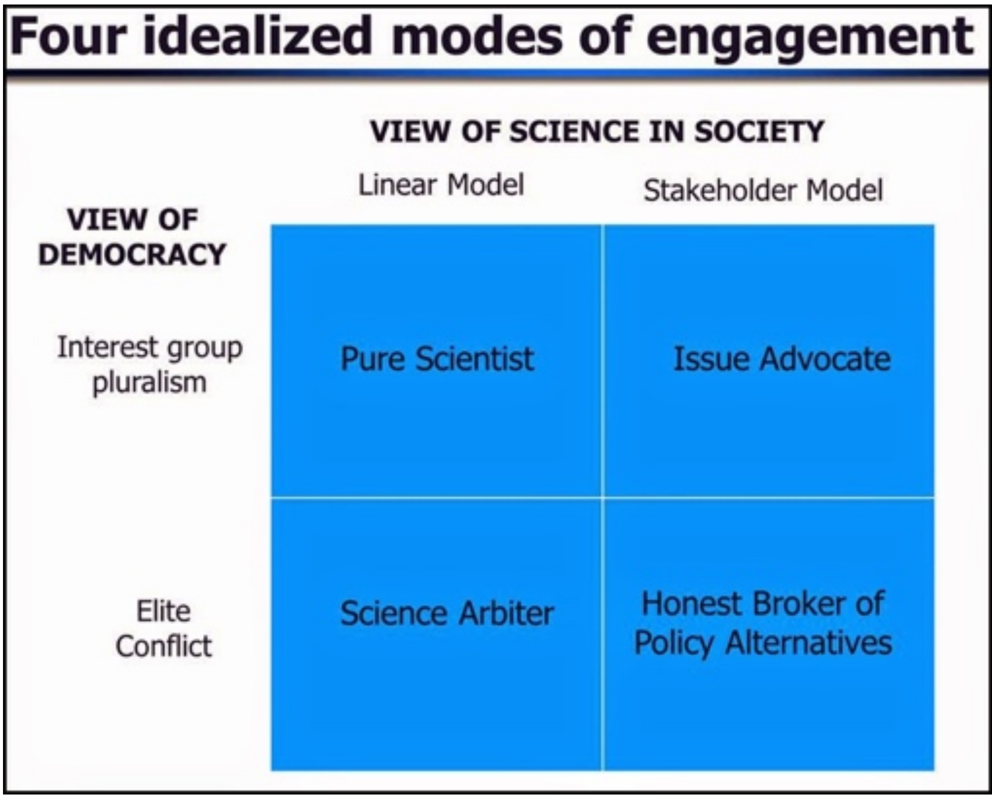
Science in Society

The market model has limited capacity for driving real political change. Its focus is on adaptation, the apolitical self-interest of competing individual actors, and not on building cross-institutional networks and collective capacity around shared political goals. According to the market model data has an inherent and absolute value that stipulates solutions whereas the polis model supports a more critical use of data in the public interest.

Given the market model's singular orientation towards the use of data and information according to the preferences of the private sector, the polis model has much greater potential in achieving the fundamental transformations in policy and public consciousness that data-intensives environmental infrastructures such as OOI and NEON had themselves set out to accomplish.

In the following, we will discuss Roger Pielke's "honest broker" model for science policy advisory arguing that it is reflective with the collaborative approach towards change and reform envisioned by Stone's polis model.

Roger Pielke Jr.'s book *The Honest Broker* has outlined concepts that have become central to the debate around the role of science in democracy. Based on Pielke's scheme of the four roles of a scientist in policy and advisory contexts (see Figure 18), we can ask how a research infrastructure could act as an 'honest broker'. Is there a specific notion or stance of science policy advisory 'built into' a scientific research infrastructure? And how could such an infrastructure be designed in order to aspire to the ideal of the honest broker of policy alternatives within evidence-based political deliberation?



[433]

Figure 18 Four idealized modes of engagement

[433] Figure Source: http://rogerpielkejr.blogspot.de/2015/01/five-modes-of-science-engagement.html

Roger Pielke Jr. has established a matrix of four idealized roles that scientists can play in the process of an evidence-based science advisory process. The two axes of Pielke's schematic represent two different understandings of democracy (interest group pluralism versus elite conflict) and two different views of the role of science in a society (linear model versus stakeholder model). One could regard democracy as the competition of various groups attempting to push individual agendas that they care about or issues they are affected by, as could be the case when environmental groups attempt to lobby against local pollution or for the protection of a specific habitat. Pielke calls this "interest group pluralism", which he contrasts with "elite conflict", the process of institutionalized policy-making, affecting not just individual issues but the set of rules and laws governing the relation between science and other spheres of life.

The second axis concerns the view of science in society. One can regard science as a linear pursuit along a cumulative trajectory of ever-expanding knowledge, building blocks of information being constantly added to the overall body of knowledge. Pielke opposes the "linear model" to the "stakeholder model", in which consensus is established by various stakeholders according to their interests, standpoints, and particular arguments. Pielke then derives the four idealized roles that science or individual scientists could play in the process of science advice: pure scientist, issue advocate, science arbiter, and honest broker of policy alternatives.

First of all, one might object that an inanimate research infrastructure could not act as an advisor in any policy process. Evidence-based advice implies agency, judgment, and reflection which are not properties of a knowledge infrastructure. On the one hand, a scientific infrastructure has no such agency as an individual scientist or expert committee could have. On the other hand, the structure of a knowledge infrastructure could impede the possibility of honest brokerage simply by virtue of its setup, unreliability, or lack of transparency. Thus, a research infrastructure's design could potentially be improved to facilitate the work of scientists aiming to act as honest brokers. In sum, while an infrastructure by itself has no agency, the setup of the research infrastructure should empower its users to act as honest brokers. The central question an honest broker should pose is: "What policy alternatives are consistent and inconsistent with scientific results?"[434] Environmental research infrastructures such as OOI and NEON certainly have a contribution to make to answer this question, yet we need to think about how this contribution can best be integrated into the decision-making of a democratic society.

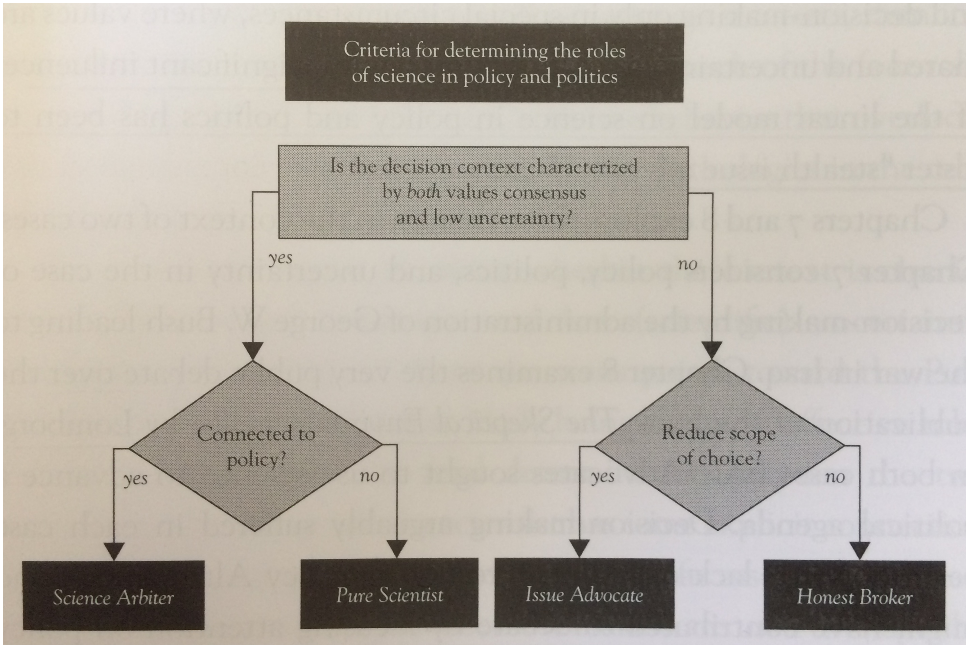


Figure 19 Criteria for determining the role of science in policy and politics[435]

To figure out which role is appropriate for science in a particular question of policy and politics one can consult Pielke's decision tree pictured above (Figure 19). The environmental research infrastructures surveyed in the case studies were built with the explicit goal of addressing and managing the impacts of climate change on the United States environment. Considering issues of climate change mitigation and adaptation in the United States, we can now

---

[434] Pielke, *The Honest Broker. Making Sense of Science in Policy and Politics.* p. 151

[435] Ibid. p. 19.

apply the decision tree to derive the ideal role for science in climate change mitigation policy and then question if and how an environmental research infrastructure could serve this role.

First, we need to consider the level of value consensus and scientific uncertainty concerning the question of climate change mitigation and adaptation. The level of scientific uncertainty about the fact of climate change in general is very low, while the uncertainty about the specific local impacts of a changing climate is higher. Climate 'skeptics' are not actually able to lower this certainty since their stance is actually one of values, i.e. disagreement about whether climate change should be avoided at all. Thus, at least in the United States, there is no value consensus concerning climate change mitigation. If such a value consensus existed in conjunction with the low level of uncertainty on the evidence of climate change, science would merely play the role of a "science arbiter" in climate change mitigation policy.

Since issues of climate change mitigation are inherently related to policy, the role of "pure scientist" is impossible and one should be weary of regarding environmental research infrastructures as advocates of pure science independent of values and uncertainties. Thus, under the condition of a lack of value consensus and fairly low uncertainty, we need to next consider the rightward branch of the decision tree. The next question to ask is whether we aim to broaden or reduce the scope of policy choices. Pielke stresses how problematic it is "to conflate scientific uncertainty with political uncertainty, and then to suggest that a reduction in the former compels a reduction in the latter."[436]

---

[436] Ibid.p. 19

Of course, concerning the environmental research infrastructures regarded in this study, the question of facilitating sound policy advice on climate change mitigation and adaptation without framing the debate beyond other alternative paths of action becomes even more contagious. Pielke points out that "in situations of gridlock, policy-makers frequently need new options, and not more science."[437] Yet, how do we know when we need more policy options rather than more knowledge? This is partly a matter of choice. An individual or a group of scientists can seek to promote specific choices and act as evidence-based "issue advocates". However, a true issue advocate has to argue based on values as well as evidence to support specific policy choices. Issue advocacy pretending to only speak about 'facts' and 'pure science' without making its ethical standpoint explicit turns into what Pielke calls "stealth issue advocacy."[438] Such stealth advocacy, while arguing based on sound evidence, cannot offer compelling arguments on the level of values since it denies any connection to questions of value in the first place.

Environmental research infrastructures could appear to be playing the role of 'Pure Scientist' at first glance. An infrastructure appears indifferent to questions of value, is not directly involved in policy debates, and was built for the explicit purpose of reducing uncertainty by expanding knowledge about environmental change. However, regarding an infrastructure such as the Ocean Observatories Initiative of the National Ecological Observatory Network as 'pure science' actors would be an instance of "stealth issue advocacy".

In 'stealth advocacy' there is no sinister stealthy conspiracy at work, and yet I have shown in the case study chapters that OOI and NEON were conceived with the explicit aim of informing environmental policy, environmental management, and the relating debate on values. When we consult the virtual observatories of OOI and NEON data portals the environmental research infrastructure will never tell me that it was built with a specific purpose.

[437] Ibid.p. 140

[438] Ibid.p. 20

However, the virtual observatory can be a tool to shape the scope of choices in addressing climate change mitigation and its impacts on the environment. If a user wants to reduce the scope of choice and advocate for specific policies, environmental research infrastructures and virtual observatories will give her plenty of information to make her case. Yet, in order to enable this, the infrastructure itself has to create the potential to broaden the scope of policy options.

Thus, an environmental research infrastructure can be an honest broker and a valuable tool for those aiming to act as honest brokers in the debate on climate change mitigation and environmental policy if it can be used to broaden the scope of policy choices. To ensure that an environmental research infrastructure used by citizens and policy-makers can be employed as an honest broker to broaden the scope of choice, we must come up with safeguards against a regression into stealth advocacy.

Firstly, an environmental research infrastructure should be built in a way that has the purpose in mind to empower those that aim to use it as honest broker. Secondly, knowledge about the genesis of the infrastructure itself, i.e. infrastructure inversion, is necessary to make arguments not just about evidence but also about values. A user, be it the lay public, decision-makers, scientists or the private sector, must be aware of how knowledge is generated and maintained to convincingly argue for the scope of policy choices and their assessment. This requires an awareness of the context and frame that the research infrastructure embodies, i.e. it requires a reflection upon what it means to take the perspective of a virtual observatory.

5.6 Recommendations: Environmental Research Infrastructures in the Command and Control Anthropocene

The previous sections of this chapter have problematized the ways in which environmental research infrastructure projects envisioned narratives of environmental risk management, security, and public participation. I had contrasted the market model with the polis model and proposed the honest broker style of science policy making to attempt to envision other ways of using environmental research infrastructures beyond Command and Control of the Anthropocene environment. I will now assess the three activities of a knowledge infrastructure – knowledge generation, knowledge sharing, and knowledge maintenance (compare Infrastructure Matrix in chapter 3) – in light of the polis model and outline the main challenges.

Knowledge Generation

The case studies have shown a repeated conflict between network and systems approaches to designing environmental research infrastructures. Inherent contradictions between a top-down approach by funding agencies and ambitions for more bottom-up participation by individual researchers were never fully resolved. We saw that, on the one hand, *laissez-faire* management led to delays and even financial mismanagement. On the other hand, strict oversight by the National Science Foundation caused a good deal of frustration on part of the individual researchers.

It will be crucial to ask early on who the actual users, both in public and within the science community, are going to be. The process of institutional deliberation in the design of a knowledge infrastructures should be improved with the aim of taking into account the existing structures of the research community as much as individual ideas. In the case of NEON, top-down management may have been necessary, but neither a large funding body such as the NSF nor the ecology community itself was up to the job. The frequency of audits and standards to ensure an efficient use of funds needs to be improved.

In the case of NEON, auditing by the funding organization, the National Science Foundation, was not sufficient since the NSF faced a conflict of interest in preserving the public image of its flagship infrastructure program. To reach this goal, it may be beneficial to cooperate with private sector companies with relevant expertise much earlier on in the planning process of the environmental research infrastructure. By separating auditing from funding, conflicts of interest by funding institutions could be circumvented.

Individuals need to be considered in designing an environmental research infrastructure in a more comprehensive way as well. Those involved in designing research infrastructures (i.e. individual scientists, funding bodies, and political actors) should define the target audience of their data products early on, be it the public, educators, private or political actors or the scientific community. Only then can data products mediated through the research infrastructures' data portal have an impact and be accessible beyond an expert audience. The individuals involved in the planning process of the infrastructures in the case studies were mostly representatives of the most established institutions in their disciplinary fields as well as administrators from funding agencies. Especially since the aim of the infrastructure was to provide a platform for all scientists from related disciplines, more efforts should have been made to involve researchers from other disciplines and less established institutions.

Secondly, laypeople are unlikely to actually utilize the virtual observatories provided by OOI and NEON in the ways that their planners had envisioned. Users are going to have at least some level of expertise, either as private industry experts, educators, or actual scientists themselves. Thus, these groups, especially potential users from private companies and educators and communicators, need to be better integrated in the design process from the start.

Technologies for environmental research and monitoring are likely to evolve rapidly over the coming decades. It is important to build a modular system to allow for openness for innovation. However, cost is an important factor as well as expert personnel to install new technologies, which caused major issues for NEON. Sufficient availability of technical experts to actually set up and operate the infrastructure components is crucial and needs to be taken into account early on.

Knowledge Sharing

Co-operation with local institutions should be a 'two-way street' relationship to enable local stakeholders to actually use the available data that could contribute to climate change adaptation and mitigation efforts. Co-operation between public environmental research infrastructures and private companies drawing value from the available data can be beneficial. However, clear protocols need to be in place to be, on the one hand, as open as possible for innovative ideas to create value from data, and, on the other hand, to ensure that the public has a fair share in profits. To ensure this fair share, models for a valuation of data and means to trace the use of such datasets need to be improved and developed since there is currently no standard in place.

Data sharing platforms and virtual observatories need to be substantially more user friendly and geared toward appropriate target audiences. A data portal is unlikely to be both a useful platform for experts and an appealing public engagement tool all at once. Data Portals and cyberinfrastructure are a major challenge due to their high complexity, cost, and need for expert personnel. Whenever possible, existing and shared structures should be utilized to avoid extra cost for data sharing infrastructure. Co-operation with private companies are also thinkable, although there are currently few examples for such co-operation.[439]

Knowledge Maintenance

Research funding institutions and programs in the United States are built in a way that separates infrastructure construction from the scientific research project to be carried out with this infrastructure. This puts the maintenance of the infrastructure at risk. Funding structures need to address this issue to guarantee long-term maintenance while still fostering competitive funding for research projects. However, there is no 'golden bullet' to securing long-term funding and fostering a competitive funding system at the same time. An unforeseeable breakdown of the infrastructure could also be caused by a breakdown of political support or ideological opposition to the aims a certain knowledge infrastructure is supposed to serve.

In addition, metadata is central to knowledge maintenance. The high cost of maintaining and updating sensor technologies is one issue, maintaining a network architecture yet another. However, maintaining compatibility with future methods and hardware and software standards is going to be both challenging and expensive. Long-term data curation could require new kinds of institutions tasked with data stewardship, metadata standards, and institutional co-operation.

[439] Abad Chabbi, Henry W. Loescher, and Margaux Dillon, "Integrating Environmental Science and the Economy: Innovative Partnerships between the Private Sector and Research Infrastructures," *Frontiers in Environmental Science* 5, no. 49 (2017).

5.7 In Closing: Making Data Terrestrial

There is a key difference in the kinds of infrastructure I have examined in this study in comparison to the kind of climate science infrastructures Paul Edwards describes in *A Vast Machine*. Environmental research infrastructures such as OOI and NEON are, on the one hand, knowledge infrastructures very similar to what Edwards is describing in his process of 'making data global'. On the other hand, these kinds of infrastructures can accomplish something beyond making data global, or national; they create data that is at once global and local, situated at neither end of the dichotomy of situatedness I have discussed above. Thus, I suggest we regard these Fourth Paradigm Science inspired knowledge infrastructure as an example of what one could call 'making data terrestrial'.

In his short treatise *The Terrestrial Manifesto*, Bruno Latour calls for bringing the current political debate about neoliberalism, inequality, and climate change 'down to earth'. Whether we want it or not, Latour believes, we are already part of a new climate regime, a time in which political ideologies have to respond to the question of how to live on a finite and warming planet. The terrestrial territory has started to strike back and participate in history in response to human actions in the Anthropocene.[440]

However, the new climate regime presents us with a wicked dilemma that seems to leave no way out. Latour states, "there is no EARTH corresponding to the endless horizon of the GLOBAL. At the same time, the LOCAL is way to narrow and tiny to contain the multitudes of entities of the terrestrial world."[441] There simply is not enough planet for the global project of endless growth in ever expanding neoliberal societies, while a resurgence of localist

[440] "Was aber soll man tun, wenn das Territorium selbst an der Geschichte teilzunehmen beginnt, Schlag auf Schlag zurückgibt, kurzum: sich mit uns beschäftigt?" Bruno Latour, *Das Terrestrische Manifest* (Frankfurt a.M.: Suhrkamp, 2018). p. 53

[441] "Es gibt keine ERDE die dem endlosen Horizont des GLOBALEN entspricht, zugleich aber ist das LOKALE viel zu eng und zu winzig, als dass es die Mannigfaltigkeit der Wesen der irdischen Welt halten könnte." Ibid. p. 66

isolationism can also neither deny nor address the global issues of the Anthropocene.

Some may hope for a retreat to national boundaries, to retrench, regroup, and address the impacts of a changing climate nationally. In fact, this narrative is present in the discourse around the National Ecological Observatory Network and the image of the environmental research infrastructure as a 'national EKG'. Yet, while the nation state has long been a "vector of modernization", it is "now merely a different name for the LOCAL – and not for an inhabitable world."[442]

Latour's proposal to put both feet back on the ground and start living on a finite planet is to become 'terrestrial'. To become 'terrestrial', Latour writes, "one has to be willing to define the terrain of life as that on which an earth-dweller depends for his survival, and then to ask which other earth-dwellers are dependent on him."[443]

To define terrains of life and ascertain which other earth-dwellers are dependent on us are questions that environmental research infrastructures and virtual observatories could help to answer. My proposal, in conclusion, is not to make data global but to make data terrestrial, with the hope of surpassing the global and the local and drag our associations with other earth-bound entities out of the virtual and digital realm of 'idiocy' into a public realm of terrestrial entities and associations that are bolstered by 'terrestrial data'.

How do we 'make data terrestrial'? It does not simply start with the data, it starts with the intention of tracing a real terrestrial association that I want to connect. Data from virtual observatories can help do this, but they do not themselves accomplish this, as Al Gore may have hoped. Thus, environmental

[442] "Stellte der Nationalstaat lange Zeit über den Vektor der Modernisierung gegenüber den überlieferten Zugehörigkeiten dar, ist er jetzt nur noch ein anderer Name für das LOKALE – und nicht mehr für die bewohnbare Welt." Ibid. p. 116

[443] "Dafür muss man bereit sein, die Lebensterrains als das zu definieren, *wovon ein Erdverbundener für sein Überleben abhängt*, und sich dann zu fragen, *welche anderen Erdverbundenen von ihm abhängig sind.*" Ibid. p. 110

research infrastructures can be of use beyond being a tool of the Command and Control Anthropocene, they could be a point of entry to making data terrestrial. And data is only becoming terrestrial by the real associations arising out of the connections between earth-dwelling entities, which this data can help to illuminate and kindle.

Yet, environmental research infrastructures and Fourth Paradigm Science as such are ambivalent tools, since they are first and foremost a product of the neoliberal and paranoid military logic of the Command and Control Anthropocene. This study urges an elaborate discussion of how virtual observatories and environmental research data could be used not to fuel control fever and earth-systems thinking, but how to "make data terrestrial" and get back "down to earth".

Bibliography:

Abbate, Jane. *Inventing the Internet.* Cambridge, MA: MIT Press, 1999.

Agar, Jon. *The Government Machine. A Revolutionary History of the Computer.* Cambridge, MA: MIT Press, 2003.

American Institute of Biological Sciences. "Plan for Developing and Governing the National Ecological Observatory Network (Neon). Report from the Neon Coordination and Implementation Conference." In *IBRCS White Paper*, edited by AIBS. Washington, DC, 2004.

Anderson, Chris. "The End of Theory: The Data Deluge Makes the Scientific Method Obsolete." *Wired Magazine*, no. 16.07.2008 (2008).

Aronova, Elena, Karen Baker, and Naomi Oreskes. "Big Science and Big Data in Biology: From the International Geophysical Year through the International Biological Program to the Long-Term Ecological Research Program, 1957-Present.". *Historical Studies in the Natural Sciences* 40, no. 2 (2010): 183-224.

Aronova, Elena, Christine von Oertzen, and David Sepkoski, eds. *Data Histories* Vol. 32, Osiris. Chicago: University of Chicago Press, 2017.

Bailis, Peter, Joseph M. Hellerstein, and Michael Stonebraker, eds. *Readings in Database Systems*. Fifth Edition ed, 2015.

Barbrook, Richard, and Andy Cameron. "Californian Ideology." In *Crypto Anarchy, Cyberstates, and Pirate Utopias*, edited by Ludlow, Peter, 373. Cambridge, MA: The MIT Press, 2001.

Barclay, Tom. "Terraserver and the Russian Adventure." *SIGMOD Record* 37, no. 2 (2008): 59-60.

Barclay, Tom, Jim Gray, and Don Slutz. "Microsoft Terraserver: A Spatial Data Warehouse." In *Technical Report*: Microsoft Research, 1999.

Barlow, John Perry. "A Declaration of the Independence of Cyberspace." https://www.eff.org/de/cyberspace-independence.

Beasley, Anthony. "Written Testimony of Dr. Anthony Beasley, Chief Operating Officer and Project Manager, National Ecological Observatory Network (Neon), Inc. Before the United States House of

Representatives Subcommittee on Research and Science Education Hearing Entitled "Nsf Major Research Equipment and Facilities Management: Ensuring Fiscal Responsibility and Accountability"." Washington D.C.: NEON, 2012.

Bell, Gordon. "Bell's Law for the Birth and Death of Computer Classes: A Theory of the Computer's Evolution." San Francisco: Microsoft Research, 2007.

Bellingham, James G. "Collaborative Oceanography and Virtual Experiments. Distribution Statement A.": Monterey Bay Aquarium Research Institute, 2009.

Benson, Etienne. *Wired Wilderness: Technologies of Tracking and the Making of Modern Wildlife.* Baltimore: The Johns Hopkins University Press, 2010.

Bergin, Thomas J., and Thomas Haigh. "The Commercialization of Database Management Systems, 1969-1983." *Annals of the History of Computing* IEEE 31, no. 4 (2009): 26-41.

Best, M., P. Favali, L. Beranzoli, M. Cannat, and N. Cagatay. "Emso: A Distributed Infrastructure for Addressing Geohazards and Global Ocean Change." *Oceanography* 27, no. 2 (2014): 167-69.

"Big Data." *Nature* 455, no. Special Issue (09/03/online 2008).

"Big Data and Natural Disasters: New Approaches for Spatial and Temporal Massive Data Analysis." *Computers & Geosciences* 115, no. Special Issue (2018/06/01/ 2018).

Bolter, Jay David, and Richard Grusin. *Remediation. Understanding New Media.* Cambridge, MA: MIT Press, 2000.

BON. "Final Report: Biodiversity Monitoring Workshop. An National Biodiversity Observation Network." University of Virginia' Blandy Experimental Farm: sponsored by the National Science Foundation, 1998.

———. "Report of the Fourth Workshop on the Biodiversity Observation Network." National Center for Ecological Analysis and Synthesis, Santa Barbara, California: sponsored by the National Science Foundation, 1999.

———. "Report of the Second Workshop on the Biodiversity Observation Network." National Center for Ecological Analysis and Synthesis, Santa Barbara, California: sponsored by the National Science Foundation, 1999.

———. "Report of the Third Workshop on the Biodiversity Observation Network." California Academy of Sciences, San Francisco, California: sponsored by the National Science Foundation, 1999.

Borck, Cornelius. "Big Data. Praktiken Und Theorien Der Datenverarbeitung Im Historischen Querschnitt.". *NTM Zeitschrift für Geschichte der Wissenschaften, Technik und Medizin* 25, no. 4 (2017): 399-406.

Bowker, Geoffrey C. *Memory Practices in the Sciences.* Cambridge, MA: MIT Press, 2005.

Bowker, Geoffrey C., and Karen Baker. "Information Ecology: Open System Environment for Data, Memories, and Knowing." *Intelligent information systems* 29 (2007): 127-44.

Bowker, Geoffrey C., Karen Baker, Florence Millerand, and David Ribes. "Toward Information Infrastructure Studies: Ways of Knowing in a Networked Environment." In *International Handbook of Internet Research*, edited by J. Hunsinger, 97-118. New York: Springer.

Bowker, Geoffrey C., and Susan Leigh Star. *Sorting Things Out. Classification and Its Consequences.* Cambridge, MA: MIT Press, 1999.

boyd, danah. "Privacy and Publicity in the Context of Big Data." In *WWW 2010.* Raleigh, North Carolina, 2010.

boyd, danah, and Kate Crawford. "Critical Questions for Big Data." *Information, Communication & Society* 15, no. 5 (10 May, 2012 2012): 662-79.

Boyle, T.C. *The Terranauts.* Bloomsbury Publishing, 2016.

Burch, Patricia. "The Professionalization of Instructional Leadership in the United States: Competing Values and Current Tensions." *Journal of Education Policy* 22, no. 2 (2007/03/01 2007): 195-214.

Bush, Vannevar. "As We May Think." *The Atlantic*, July 1945.

Camprubí, Lino. "The Sonic Construction of the Ocean as the Navy's Operating Environment." In *Navigating Noise*, edited by N. v. Dijk, K. Ergenzinger, C. Kassung and S. Schwesinger, 219-45. Köln: Walther König, 2017.

Carnes, Donna. "Ode to a Sailor." *SIGMOD Record* 37, no. 2 (2008): 16-18.

Carson, Rachel. *The Sea around Us.* Oxford University Press, 1989.

Ceruzzi, Paul E. *A History of Modern Computing.* Cambridge, MA: The MIT Press, 1998.

Cesare, Chris. "Ecology Aims High." *Nature* 529, no. 21 January 2016 (2016): 274-76.

Chabbi, Abad, Henry W. Loescher, and Margaux Dillon. "Integrating Environmental Science and the Economy: Innovative Partnerships between the Private Sector and Research Infrastructures." *Frontiers in Environmental Science* 5, no. 49 (2017).

Christensen, Clayton M. *The Innovator's Dilemma. The Revolutionary Book That Will Change the Way You Do Business.* New York: Harper Business, 2000.

Clemson, Gaye I. *Tandem Computers Unplugged. A People's History.* Fast Pencil Inc., 2012.

"Closing the Climategate." *Nature*, 18 November 2010 2010, 345.

Codd, E.F. "A Relational Model of Data for Large Shared Data Banks." *Communications of the ACM* 13, no. 6 (1970): 377-87.

Committee, House of Commons Science and Technology. "The Reviews into the University of East Anglia's Climatic Research Unit's E-Mails." edited by Government Office for Science. London, 2011.

Committee on Innovations in Computing and Communications: Lessons from History, National Research Council. "Funding a Revolution: Government Support for Computing Research." Washington, D.C.: National Research Council, 1999.

Cornell, Sarah E., I. Colin Prentice, Joanna I. House, and Catherine J. Downy, eds. *Understanding the Earth System. Global Change Science for Application.* Cambridge: Cambridge University Press, 2012.

Cosgrove, Denis. *Apollo's Eye: A Cartography of the Earth in the Western Imagination.* Baltimore: Johns Hopkins University Press, 2001.

Crawford, Kate. "Think Again: Big Data. Why the Rise of Machines Isn't All It's Cracked up to Be." *Foreign Policy*, 10 May 2013 2013.

Crutzen, Paul J. "Geology of Mankind." *Nature* 415 (31 January 2002 2002): 23.

Crutzen, Paul J., and Eugene F. Stoermer. "The Anthropocene." *IGBP Newsletter* 41 (2000): 17-18.

Cullen, V. *Down to the Sea for Science: 75 Years of Ocean Research, Education, and Exploration at the Woods Hole Oceanographic Institution.* Woods Hole Oceanographic Institution, 2005.

Cushman, Gregory T. "Choosing between Centers of Action." In *The Machine in Neptune's Garden: Historical Perspectives on Technology and the Marine Environment*, edited by Helen M. Rozwadowski and David K. van Keuren. Sagamor Beach: Science History Publications/USA, 2004.

Daly, Kendra. "The Ocean Observatories Initiative and Orion Organization. Ooi Conceptual Design Review." Moss Landing, CA, 2006.

Daston, Lorraine, ed. *Science in the Archives. Pasts, Presents, Futures.* Chicago: The University of Chicago Press, 2017.

"The Data Deluge. And How to Handle It: A 14-Page Special Report." *The Economist*,, February 27 2010.

Davenport, Thomas H. *Big Data @ Work. Dispelling the Myths, Uncovering the Opportunities.* Boston: Harvard Business Review Press, 2014.

———. *Enterprise Analytics : Optimize Performance, Process and Decisions through Big Data.* Upper Saddle River, N.J.: Financial Times/Prentice Hall, 2013.

Davenport, Thomas H., and Jeanne G. Harris. *Competing on Analytics. The New Science of Winning.* Boston, MA: Harvard Business Review Press, 2007.

Dawson, Todd, Serita Frey, Dave Schimel, and Susan Stafford. "Neon Scope Management Recommendations." NEON, 2015.

Delgado, J.P., and C. Cussler. *Silent Killers: Submarines and Underwater Warfare.* Bloomsbury Publishing, 2011.

Delibašić, B., J.E. Hernández, J. Papathanasiou, F. Dargam, P. Zaraté, R. Ribeiro, S. Liu, and I. Linden. *Decision Support Systems V – Big Data Analytics for Decision Making: First International Conference, Icdsst 2015, Belgrade, Serbia, May 27-29, 2015, Proceedings.* Springer International Publishing, 2015.

Derosières, Alain. *Die Politik Der Großen Zahlen. Eine Geschichte Der Statistischen Denkweise.* Berlin: Springer, 2005.

DeWitt, David J., and Charles Levine. "Not Just Correct, but Correct and Fast. A Look at Jim Gray's Contributions to Database System Performance." *SIGMOD Record* 37, no. 2 (2008): 45-49.

Dickson, P. *Sputnik: The Shock of the Century.* Walker, 2001.

Diebold, Francis X. "A Personal Perspective on the Origin(S) and Development of "Big Data": The Phenomenon, the Term, and the Discipline." 2012.

Dionne, E. J. . "Greening of Democrats: An 80's Mix of Idealism and Shrewd Politics." *The New York Times*, June 14, 1989 1989.

Drake, N. "Underwater Volcano Offers Rare Look at Eruption in Real Time." *National Geographic*, 15 December 2016 2016.

Dyson, George. *Turing's Cathedral. The Origins of the Digital Universe.* London: Penguin, 2012.

Edwards, Paul N. *The Closed World. Computers and the Politics of Discourse in Cold War America.* Cambridge, MA: MIT Press, 1996.

———. *A Vast Machine. Computer Models, Climate Data, and the Politics of Global Warming.* Cambridge, MA: MIT Press, 2010.

Edwards, Paul N., Steven J. Jackson, Melissa K. Chalmers, Geoffrey C. Bowker, Christine L. Borgman, David Ribes, Matt Burton, and Scout Calvert. "Knowledge Infrastructures: Intellectual Frameworks and Research Challenges. Report of a Workshop Sponsored by the National Science Foundation and the Sloan Foundation, 25-28 May 2012." **University of Michigan School of Information**, 2012.

Ezrahi, Yaron. "Science and the Political Imagination in Contemporary Democracies." In *States of Knowledge. The Co-Production of Science and the Social Order*, edited by Sheila Jasanoff, 254-73. London: Routledge, 2004.

Foer, F. *World without Mind: The Existential Threat of Big Tech.* Penguin Publishing Group, 2017.

Friedman, Thomas L. *Thank You for Being Late. An Optimist's Guide to Thriving in the Age of Accelerations.* New York: Farrar, Straus and Giroux, 2016.

Fuller, R.B. *Operating Manual for Spaceship Earth.* Lars Muller Publishers, 2008.

Garcia-Molina, Hector, Jeffrey D. Ullman, and Jennifer Widom. *Database Systems: The Complete Book.* New International Edition of Second Revised Edition ed. New York: Pearson Education, 2013.

Gartner. "Hype Cycle for Emerging Technologies, 2008." Gartner, Inc., 2008.

———. "Hype Cycle for Emerging Technologies, 2009." Gartner, Inc., 2009.

———. "Hype Cycle for Emerging Technologies, 2011." Gartner, Inc., 2011.

———. "Hype Cycle for Emerging Technologies, 2012." Gartner, Inc., 2012.

———. "Hype Cycle for Emerging Technologies, 2013." Gartner, Inc., 2013.

———. "Hype Cycle for Emerging Technologies, 2014." Gartner, Inc., 2014.

Gehl, Robert W. "Sharing, Knowledge Management and Big Data: A Partial Genealogy of the Data Scientist." *European Journal of Cultural Studies* 18, no. 4-5 (2015): 413-28.

Gitelman, Lisa. *"Raw Data" Is an Oxymoron.* Cambridge, MA: MIT Press, 2013.

"Global Warming of 1,5 °C. An Ipcc Special Report on the Impacts of Global Warming of 1.5 °C above Preindustrial Levels and Related Global Greenhouse Gas Emission Pathways, in the Context of Strengthening the Global Response to the Threat of Climate Change, Sustainable Development, and Efforts to Eradicate Poverty." Intergovernmental Panel on Climate Change, 2018.

Gore, Al. "The Digital Earth: Understanding Our Planet in the 21st Century." California Science Center, Los Angeles, 1998.

———. *Earth in the Balance. Forging a New Common Purpose.* London: Earthscan, 2007.

Gram, W. "National Ecological Observatory Netowork. Education and Public Engagement." NEON, 2009.

Gramelsberger, Gabriele. "Big Data-Revolution Oder Datenhybris? Überlegungen Zum Datenpositivismus Der Molekularbiologie." *NTM Zeitschrift für Geschichte der Wissenschaften, Technik und Medizin* 25, no. 4 (2017): 459-83.

———. *Computerexperimente. Wandel Der Wissenschaft Im Zeitalter Des Computers.* Science Studies. Bielefeld: Transcript, 2010.

———, ed. *From Science to Computational Studies. Studies in the History of Computing and Its Influence on Today's Sciences.* Zürich: diaphanes, 2011.

Gray, Jim. "A Critique of Ibm's Computer Science Research." edited by IBM Research. San Jose: IBM Research, 1980.

———. "Data Management: Past, Present, and Future." *IEEE Computer* 29, no. 10 (1996): 38-46.

———. "How Do You Know?" In *http://jimgray.azurewebsites.net/default.htm*, edited by Microsoft Research, 2: Microsoft Research, 2003.

———. "A Measure of Transaction Processing Power." *Datamation* (April 1, 1985 1985).

———. "Mip Envy: A Programming Complex." edited by IBM Research. San Jose: IBM Research, 1980.

———. "Tape Is Dead, Disk Is Tape, Flash Is Disk, Ram Locality Is King." Microsoft Research, 2006.

———. "What's Next? A Dozen Information-Technology Research Goals." In *Technical Report*: Microsoft Research, 1999.

Gray, Jim, Bob Good, Dieter Gawlick, Pete Homan, and Harald Sammer. "One Thousand Transactions Per Second." Paper presented at the Proceedings of IEEE Compcon-85, San Francisco, 1984.

Gray, Jim, and Andreas Reuter. *Transaction Processing: Concepts and Techniques.* The Morgan Kaufmann Series in Data Management Systems. San Francsico, California: Morgan Kaufmann Publishers, 1992.

Grevsmühl, Sebastian Vincent. *La Terre Vue D'en Haut. L'invention De L'environnement Global.* Paris: Éditions du Seuil, 2014.

Grochow, Keith. "The Design of Cove: A Collaborative Ocean Visualization Environment ", University of Washington, 2011.

Guldi, J., and D. Armitage. *The History Manifesto.* Cambridge University Press, 2014.

Hafner, Katie. "Silicon Valley's High-Tech Hunt for Colleague." *The New York Times*, February 3 2007.

Haigh, Thomas. "How Data Gots Its Base: Information Storage Software in the 1950s and 1960s." *Annals of the History of Computing* IEEE 31, no. 4 (2009): 6-25.

Hanwahr, Nils C. "Marine Animal Satellite Tags." In *Future Remains. A Cabinet of Curiosities for the Anthropocene*, edited by Gregg Mitman, Marco Armiero and Robert S. Emmett. Chicago: University of Chicago Press, 2018.

———. "„Mr Database“. Jim Gray and the History of Database Technology." *NTM Zeitschrift für Geschichte der Wissenschaften, Technik und Medizin* 25, no. 4 (2017): 519-42.

Harari, Yuval Noah. "Dataism Is Our New God." *New Perspectives Quarterly* 34, no. 2 (2017): 36-43.

Haraway, Donna. "Situated Knowledges: The Science Question in Feminism and the Privilege of Partial Perspective." *Feminist Studies* 14, no. 3 (1988): 575-99.

Harrison, Michael A. "Jim Gray at Berkeley." *SIGMOD Record* 37, no. 2 (2008): 26-27.

Heise, Ursula K. *Nach Der Natur. Das Artensterben Und Die Moderne Kultur.* Berlin: Suhrkamp Verlag, 2010.

———. *Sense of Place and Sense of Planet: The Environmental Imagination of the Global.* New York: Oxford University Press, 2008.

Helland, Pat. "Knowledge and Wisdom." *SIGMOD Record* 37, no. 2 (2008): 28-29.

Hellerstein, Joseph M., and David L. Tennenhouse. "Searching for Jim Gray: A Technical Overview." *Communications of the ACM* 54, no. 7 (2011): 77-89.

Helmreich, Stefan. *Alien Ocean: Anthropological Voyages in a Microbial Sea.* Berkeley: University of California Press, 2009.

Henke, Nicolaus, Jacques Bughin, and Michael Chui. "The Age of Analytics: Competing in a Data-Driven World." McKinsey Global Institute, 2016.

Hey, Tony, Stewart Tansley, and Kristin Tolle, eds. *The Fourth Paradigm. Data-Intensive Scientific Discovery.* Redmont, Washington: Microsoft Research, 2009.

Hoffman, S., and A. Podgurski. "Big Bad Data: Law, Public Health, and Biomedical Databases." [In eng]. *J Law Med Ethics* 41 Suppl 1 (Mar 2013): 56-60.

Höhler, S. *Spaceship Earth in the Environmental Age, 1960-1990.* Taylor & Francis, 2015.

Howarth, Richard B., and Richard B. Norgaard. "Environmental Valuation under Sustainable Development." *The American Economic Review* 82, no. 2 (1992): 473-77.

Hu, Tung-Hui. *A Prehistory of the Cloud.* Cambridge, MA: The MIT Press, 2015.

Hutchins, Edwin. *Cognition in the Wild.* Cambridge, MA: The MIT Press, 1995.

Institute, Portola. *The Last Whole Earth Catalog: Access to Tools.* Portola Institute, 1971.

Jordan, Stephen J., Sharon E. Hayes, David Yoskowitz, Lisa M. Smith, J. Kevin Summers, Marc Russell, and William H. Benson. "Accounting for Natural Resources and Environmental Sustainability: Linking Ecosystem Services to Human Well-Being." *Environmental Science & Technology* 44, no. 5 (2010/03/01 2010): 1530-36.

Kaiser, David, and Lee Wasserman. "The Rockefeller Family Fund Vs. Exxon." *The New York Review of Books*, 8 December 2016 2016.

Kaplan, S. "A Massive Underwater Volcanic Eruption Is Captured in Real Time." *Washington Post*, 16 December 2016 2016.

Kay, L.E. *Who Wrote the Book of Life?: A History of the Genetic Code.* Stanford University Press, 2000.

Kormann, Carolyn. "The Dire Warnings of the United Nations' Latest Climate-Change Report." *The New Yorker*, 8 October 2018 2018.

Lanier, Jaron. *Who Owns the Future?* London: Allen Lane, 2013.

Latour, B. *Das Parlament Der Dinge: Für Eine Politische Ökologie.* Suhrkamp, 2010.

Latour, Bruno. *Das Terrestrische Manifest.* Frankfurt a.M.: Suhrkamp, 2018.

———. *Existenzweisen: Eine Anthropologie Der Modernen.* Frankfurt am Main: Suhrkamp, 2014.

———. *Face À Gaïa. Huit Conférences Sur Le Nouveau Régime Climatique.* Paris: Éditions La Découverte, 2015.

———. *Politics of Nature. How to Bring the Sciences into Democracy.* Translated by Catherine Porter. Cambridge, MA: Harvard Univerity Press, 2004.

Lawrence, David M. *Upheaval from the Abyss: Ocean Floor Mapping and the Earth Science Revolution.* New Brunswick: Rutgers University Press, 2002.

Leigh Star, Susan, and Karen Ruhleder. "Steps toward an Ecology of Infrastructure: Design and Access for Large Information Spaces." *Information Systems Research* 7, no. 1 (1996): 111-36.

Leonelli, Sabina. *Data-Centric Biology. A Philosophical Study.* Chicago: University of Chicago Press, 2016.

———. *La Ricerca Scientifica Nell'era Dei Big Data.* Rome: Meltemi, 2018.

Lerner, Allison C. "Statement of Allison C. Lerner, Inspector General National Sciece Foundation before a Hearing of the Committee on Science, Space, and Technology." edited by United States House of Representatives. Washington, D.C., 2014.

Lindsay, Bruce G. "Jim Gray at Ibm. The Transaction Processing Revolution." *SIGMOD Record* 37, no. 2 (2008): 38-40.

Lohr, Steve. "How Data Became So Big." *The New York Times*, 11 August 2012 2012.

———. "The Origins of 'Big Data': An Etymological Detective Story." *The New York Times*, 1 February 2013 2013.

MacCormick, John. *9 Algorithms That Changed the Future: The Ingenious Ideas That Drive Today's Computers.* Princeton: Princeton University Press, 2012.

Maier, Dave, Eric Landis, Judy Cushing, Anne Frondorf, Avi Silberschatz, and John Schnase. "Research Directions in Biodiversity and Ecosystem Informatics. Report of an Nsf, Usgs, Nasa Workshop on Biodiversity and Ecosystem Informatics Held at Nasa Goddard Space Fight Center, June 22-23, 2000." 2000.

Manyika, James, Michael Chui, and Brad Brown. "Big Data: The Next Frontier for Innovation, Competition, and Productivity." McKinsey Global Institute, 2011.

Markoff, John. "In Computing, Weighing Sheer Power against Vast Pools of Data." *The New York Times*, 2 June 2003 2003.

Matthews, Christine M. "U.S. National Science Foundation: Major Research Equipment and Facility Construction." In *Report for Congress*, edited by Congressional Research Service, 2012.

Mayer-Schönberger, Viktor, and Kenneth Cukier. *Big Data. A Revolution That Will Transform How We Live, Work and Think.* London: John Murray, 2013.

McAfee, A., and E. Brynjolfsson. "Big Data: The Management Revolution." [In eng]. *Harv Bus Rev* 90, no. 10 (Oct 2012): 60-6, 68, 128.

McGill, Joshua Kirk. "Disclosure Pursuant to Allegations Related to Suspected: Fraud, Gross Mismanagement, Gross Waste of Funds, Abuse of Authority Etc." edited by United States Congress, 179. Washington D.C., 2015.

McKibben, Bill. *The End of Nature.* New York: Random House Trade Paperbacks, 2006.

Mervis, J. "Ecology's Megaproblem. Fledgling National Observing Network Faces Harsh Realities." *Science*, 24 September 2015 2015.

———. "Neon Contractor Hanging by a Thread, Nsf Tells Congress." *Science*, 21 September 2015 2015.

———. "Nsf Fires Managers of Troubled Neon Ecology Project." *Science*, 11 December 2015 2015.

———. "Nsf Picks Battelle to Run Neon." *Science*, 14 March 2016 2016.

Miller, S. *Dual Use Science and Technology, Ethics and Weapons of Mass Destruction.* Springer International Publishing, 2018.

Mills, William J. "Metaphorical Vision: Changes in Western Attitudes to the Environment." *Annals of the Association of American Geographers* 72, no. 2 (1982): 237-53.

Mitman, Gregg. "When Nature Is the Zoo: Vision and Power in the Art and Science of Natural History." *Osiris* 11 (1996): 117-43.

Morozov, Evgeny. *To Save Everythink, Click Here. The Folly of Technological Solutionism.* New York: Public Affairs, 2013.

Morton, Timothy. *Ecology without Nature: Rethinking Environmental Aesthetics.* Cambridge: Harvard University Press, 2009.

Nagel, Thomas. *The View from Nowhere.* New York ; Oxford: Oxford University Press, 1986.

Nash, Roderick F. *Wilderness and the American Mind.* New Haven: Yale University Press, 1967.

National Research Council. *Neon: Addressing the Nation's Environmental Challenges.* Washington D.C.: The National Academies Press, 2004.

National Research Council, Ocean Studies Board. "Enabling Ocean Research in the 21st Century: Implementation of a Network of Ocean Observatories." edited by National Academy Press. Washington, DC, 2003.

———. "Illuminating the Hidden Planet: The Future of Seafloor Observatory Science." edited by National Academy Press. Washington, DC, 2000.

National Science Foundation. "National Ecological Observatory Network: Revolutionizing Ecological Research ": National Science Foundation, 2013.

National Science Foundation, Decadal Commission. "Ocean Sciences at the New Millennium." edited by Inc. Geosciences Professional Services, 2001.

Nauman, John. "Jim Gray's Tandem Contributions." *SIGMOD Record* 37, no. 2 (2008): 41-44.

NEON. "Biodiversity, Species Composition, and Ecosystem Functioning." Carmel Valley, CA, 2004.

———. "Final Report of the Neon-V: Cripton Workshop. Collections, Research, Inventories, and People for Taxonomic Opportunities in Neon." Field Museum of Natural History, Chicago, 2002.

———. "Integrated Science and Education Plan for the National Ecological Observatory Network." 2006.

———. "Neon, Inc. Announces the Selection of the National Core Sites and Research Question Designs." 2007.

———. "Report on First Workshop on the National Ecological Observatory Network (Neon)." Archbold Biological Station, Lake Placid, Florida: Sponsored by the National Science Foundation, 2000.

———. "Report to the National Science Foundation from the Fourth Workshop on the Development of a National Ecological Observatory Network (Neon): Standard Measurements and Infrastructure Needs." The Millennium Hotel, Boulder, Colorado, 2002.

———. "Report to the National Science Foundation from the Second Workshop on the Development of a National Ecological Observatory Network (Neon)." San Diego Supercomputer Center, La Jolla, California, 2000.

———. "Report to the National Science Foundation from the Sixth Workshop on the Development of a National Ecological Observatory Network (Neon): Information Management." National Center for Ecological Analysis and Synthesis, University of California Santa Barbara, Santa Barbara, CA, 2002.

———. "A Workshop on Modelling in Neon." Woods Hole, MA: The Ecosystems Center, Marine Biological Laboratory, 2005.

"Neon Warning Signs: Examining the Management of the National Ecological Observatory Network.". Washington D.C.: U.S. House of Representatives Committee on Science, Space, and Technology, 2015.

Nicolet, M. *The International Geophysical Year Meetings: Annals of the International Geophysical Year*. Elsevier Science, 2013.

Nielsen, Michael. *Reinventing Discovery. The New Era of Networked Science.* Princeton: Princeton University Press, 2011.

Nixon, Rob. "The Anthropocene. The Promise and Pitfalls of an Epochal Idea." In *Future Remains. A Cabinet of Curiosities for the Anthropocene*, edited by Gregg Mitman, Marco Armiero and Robert S. Emmett, 1-20. Chicago: The University of Chicago Press, 2018.

———. *Slow Violence and the Environmentalism of the Poor*. Harvard University Press, 2011.

Nooner, Scott L., and William W. Chadwick Jr. "Inflation-Predictable Behavior and Co-Eruption Deformation at Acial Seamount." *Science* 354, no. 6318 (2016): 1399-403.

Nye, D.E. *American Technological Sublime.* MIT Press, 1996.

O'Neil, C. *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy.* Crown/Archetype, 2016.

Oceanography, Scripps Institution of. *Celebrating 100 Years of Oceanography at Scripps Institution of Oceanography.* Oceanography Society, 2003.

Oldham, C. *Underway on Nuclear Power: 50th Anniversary of Uss Nautilus.* Faircount LLC, 2004.

Oliver, Ralph, and Olaf Storbeck. "Allianz to Stop Selling Insurance to Coal Companies." *Financial Times*, 4 May 2018 2018.

OOI. "Final Network Design." edited by Consortium for Ocean Leadership. Washington, DC, 2011.

———. "National Science Foundation Panel Report on the Final Design Review of the Ocean Observatories Initiative (Ooi) ": National Science Foundation, 2008.

———. "Scientific Objectives and Network Design: A Closer Look." 2007.

Oreskes, Naomi, and John Krige, eds. *Science and Technology in the Global Cold War*. Cambridge, MA: The MIT Press, 2014.

ORION Executive Steering Committee. "Ocean Observatories Initiative Science Plan." Washington, D.C., 2005.

Patil, Dhanurjay. *Building Data Science Teams.* Sebastopol, CA O'Reilly Media, 2011.

Philip Chen, C. L., and Chun-Yang Zhang. "Data-Intensive Applications, Challenges, Techniques and Technologies: A Survey on Big Data." *Information Sciences* 275 (2014/08/10/ 2014): 314-47.

Pielke, Roger A. Jr. *The Honest Broker. Making Sense of Science in Policy and Politics.* Cambridge: Cambridge University Press, 2007.

Plattner, Hasso. "Oral History of Hasso Plattner." By John Hollar. *Computer History Museum* (2012).

Policy, United States. Commission on Ocean. *An Ocean Blueprint for the 21st Century: Final Report.* U.S. Commission on Ocean Policy, 2004.

Porter, Theodore M. *Trust in Numbers. The Pursuit of Objectivity in Science and Public Life.* Princeton: Princeton University Press, 1995.

Pschera, Alexander. *Vom Schweben. Romantik Im Digitalen.* Berlin: Matthes & Seitz, 2013.

Purdy, J. *After Nature: A Politics for the Anthropocene.* Harvard University Press, 2015.

Rainger, Ronald. "‚a Wonderful Oceanographic Tool'. The Atomic Bomb, Radioactivity and the Development of American Oceanography." In *The Machine in Neptune's Garden: Historical Perspectives on Technology and the Marine Environment*, edited by Helen M. Rozwadowski and David K. van Keuren. Sagamore Beach, MA: Science History Publications/USA, 2004.

"Report of the Ioc Blue Ribbon Panel for a Global Ocean Observing System (Goos). The Case for Goos." Intergovernmental Oceanographic Commission (of UNESCO), 1993.

"Responding to Anthrax Attacks." *The New York Times*, 16 October 2001 2001.

Reuter, Andreas. "Is There Life Outside Transactions? Writing the Transaction Processing Book." *SIGMOD Record* 37, no. 2 (2008): 54-58.

Richtel, Matt. "Huge Microsoft Photo File Is Part of a Bigger Picture." *The New York Times*, June 25, 1998 1998.

Rid, T. *Rise of the Machines: The Lost History of Cybernetics.* Scribe Publications Pty Limited, 2016.

"Riding the Wave. How Europe Can Gain from the Rising Tide of Scientific Data." In *High Level Expert Group on Scientific Data*: European Commission, 2010.

Rigal, Frédérique, and Herbert Pümpel. "A Brief Overview of Climate Change Impacts on Aviation, Industry Needs, and the Resulting „Home Work" for Scientists." In *WMO Aeronautical Meteorology Scientific Conference 2017*. Toulouse, France, 2017.

Rosenberg, Daniel. "Data before the Fact." In *„Raw Data" Is and Oxymoron*, edited by Lisa Gitelman, 15-40. Cambridge, MA: MIT Press, 2013.

Rowley, Jennifer. "The Wisdom Hierarchy: Representations of the Dikw Hierarchy." *Journal of Information Science* 33, no. 2 (2007): 163-80.

Royal Society. "Observing the Earth. Expert Views on Environmental Observation for the Uk." London, 2015.

Rozwadowski, Helen M. *Fathoming the Ocean: The Discovery and Exploration of the Deep Sea.* Cambridge: Harvard University Press, 2005.

Schlosser, Eric. *Command and Control. Nuclear Weapons, the Damascus Accident, and the Illusion of Safety.* New York: Penguin Books, 2013.

"Science as an Open Enterprise." London: Royal Society, 2012.

Sciences, American Institute of Biological. "Ecological Impacts of Climate Change: Report from a Neon Science Workshop." Washington D.C.: AIBS, 2004.

Silberman, Steve. "Inside the High Tech Hunt for a Missing Silicon Valley Legend." *Wired*, July 24 2007.

Silberschatz, Avi, Michael Stonebraker, and Jeffrey D. Ullman. "Database Systems: Achievements and Opportunities." *ACM Sigmod Record* 19, no. 4 (1990): 6-22.

Stone, Deborah. *Policy Paradox. The Art of Political Decision Making. Third Edition.* New York: W.W. Norton & Company, 2011.

Stonebraker, Michael. "Oral History of Michael Stonebraker." By Burton Grad (23 August 2007 2007).

Strasser, Bruno J. *Collecting Experiments. Making Big Data Biology.* Chicago: University of Chicago Press, 2019.

———. "The Experimenter's Museum: Genbank, Natural History, and the Moral Economies of Biomedicine.". *Isis* 102 (March 2011 2011): 60-96.

Strasser, Bruno J., and Paul N. Edwards. "Big Data Is the Answer ... but What Is the Question?". *Osiris* 32, no. 1 (2017): 328-45.

Tawa, R. "National Ecological Observatory Network Cyber Infrastructure (Ci)." NEON, 2009.

The Economist. "Visiting Neptune's Kingdom. Outer Space Has Hogged the Limelight for Too Long. Welcome to Inner Space." *The Economist*, 15 November 2007 2007.

"Tribute to Honor Jim Gray." *ACM Sigmod Record* 37, no. 2 (2008).

Tucker, J.B., R. Danzig, L.P. Knowles, G.L. Epstein, C. Jefferson, F. Lentzos, P. Silver, *et al. Innovation, Dual Use, and Security: Managing the Risks of Emerging Biological and Chemical Technologies.* MIT Press, 2012.

Turner, Fred. *From Counterculture to Cyberculture. Stewart Brand, the Whole Earth Network, and the Rise of Digital Utopianism.* Chicago: The University of Chicago Press, 2006.

Vaskevitch, David. "Jim Gray: His Contribution to Industry." *SIGMOD Record* 37, no. 2 (2008): 35-36.

Vermeulen, Niki. "Big Biology: Supersizing Science During the Emergence of the 21st Century." *NTM* 24, no. 2 (06/28/nihms-submitted 07/10/pmc-release 2016): 195-223.

Verne, Jules. *20,000 Leagues under the Seas.* Translated by William Butcher. Oxford: Oxford University Press, 1998.

von Willemoes-Suhm, R. *Die Challenger-Expedition: Zum Tiefsten Punkt Der Weltmeere : 1872-1876.* Edition Erdmann, in der Verlagshaus Römerweg GmbH, 2015.

Weinberger, David. *Too Big to Know. Rethinking Knowledge Now That the Facts Aren't the Facts, Experts Are Everywhere, and the Smartest Person in the Room Is the Room.* New York: Basic Books, 2011.

Wiener, N. *Cybernetics or Control and Communication in the Animal and the Machine.* M.I.T. Press, 1961.

Wilkinson, Alec. "Remember This? A Project to Record Everything We Do in Life." *The New Yorker*, May 28 2007.

Willinsky, John. "The Unacknowledged Convergence of Open Source, Open Access, and Open Science." *2005* (2005-08-01 2005).

Witze, Alexandra. "Massive Ocean-Observing Project Launches — Despite Turmoil. Network of Deep-Water Observatories Streams Data in Real Time.". *Nature* 534, no. 7606 (2016): 159–60.

———. "Ocean Observatory Project Hits Rough Water." *Nature*, 27 November 2014 2014.

———. " Us Ocean Sciences Told to Steer a New Course." *Nature* 517, no. 7536 (2015): 538–39.

Zachary, G. Pascal. *Endless Frontier: Vannevar Bush, Engineer of the American Century.* New York: The Free Press, 1997.