Article:

Bates, J., Goodale, P., Lin, Y.-W. et al. (2019) Assembling an infrastructure for historic climate data recovery: data friction in practice. Journal of Documentation, 75 (4). pp. 791-806. ISSN: 0022-0418

https://doi.org/10.1108/JD-08-2018-0130

Assembling an infrastructure for historic climate data recovery: data friction in practice

Journal:	*Journal of Documentation*
Manuscript ID	JD-08-2018-0130.R1
Manuscript Type:	Article
Keywords:	Data Friction, Data Assemblage, Data Infrastructure, Climate data, Historical Records, Critical Data Studies

Assembling an infrastructure for historic climate data recovery: data friction in practice

Abstract

Purpose

We adopt an assemblage theory lens to examine the socio-material forces shaping the development of an infrastructure for the recovery of archived historical marine weather records for use in contemporary climate datasets.

Design/methodology/approach

We adopted a Data Journeys approach to research design, conducting in-depth semi-structured interviews with climate scientists, citizen scientists, and a climate historian who were engaged at key sites across the journey of data from historical record to the ICOADS database. Interview data were complemented by further qualitative data collected via observations of working practices, a digital ethnography of citizen scientists' online forums, and documentation relevant to the circulation and governance of climate data across emergent data infrastructures. Data were thematically analysed (Ryan and Bernard, 2003), with themes being informed primarily by the theoretical framework.

Findings

We identify and critically examine key points of friction in the constitution of the data recovery infrastructure and the circulation of data through it, and identify the reflexive and adaptive nature of the beliefs and practices fostered by influential actors within the assemblage in order to progress efforts to build an infrastructure despite significant challenges. We conclude by addressing possible limitations of some of these adaptive practices within the context of the early twenty-first century neoliberal state, and in light of current debates about data justice.

Originality/value

The paper draws upon original empirical data and a novel theoretical framework that draws together Deleuze and Guattari's assemblage theory with key concepts from the field of Critical Data Studies (Data Journeys, Data Friction, Data Assemblage) to illuminate the socio-material constitution of the data recovery infrastructure within the context of the early twenty-first century neoliberal state.

Key words: Data Friction; data assemblage; data infrastructure; climate data; historical records.

Introduction

> "Climatology requires long-term data from many locations, consistent across both space and time. This requirement implies a lengthy chain of operations, including observation, recording, collection, transmission, quality control, reconciliation, storage, cataloguing, and access. Every link in this chain represents an information interface subject to data friction." (Edwards, 2010, p. 84)

Over the last decade, climate scientists have engaged in a variety of international efforts to enhance the quality and quantity of meteorological data that constitute important global climate datasets such as the International Comprehensive Ocean-Atmosphere Data Set (ICOADS). Funding has been provided to aid the construction of some datasets, and various meteorological agencies have endorsed principles aimed at the full integration of weather and climate data networks in order to create a robust and sustainable climate data infrastructure. However, there remain many barriers to realising such a vision, and as Edwards (2010, p. 302) observes, a particularly “daunting task” is that of “refining and reconstructing the historical record”.

The current climatological record is relatively strong in some geographical regions such as the land regions of the northern hemisphere where detailed observational records can exist back to the late 19th century. However, the uneven development of meteorological observation and data archiving infrastructures around the world means that the historical data of many regions is sparse or non-existent. In order to deal with these gaps in the historical record, climate scientists can either generate synthetic data or recover archived data that are not yet integrated into international climate datasets. Scientists argue that by increasing the quantity and quality of observational data held in databases, the baseline for state of the art climate ‘reanalysis’ will be enhanced, improving the integrity and reliability testing of climate models. One source of such data are historical ship logbooks in which crew members recorded detailed accounts of weather conditions at sea, and which are now stored as archival documents in different sites around the world.

While early efforts to integrate meteorological observation data from different archival sources concentrated on variables such as pressure, more recently work has been directed at synthesising marine and terrestrial surface temperature observation records from the last 200 years. The Atmospheric Circulation Reconstructions over the Earth (ACRE) initiative aims to bring together people and organisations from around the world in order to achieve this ambition. In this paper, we focus on the efforts of those who are engaged, both directly and indirectly, through the ACRE initiative in work to develop a robust infrastructure for the recovery of historical marine surface temperature data from regions such as the Southern Ocean, South Pacific and Arctic. Our particular interest is in efforts to organise the recovery of meteorological observation records from the archived logbooks of historical ships, primarily those of the UK’s Royal Navy and the US Navy (c. 1800-1950). The intention of recovering this data is that they can be integrated into the ICOADS, and ultimately be used by climate modellers and application developers to better predict and respond to changes in the Earth’s climate. While there are significant quantities of meteorological records available to be recovered from archives, there are many challenges faced by climate scientists working to discover, digitise and transcribe such data and integrate them into the ICOADS.

The socio-material factors that influence efforts to develop data infrastructures for gathering, preserving and using historical meteorological records have been observed by Edwards (2010) and AUTHORS (2016). In our previous work, we identify the movement of historic meteorological observation data from ship logbooks into the ICOADS database as an example of a “data journey”. In our work on “data journeys” (AUTHORS, 2016), we developed a methodological approach for illuminating the socio-material ‘life of data’ as they move across different data infrastructures, from sites of production through to multiple sites of re-use in different sectors. By following data on their journey between different sites of practice, we were able to examine some of the complex socio-material forces shaping the constitution of data movements across space and time. An aim of developing the Data Journeys approach was to move beyond STS-influenced rich descriptive accounts of data infrastructures and the intra-network politics of their development as seen, for example, in the work of Edwards (2010), Leonelli (2013) and Ruppert et al., (2015). Instead, we aimed to address how data come to be constituted and used across multiple interconnected sites of data practice, and how the movement of data between such sites brings social actors into qualitatively new forms of relation with one another. The findings from our work on Data Journeys identified that shifting sites and moments of "friction" (Edwards, 2010) in the circulation of data were important to examine in order to develop critical insight into the power dynamics shaping the movement of data between different actors, and thus emergent social relations. In later work, we further developed Edwards' concept of "data friction" to identify how the interrelated socio-material dynamics of infrastructure, socio-cultural forces and regulatory constraints, framed by deeper political economic dynamics, function to shape the nature of data frictions in the circulation of research data and online communications data (Author, 2017).

This current paper builds on this earlier work to provide an in-depth account of the socio-material data frictions that work to shape efforts to develop an infrastructure for the discovery, recovery and digitisation of historical ship log book records and integrate them into the global circulation of digital climate data, and aims to respond to calls for more critical empirical work on the constitution of data infrastructures (Kitchin, 2014). Our observations point to the ways in which tensions that emerge in efforts to progress scientific values and practices in the context of the early twenty-first century neoliberal state generate significant frictions in the journey of data from archived ship logbooks into the ICOADS database. Our findings illuminate how climate scientists and their collaborators adapt to this context in their efforts to overcome such frictions, and we examine the strengths and limitations of these adaptive practices.

Methodology: Capturing the socio-material dynamics of data infrastructures

As Larkin (2013, p. 330) observes, to define an infrastructure as an object of study is "a categorizing moment"; an infrastructure should not be understood "in any positivist sense, [as] simply 'out there'", but as something constructed in part by the perspective of the researcher. There are many potential ways of entering, and defining the limits of, the field of study in order to understand better the socio-material factors shaping the development of a data infrastructure and the movement of data through it. In our research, we identified the ACRE initiative as a central node in the assemblage of material and symbolic forces shaping the journey of data from logbooks into the ICOADS, and entered the field from this site. Through this entry point, we then connected with other projects and individuals that were actively engaged in the data recovery effort at different sites along the "journey" of data from historical log books into the ICOADS database. This research was conducted as part of the [anonymised] project. As part of the project, we conducted in depth, semi-structured interviews with nineteen key actors heavily engaged at different sites in the UK's climate data infrastructure and observations of working practices and cultures at five of these sites including a digital ethnography of the forums used by citizen scientists engaged in the Old Weather project. We also analysed primary documentation relevant to the circulation and governance of climate data across emergent data infrastructures (e.g. organisational and project websites, policy documents). This research was ethically approved by [removed for anonymisation] Ethics Committee, and transcripts of many interviews are available at: [removed for anonymisation].

Our observations in this paper draw primarily upon data collected from expert informant interviews with leading climate scientists behind the ACRE initiative and the related Old Weather citizen science project, a climate historian whose work involves searching for suitable logbooks to be digitised in archives around the world, and citizen scientists who volunteer their labour by transcribing digitised records as part of the Old Weather project. These interviews were complemented by discussions with climate scientists during observations of their working practices at the Met Office Hadley Centre for Climate Prediction and Research and the University of East Anglia's Climatic Research Unit, and analysis of the Old Weather project forums. Analysis of online documentation relevant to ACRE, Old Weather and related initiatives was also examined in order to develop our understanding of the nature of the projects and the networks they were embedded in.

Data were thematically analysed (Ryan and Bernard, 2003) in order to draw out the key socio-material factors that (1) work to constitute the assemblage of human and non-human entities that make up the initiative and shape the development of the data recovery infrastructure, and (2) generate and frame responses to frictions in the journey of data from archived log books into the ICOADS. Examples of themes include, 'labour – paid and voluntary', 'archival practices', 'relations between people', 'motivation for engaging in project', 'affective', 'potential uses of the data', 'moments of friction' 'identifying records', 'funding the project'. Quotations from interviews are reported anonymously with each interviewee assigned a code representing the organisation or project they were part of (MO: Met Office (UK); CRU: Climate Research Unit (UK); OW: Old Weather project) and their position in the sequence of interviews at that site e.g. MO06 identifies the sixth person interviewed at the Met Office. While an exhaustive study of the entire network of actors involved in the data recovery assemblage is impractical and beyond the scope of this study, through developing an understanding of the data recovery initiative from the perspective of those heavily engaged at key sites, we illuminate how complex and interrelated socio-material dynamics come into play, shaping the assembly of the data recovery infrastructure and the circulation of data across it.

Theoretical framework: Data Infrastructures and Assemblages

Our work on Data Journeys echoes work in cognate disciplines to illuminate the socio-material constitution and consequences of data movements in a variety of contexts (e.g. Leonelli, 2016; Merrick White, 2017; Beer, 2013). Despite their disciplinary differences, a common strand that draws many of these approaches together is an interest in the socio-material constitution of the data infrastructures that enable digital data flows, and how these contribute to the constitution of emergent forms of knowledge and/or social relations. In some cases, such work falls under the banner of "critical data studies"; an interdisciplinary field aimed at turning critical attention to the complex "technological, political, social and economic" context in which data are "produced, organised, analysed and employed" (Kitchin and Lauriault, 2014).

Academic studies of infrastructure have emerged across a variety of disciplines in recent decades with many scholars' identifying Star and Ruhleder's (1996) description of infrastructure as the seminal text of the field. For Star and Ruhleder, infrastructures are relational, rather than concrete, entities. As Larkin (2013, p. 329) astutely observes: "Infrastructures are matter that enable the movement of other matter. Their peculiar ontology lies in the fact that they are things and also the relation between things". Dourish and Bell (2007) describe how scholars have understood infrastructure to both "crystalliz[e]… institutional relations" and "shape individual actions and experience" (p. 416-7).

Scholars from various disciplines have adopted such understandings to examine the historical context for the development of infrastructure (e.g. Bowker et al, 2010). They have observed, for instance, the ways that infrastructures act as vehicles for a variety of socio-material forces: colonialism, neoliberalism, racism, and so on (Wilson, 2016). For Edwards (2003, p. 2-6), infrastructures "co-construct the condition of modernity". He observes how the development of infrastructure, in all its various forms, has been "constitutive of the modern condition", and how ideologies of modernity have contributed to defining its purpose and characteristics.

Drawing upon this wide body of work on infrastructure, Kitchin (2014) defines *data infrastructure* as the "the institutional, physical and digital means for storing, sharing and consuming data across networked technologies" (p. 32). He observes that both data and data infrastructures are produced by complex and contingent "data assemblages" made up of a variety of interrelated social, economic, political, material and cultural components (p. 25). To speak of a data infrastructure being "assembled" thus means to observe that it – as with most things - is "made up of precarious socio-material relations" (Muller and Schurr, 2015, p. 217).

Kitchin's concept of a "data assemblage" is inspired by work in the field of Science and Technology Studies (STS) which has adopted Deleuze and Guattari's (1980) philosophical concept of "assemblage" as largely synonymous with Latour's empirically grounded "actor-network" (e.g. Latour, 2005; Law, 2009). However, Deleuze and Guattari's assemblage thinking forms part of a more radical and politically-driven intellectual project than the heavily descriptive accounts of socio-material relations advocated by Latour and commonly seen in studies of infrastructure. For example, *A Thousand Plateaus,* the book in which the concept of assemblage is articulated by Deleuze and Guattari, is a ground-breaking post-structuralist work based on a deep critique of hierarchical systems of organisation, and with an emphasis on the potential for objects to be re-assembled into new formations, rather than rich description of what currently exists as seen in work inspired by Latour. This broader conceptualisation of assemblage has been adopted and developed by a range of scholars outside of the field of STS (Muller and Schurr, 2015) to illuminate the ways in which complex associations amongst human and non-human elements constitute the world (e.g. DeLanda, 2006; DeLanda, 2016; Li, 2014).

DeLanda (2006; 2016) has arguably articulated the richest theoretical development of Deleuze and Guattari's concept of 'assemblage', presenting a realist social ontology that aims to avoid the limitations of over determination at both the micro and macro levels. While there are certainly limitations to DeLanda's 'flat ontology', his ideas around assemblages do offer a valuable lens through which we might imagine the ways in which things – in our case a data recovery infrastructure - are constituted through complex and dynamic relations between human and non-human entities. For DeLanda, assemblages operate at different scales – an assemblage can be an individual person in the same way it can be an individual city. All these assemblages have "a fully contingent historical identity" and are composed of heterogeneous components including material and symbolic artefacts as well as people. These assemblages interact with other assemblages – they can be component parts of larger assemblages, and different scales of assemblage e.g. a person and a city can interact with one another. Crucially, these assemblages are emergent and relational: each assemblage emerges from interactions between component parts, and once it exists it acts as a source of limitations and opportunities for components – the majority of which will come into existence once the assemblage has constituted itself.

Within the field of LIS, Deleuze and Guattari's concepts have not been adopted as a theoretical lens to the same extent as other post-structuralist theorists such as Foucault and Bourdieau (Robinson and McGuire, 2009). However, in recent years a small number of LIS scholars have begun exploring the potential of assemblage approaches for critical examination of LIS institutions and objects. Faucher (2014), for instance, examines documents as assemblages "that are in constant negotiation with an environment", and Gerolami (2015) adopts the assemblage concept to consider the library as a potential institution of social justice. Synthesising such insights, Bilmyer (2018) examines archives and their contents as "assemblages of politicized decisions". This recent body of LIS scholarship on assemblages, contributes to a wider, if limited, engagement with other concepts of Deleuze and Guattari within the field. For example, Robinson and McGuire's (2010) examination of the concept of "rhizome" for re-thinking information organisation in the era of the web, and Day's (2010) work exploring how Deleuzian perspectives may open up new ways of imagining the affective nature of information and LIS' construction of the user.

As others have pointed out, despite their limitations, assemblage approaches can help to illuminate and "make sense of" (Kitchin, 2014, p. 24) the various components that shape data objects and flows. In this paper, we adopt the concept of an assemblage as a conceptual lens through which to illuminate UK-based climate scientists' efforts to develop a data infrastructure for the recovery of climate data from historic ship log books, situating these efforts within the context of the deepening neoliberalisation of the states that many of the contributing scientists are embedded within.

Findings: Assembling an infrastructure for climate data recovery

Observing those involved in the journey of data from archived ship logbooks into the ICOADS database highlights that efforts to recover data are dependent upon the joining up of a number of already existing data and knowledge infrastructures. These include the archival infrastructures for storing and cataloguing historical records within different countries, the data infrastructure for the production and sharing of global climate datasets such as the ICOADS, and the infrastructure behind the Zooniverse platform which supports various online citizen science projects including Old Weather. The assemblages of material and symbolic forces that shape how this linking up happens in practice, and that constitute how data are recovered and move between log book and ICOADS are examined in detail below. Here, it is enough to observe that at the core of the assemblage are the climate scientists actively working to drive forward the initiative, and towards the periphery of the assemblage are a heterogeneous range of component assemblages upon which the project is dependent. For example, the archival community and its artefacts can be said to constitute one set of assemblages, those that hold the material means of production i.e. government funding agencies and commercial organisations form another. The communities and tools of volunteer 'citizen scientists' make up a further set of assemblages. Furthermore, much of this linking up takes place within the context of the assemblage of socio-material forces that constitute the contemporary neoliberal states of, for example, the UK and USA.

In the following sections, we examine the socio-material dynamics that bring into being and hold together this assemblage, thus shaping the development of the data recovery infrastructure and the circulation of data across it, from the perspective of those actively engaged in driving the initiative forward. We begin by describing the production and nature of the historical records, before moving on to examine the motivation to integrate these records into the datasets which form a foundation for climate research around the world. We then examine the sources of friction that frustrate data recovery and circulation, and end by examining how the assemblage of actors engaged in the data recovery initiative responds to - and constitutes itself in relation to - these frictions and takes advantage of opportunities to progress data recovery despite the challenges.

Historical meteorological data

The historical records that climate scientists are aiming to recover and digitise were created by the crew members of ships, most of which were active during the period 1800-1950. The records were created so that ships' crews could monitor weather conditions at sea in order to ensure safe and efficient navigation. The log books contain a page for each day in which handwritten records of a variety of meteorological measurements (e.g. sea temperature, wind speed, air pressure) were made, in some cases as regularly as every 4 hours. Complementing these quantitative measurements are a variety of qualitative observations of conditions such as ice and cloud cover. The climate historian we spoke to observed the high quality of these records:

> “they were very good observers, but they didn’t necessarily understand what they were observing. But their observations were excellent, make no mistake about that” (CRU_03)

Once finished with, the log books were archived in a variety of places around the world. In some cases, they were archived in national archives (e.g. The National Archives and The Met Office archives in the UK), while others were placed in specialist archives e.g. whaling industry archives.

Specifically how data were recorded in the log books varies from country to country, particularly in the case of older ships that were active in an era when there was little standardisation of measurements. For example, ships from different countries tended to use different instruments and scales for measuring weather conditions. Until 1935, for instance, Norway used a different wind scale to other countries. Countries also had very different ways of recording a ship’s location – a vital data point necessary to integrate the observation into climate records.

These conditions of production and archiving mean that while a rich collection of records for understanding historical climate conditions exists, the records are not standardised and they are widely distributed across a large number of, often unknown, physical sites around the world.

The drive to recover historical meteorological data

Initial efforts to develop the ACRE data recovery infrastructure emerged in the Australian agricultural sector, as a response to needs of crop and pasture modellers working for the Queensland government’s prime industry section for a “bigger, better” (MO06) meteorological database from which to develop their models. The Queensland State Government provided three years of funding during the period 2007-10 to a climate scientist based at the Met Office’s Hadley Centre (UK) to begin work on what became the ACRE project, now an international initiative involving 9 core partners including the UK’s Met Office.

This beginning to the initiative demonstrates how the value of the data recovery effort is largely framed from a technological application perspective. It began with the need to develop better agricultural applications to inform Queensland’s farming industry. Today, the recovered data that are integrated into the ICOADS database are similarly valued in relation to how they are used. They are understood by climate scientists to be the “building foundations…the basic building blocks” (MO03) for state of the art climate reanalysis models such as the 20th Century Reanalysis project, as well as a wide range of climate applications used in industries including, but not limited to, agriculture. For the central actors in ACRE and its collaborating projects, the key drivers for data recovery can therefore be understood as revolving around efforts to deepen scientific understanding and develop technological responses to the Earth’s changing climate.

Beyond the scientific core of the ACRE assemblage, the drivers for those actively engaged in recovering and integrating data into the ICOADS database tend to fall in line with these scientists’ perceptions of the scientific value of the project. In assemblage theory terms, the assemblage is relatively “coded” (DeLanda, 2016); that is, participants’ share a relatively coherent identity as a group of people engaged in important infrastructural work necessary to the future of climate related research and technological development. For example, individual volunteers had many different motivations for working on the Old Weather project and dedicating much time and energy to transcribing digitised records, however they tended to do so on the basis that they trusted the climate scientists that this was important work necessary for addressing the challenge of climate change:

> "I mean to my mind the name of the game is trying to find out what's going on with the climate." (OW_02)

This was a trust that was actively fostered by leading climate scientists, who regularly engaged with volunteers within the various online forums that are crucial artefacts in the Old Weather project assemblage that the citizen scientists are part of (AUTHORS, 2016). Similar to the Old Weather volunteers, the climate historian perceived that he was making an important contribution to the scientific effort to understand and respond to climate change:

> "My key motivation is – well, it may sound rather strange but I actually feel I'm doing something extremely useful…I feel like I'm making a contribution to help the scientists do their work. So, we all like to think we're doing something useful…I can't save the planet on my own. I can just do my little bit" (CRU_03)

As in the case of the volunteer transcribers, this perception was something that was actively supported by the climate scientists collaborating with the historian and making use of the archived records that he was skilled in being able to locate.

Data friction in the recovery and circulation of historical records

Most of the meteorological observations recorded in ship log books have laid inactive in archives for decades. While some of the log books that these data are recorded in are catalogued and well preserved, in many countries ship log books are "entirely uncatalogued" (MO03) and in some cases not properly preserved. For example, participants reported that the logbooks stored in the national archives of the UK and USA tended to be relatively well catalogued and looked after, in contrast to a building full of "mouldering" records that had been located in Mauritius that needed urgent "rescue" (CRU03). These local variations in archival infrastructures and conditions have a significant impact on efforts to shift data from their inactive analogue state, and 're-assemble' them as active digital objects circulating within present-day climate data infrastructures. These variations are therefore a key factor in the potential for data friction in the integration of a given set of records into the ICOADS.

While it is easy for climate scientists to recognise where there are gaps in the historical climate data record, it is much more difficult to know how to fill those gaps. This is particularly the case if catalogue records created by archivists about their collections contain no or partial metadata about the contents of the ship log books they hold, or if the catalogue records are difficult to access remotely. The ability to locate useful collections was perceived as a unique skill, with everyone recognising the valuable contribution of a single climate historian in undertaking this particular task:

> "The only way to really find out what's there is to actually go and look at the pages, read the documents, find out what's in them…The best thing to do is to ring up [climate historian] and send him to the archives…because he's a big asset for what we do because he's trained as a historian, he knows a lot about archives and what makes them tick, and how to find things, and how to deal with archivists…and he also knows what we are looking for" (MO03)

Such tacit knowledge about how archivists deal with such records was vital to the development of the assemblage. The culture of the archival profession was identified by the climate historian as a challenge exacerbating data friction. From his experience of working directly with archivists, he perceived that there was a tendency among archival practitioners not to place a high value on the logbooks due to their largely numerical content. His perception was that having being trained in the humanities, many archivists simply did not recognise the value of the numbers in the logbooks and did not know how to catalogue them. This he had learned over time tended to result in logbooks being catalogued under the generic heading ‘miscellaneous’, a label he had come to read as meaning the source ought to be examined for interesting data that could be integrated into the assemblage. This disciplinary difference within archival practice was perceived to account for the significant friction experienced by our climate historian in his efforts to locate logbooks deposited in archives around the world. This anecdotal observation of cultural differences between the archival profession and the requirements of climate scientists reflects research in the wider field of scientific record management. Shankar (2004, p. 371), for example, observes that the task of managing, appraising and arranging scientific records tends to cause difficulties for archivists given scientific practice “often does not conform to the ‘traditional wisdom’ of the archival community”.

Once relevant records were located in archives, we observed that a further set of frictions that slowed down the process of integrating data-objects into the assemblage began to emerge. In order to bring the records into circulation in contemporary data infrastructures, the analogue data must first be digitised. This generally means photographing the paper records. Significant frictions can therefore arise when records have been poorly preserved, for example faded inscriptions were recognised as a challenge as it was difficult to create digital representations of these records by camera. Once readable digital images had been successfully produced, the cursive writing style of the crew members that recorded the observations presented another significant source of friction. The ability to recognise these inscriptions is beyond the capabilities of current day optical character recognition technologies, meaning that in order to turn the handwritten inscriptions into machine readable digital numerals each data point has to be transcribed by a human. To ensure the accuracy of these manual transcriptions, each observation has to be completed by three independent transcribers. Once transcribed, homogenisation was also vital. As described on the Old Weather project website:

> "Don’t worry though…we will go through the results and clean up obvious mistakes, like large jumps in temperature or a ship that suddenly moves from one end of the Earth to the other” (https://classic.oldweather.org/faq)

The climate historian we spoke to explained how historically different countries had recorded observations in different ways, for example their ships used different wind scales or different ways of recording the ship’s position. This contextual knowledge was vital for the climate scientists engaged in the work of homogenising the data to ensure the records from different ships were comparable. The frictions experienced in this transcription and homogenisation process contribute to the 2-3 years timespan it takes for the records to become part of the official ICOADS database from the point the log books are initially digitised.

Clearly, a significant amount of human labour is required by the ACRE assemblage in order to overcome these infrastructural data frictions. However, participants reported being engaged in an ongoing struggle to convince funders of the importance of their work. UK based climate scientists employed by the national Met Office were unable to apply for the full range of funding available to researchers working in universities, and despite some government funding from the USA and Australia over the years, convincing government bodies of the value of the data recovery initiative was perceived to be a significant challenge. As one climate scientist noted: “at the moment [it] is not a good time to try and get money out of the government” (MO03). Participants felt frustrated by what they perceived to be conflicting understandings about what ought to be the outcome of public funding. It was widely perceived that public funding was difficult to obtain because those in a position to fund the work were only interested in “sexy…full blast modelling work” (MO06) and “concrete products…results” (MO03), and had demonstrated little interest in providing sustainable funding for the development of the data infrastructure underlying this work, despite its relatively low cost.

During the period 2000-2011, the US government had funded some digitisation activity through its Climate Database Modernisation Programme – an initiative that also created paid jobs in areas of high unemployment. However, in 2011 the funding for this programme was cut by the US Government, and since this time we observed that data recovery efforts had become more dependent upon volunteer labour and charity. In fact, other than a few individuals working in national meteorological agencies, a significant amount of labour engaged in the assemblage was currently unpaid, a situation that impacted the sustainability of the data recovery effort. For example, volunteers working with the International Environmental Data Rescue Organisation provide expertise to relevant organisations in the global south, volunteer crowd workers transcribe digitised records through participation in the 'citizen science' Old Weather project, and the freelance historian – at one point employed by the US's National Oceanic and Atmospheric Administration (NOAA) on a six year contract funded by the Climate Database Modernisation Programme to identify logbooks in archives around the world - was now working unpaid hoping that his current efforts would lead to paid contracts in the future.

This lack of sustainable funding for the initiative meant that over the years the process of assembling an infrastructure to move data from archives into the global circulations of climate data had been a slow process. One climate scientist observed that a lot of the work they were currently doing was "basic stuff [that] should have been done years ago really" (MO06) and that the pressure to get things done quickly meant that some tasks were therefore left unfinished. While the independent climate historian had actively begun looking to commercial collaborators to connect into the assemblage as a source of potential funding for data recovery projects, in general participants had not explored opening up new relations with commercial actors as potential sources of funding despite some recognising the commercial value of the applications that such data feed into.

Overcoming data frictions

The challenge of generating adequate public funding to foster the development of a climate data recovery infrastructure reflects wider concerns about lack of sustainable public investment within the neoliberal era. The perception that funders are more interested in "sexy" (MO06) models and the dependence of the initiative on the voluntary labour of a wide range of actors finds echoes across research disciplines and sectors. In response to these challenges, those at the core of the assemblage had adapted to the demands of the neoliberal context and fostered the development of ACRE as a highly decentralised project that could be run "on a shoestring" (MO06), surviving on in kind support and small amounts of money from partner organisations.

We observed that the ACRE project acted as a hub, drawing together an assemblage of organisations, projects and people with an interest in climate data recovery. Through this, ACRE was able to bring existing data infrastructures into new relations with one another rather than creating an entirely new infrastructure from scratch. For example, the ACRE initiative acts as a hub connecting national meteorological agencies, international datasets, voluntary organisations such as the International Environmental Data Rescue Organisation (IEDRO), and the Zooniverse online citizen science platform.

As one leading climate scientist observed, the lead scientist behind the initiative had been given a lot of freedom to innovate by his employer (the UK's Met Office), and in contrast to the highly bureaucratic meteorological agencies that it connected to, the ACRE initiative was run on a much more ad-hoc and flexible basis, comparing its structure to the “Pirates of the Caribbean” as opposed to “Royal Navy” (MO06) like meteorological agencies. The key climate scientists behind the initiative perceived ACRE to have developed “organically”, with “no great planning” and a lot of “serendipity” (MO06). A bottom up approach to assembling the infrastructure was fostered, meaning that there were no demands placed on any project or organisation to join the initiative, rather they joined if they were convinced ACRE was a worthwhile project to engage with. Relatedly, one participant explained how he had been toying with the idea of building a citizen science project to help transcribe log books for a while when he became aware that a professional contact knew someone involved in the GalaxyZoo project hosted on the Zooniverse platform. Through fostering a collaboration with this contact, and working within the loose structure of the data recovery assemblage, our participant was able to create the Old Weather project; a “powerful” (MO03) means of transcribing hundreds of thousands of digitised weather observations. This informal and non-hierarchical culture was perceived by participants as core to ACRE’s success. While it was recognised that the assemblage needed a core person to “keep it all linked up…the balls in the air” (MO06), there were concerns that if the project was passed to an institution such as the Met Office to lead it would become overly formalised and bureaucratic, and resultantly restrict the innovative work of the project.

Similar relations were also observable between the climate scientists and the volunteer transcribers working on the Old Weather project and the climate historian. All participation was on the basis of active interest and support for the cause, and the informal and non-hierarchical structures made for an enjoyable community for these participants to be part of. This allowed the ACRE assemblage to benefit from the enthusiasm and interests of those that it engaged (see AUTHORS, 2016b). The freelance climate historian, for example, was at heart driven by curiosity and serendipitous discovery:

> “yeah...my favourite research method is serendipity. I’ll give you an example…just now I was looking through this [book – Meteorological and Hydrography by Charles Meldrum] and I saw reference to a Swedish ship called the Eugenia, right?... He was saying it’s wonderful this ship is doing sea temperatures and all sorts of other things…So, I look through what I already have and I have a note of this ship, but not a note of it having a logbook that I knew about…So, I just did a little search…and they’ve got a collection of papers and results from the expedition…published in 3 volumes in Stockholm in 1857. So that book will exist somewhere and we can go and, even if we don’t find the log, we can find the scientific observations…So, serendipity. I just happened to look at something…it triggered a connection” (CRU_03)

The freedom he found working on the climate data recovery initiative – whilst not always for financial gain - allowed for him to satisfy these desires and needs while also making a crucial contribution to the success of the initiative. Similarly, participants in the Old Weather project perceived the voluntary nature of their work fostered their affective ties with the project, and in our analysis of their online forums we observed volunteers insisting that they did not feel exploited, nor wish to be paid for their labour as they perceived this would shift how they related to the project. Yet, volunteers’ emotions in relation to their labour were not always positive, for example, some reported feeling “guilty” (OW01) when drawn away from their transcription work by other interests and commitments on their time. Here we observe the ways that desires and affects make their way into the constitution of the assemblage, holding it together – at least for the present – despite the sometimes challenging material conditions. As Muller and Schuur (2016, p. 224) observe, it is the affective in the form of desire or wish – a "positive, productive force" – that "makes assemblages coalesce together".

The challenges generated by the socio-material conditions under which participants were working to assemble a data recovery infrastructure, we observed, fostered a determination among participants to succeed against the odds and create a culture that seemingly empowered different actors and enabled the assemblage to adapt to the frictions that it encountered. Over recent years those involved in the project have, therefore, been able to recover significant amounts of data from archived log books, albeit at a slower pace than may have been the case with a sustainable funding stream.

Discussion: tensions in the data recovery assemblage

As observed in the findings, the ACRE initiative is highly dependent not only upon fostering a productive culture for data recovery at the core of the assemblage, but also in linking with other already existing assemblages of actors and infrastructures with their own pre-existing cultures and practices. As we move from the core to the edges of the data recovery assemblage, we begin to see how frictions in the recovery and integration of data are generated due to the dependency of the core on linking with other already existing assemblages including archives, government funders, national meteorological agencies and individual volunteers. The data recovery infrastructure thus becomes shaped by a variety of elements that "frame what is possible, desirable and expected" (Kitchin, p. 24) of the data held in archived ship logbooks.

We observed that the various actors engaged in assembling the data recovery infrastructure shared a loosely "coded" (Deland, 2016) identity as a group engaged in crucial work necessary to the future of climate research and applications. We also observed the ways in which this assemblage of actors were drawn into collaboration with one another, in part, through an affective "force" (Muller and Schuur, 2016) related to their desire to engage their diverse enthusiasms and skills in the pursuit of deepening scientific understanding and technological adaption to the Earth's changing climate. These motivations reflect Edward's (2003, p. 2-6) observation that the ideologies of modernity (i.e. desires to gain control over nature and shape it by technical means to meet human needs) have long shaped the development of infrastructure. Yet, in the case of ACRE we also observe that this tendency towards modern scientific values exists within a space that simultaneously fosters a culture and practice in tension with meteorological science's existing institutional structure, which was perceived by participants to be highly formal and bureaucratic in nature.

The metaphor that ACRE was organised more like 'Pirates of the Caribbean' than the 'Royal Navy'-like bureaucratic structures of the meteorological offices (MO06), echoes Deleuze and Guattari's contrast between emergent, non-hierarchical "rhizomatic" forms of relations, and the rigid, hierarchical "tree" structures that they observe have tended to "dominate Western reality…and thought" (p. 18). We observed that the climate scientists leading the initiative had developed it at arms-length from the formal and bureaucratic structures of the core partners that support the initiative. In Deleuzian terms, these key actors can be understood as having entered a "line of flight" (p. 2) from such institutions, albeit while still remaining tethered to them through their employment. Resultantly, they have been able to develop their own adaptive and flexible culture which has enabled the initiative to overcome a variety of frictions.

For Deleuze and Guattari, a rhizome is dynamic, flat in structure, and not overcoded (p. 6). Rhizome forms, they observe "are defined by the outside…the line of flight…according to which they change in nature and ['ceaselessly' p. 6] connect with other [assemblages]" (p. 8). Given its desire to use technological means to gain control over nature and its ultimate dependence upon the meteorological institutions it distances itself from, the initiative cannot be identified as wholly rhizomatic in nature. Nonetheless, the assemblage theory lens is still useful for illuminating how rhizome-like tendencies generate a productive force that drives the data recovery assemblage, and its ability to overcome data frictions. Further, the assemblage lens helps illuminate key actors' concerns that if the project was handed over to a meteorological institution to manage, it would become constrained by these organisations' bureaucratic cultures. As Deleuze and Guattari (p. 14) observe:

> "once a rhizome has been obstructed, arborified, it's all over, no desire stirs: for it is always by rhizome that desire moves and produces".

Therefore, in terms of theorising the nature of the data recovery assemblage and its efforts to overcome data frictions, we observe that the values of modernity crystallising in the ACRE initiative – which uphold a desire to gain control over nature while also developing a critical practice at odds with the stabilising institutional forms of modernity - echo Beck et al's (2003) notion of "reflexive modernity". This concept captures not a 'rhizomatic' post-modernity, but a modernity that "has begun to modernize its own foundations…[a modernity that] has become directed at itself".

Similar to the national meteorological agencies, the archives in which the records are stored can also be understood as typically modern institutions. In many ways, archival practices have not been sufficient to fully meet the needs of early twenty-first century climate scientists who want to re-assemble historical records in new and innovative ways. As C.P. Snow observed back in 1959, western intellectual practice has been siloed into two seemingly separate cultures - that of science and that of the humanities – a development he and others have perceived to be a barrier to addressing the challenges facing societies. This binary separation, fostered and institutionalised as part of the traditions of modernity, was perceived by the climate historian who worked at the border of the sciences and humanities to have resulted in cataloguing schema and perceptions of value within the archiving community that frequently failed to satisfy the unforeseen needs of the data recovery initiative to re-assemble and "mutate" (AUTHORS, 2016) data objects into new formations for different ends. In the face of such challenges, the assemblage had to foster a new connection with an expert climate historian who was able to act as a bridge between climate science and archives, and rely upon his serendipitous methods to discover the location of valuable records.

We also observed the ways in which the assemblage adapted to a socio-material context that meant sustainable public funding to foster the development of the data recovery infrastructure was not forthcoming. Despite successes in finding ways to recover data with limited funding, it was clear from all involved that the development of the infrastructure was stunted due to a lack of sustainable funding. Further, the dependence of the assemblage upon a significant amount of volunteer labour and goodwill also left it potentially vulnerable to shocks, for instance, loss of critical personnel such as the climate historian or members of the small, but highly active core of citizen scientists involved in transcribing records as part of the Old Weather project. As assemblage theory reminds us there is an "ever present potential for breakdown and disruption" within the infrastructures that make current practices possible (Muller and Schuur, 2016, p. 222); the affects and desires that currently contribute to holding the assemblage together simultaneously have the potential to become destabilising forces producing breakdown of the data recovery infrastructure (Muller and Schuur, 2016).

To understand better the nature of the assemblage and its work to develop a climate data recovery infrastructure, we must also understand the data recovery assemblage in relation to the broader systems of assemblages that it is embedded within. As is common across all forms of infrastructure, in the assembling of the climate data recovery infrastructure, we can see the "underly[ing]…political rationality" (Larkin, 2013, p. 328) of the current wave of neoliberalism – with its fostering of tight restrictions on public investment, increased voluntarism and unpaid labour, and a demand for flexible adaptation placed on its subjects – crystallising into the constitution of the infrastructure. As Howe et al (2016, p. 549) note: "infrastructure…epitomizes the conjunction of material forms, expertise, social priorities, cultural expectations, aesthetics, and economic investments".

Despite our observation of this political rationality working its way through the assemblage into the constitution of the infrastructure, we noted that while climate scientists were critical of funding priorities and were keen to avoid fostering exploitative relations with volunteer workers, they did not forward a more systematic critique of the conditions of their work. For example, neither the underlying socio-economic drivers behind climate change nor the conditions of the late neoliberal political economic context were fundamental to the participants' understanding of their work or the data frictions they encountered. In this sense, their "reflexive modernity" was firmly focused on the institutional cultures and practices of meteorology and archives. In contrast, some participants seemed encouraged that the re-insurance industry – part of the institutional bedrock of neoliberal financial capitalism - were "really interested" (MO06) in the types of analysis that the recovered data enabled.

Relating to the assemblages of global financial capitalism in such a way suggests there is a space for increased reflexivity about the conditions of the assemblage's own existence and struggles to develop a data recovery infrastructure. It also points to a potential gap in how the assemblage tends to understand itself and its work within the frameworks of a scientific modernity which emphasises a separation between nature and society, and aims to leverage science and technology to exert control over an unpredictable environment, rather than working towards a deep and transformative understanding of the social roots of anthropogenic climate change. A potential limitation of the assemblage is, therefore, its failure to combine an advocacy for more horizontally structured and participative modes of governance with a critique of the overarching institutionalised structures of global neoliberal capitalism which continue to exacerbate the environmental crises that the data recovery initiative aims to respond to.

With its dependence upon substantial amounts of free labour and no systemic critique of context, the initiative thus risks becoming an assemblage of actors whose practices become co-opted into the dominant neoliberal response to climate change, rather than contributing to the broader project of social justice, as advocated in the context of critical data studies by e.g. Dencik et al (2016) and Taylor (2017).

Conclusion

Across sectors, diverse assemblages of socio-material forces are coalescing to develop a variety of data infrastructures that enable the integration and circulation of data across time and space. The data journeys that these emergent infrastructures enable bring social actors into new forms of relation with one another, and thus have complex societal implications. Deepening our empirical understanding of the assemblages of socio-material forces that are influencing the development of data infrastructures and the movement of data across them helps to identify the forms of value that are being crystallised through the development of infrastructure and illuminate what the potential implications of this may be. It also enables us recognise where critical praxis might need to be fostered.

Our research examining efforts to assemble a climate data recovery infrastructure begins to make a contribution in this direction, and illustrates the value of an assemblage theory framework for illuminating the means by which complex socio-material forces coalesce in temporary formations to produce particular forms of infrastructure, in our case for the recovery of historic climate records. Through adopting this approach, we are able to identify how the desires and enthusiasms of the assemblage of actors involved in the initiative relate with the socio-material context in a way that enables them to overcome data frictions and recover significant amounts of data despite a lack of sustainable funding. Placing a critical lens over this assemblage, we also observe the vulnerabilities of the assemblage as it develops in a reflexive and adaptive way to the neoliberal conditions of its existence, without fostering a deep and critical understanding of these conditions. Specifically, we observe the initiative's dependence upon voluntary labour, a drive to respond to climate change through scientific and technological means, and a lack of critique of the role of global neoliberal financial capitalism in the deepening of environmental crises. We conclude by arguing that to foster environmental and social justice - that is, the reforming of the foundations of society in a way that advances fairness, equity and positive freedoms for all - as the core value driving the constitution of the climate data recovery infrastructure, there is a need to go beyond the current reflexive stance of the assemblage towards the institutional structures it emerges from, and move towards a deeper critique of the broader socio-material context in which the work of constituting an infrastructure for climate data recovery unfolds.

Bibliography

Author (2017).

AUTHORS (2016).

AUTHORS (2016b)

Beck, U., Bonss, W. and Lau, C. (2003), "The Theory of Reflexive Modernization: Problematic, Hypotheses and Research Programme", *Theory, Culture and Society*, Vol. 20 No. 2, pp. 1-33.

Beer, D. (2013), *Popular Culture and New Media: The Politics of Circulation*, Palgrave, London.

Bowker, G. C., Baker, K., Millerand, F., and Ribes, D. (2010), 'Toward Information Infrastructure Studies: Ways of Knowing in a Networked Environment', in J. Hunsinger, L. Klastrup, & M. Allen (Eds.), *International Handbook of Internet Research*, Springer, New York, pp. 159–170.

Brilmyer, G. (2018), “Archival assemblages: Applying disability studies’ political/relational model to archival description”, *Archival Science,* Vol. 18 No. 2, pp. 95-118.

C.P. Snow (1959), *The Two Cultures,* The Rede Lecture, 1959.

Day, R. (2010), “Death of the user: Reconceptualizing subjects, objects, and their relations”, *Journal of the Association for Information Science and Technology,* Vol. 62 No. 1, pp. 78-88.

DeLanda, M. (2006), *A New Philosophy of Society: Assemblage Theory and Social Complexity*, Continuum, London.

DeLanda, M. (2016), *Assemblage Theory,* Edinburgh University Press, Edinburgh.

Deleuze, G. and Guattari, F. (1980), *A Thousand Plateaus: Capitalism and Schizophrenia,* Continuum, London.

Dencik, L., Hintz, A., and Cable, J. (2016), “Towards data justice? The ambiguity of anti-surveillance resistance in political activism”, *Big Data and Society* (Online Early)

Dourish, P. and Bell, G. (2007), “The infrastructure of experience and the experience of infrastructure: meaning and structure in everyday encounters with space”, *Environment and Planning B,* Vol. 34, pp. 414-430.

Edwards, P. (2010), *A Vast Machine: Computer Models, Climate Data, and the Politics of Global Warming*, MIT Press, Cambridge, Mass.

Edwards. P. (2003), ‘Infrastructure and Modernity: Force, Time, and Social Organization in the History of Sociotechnical Systems’ in Misa, T., Brey, P. and Feenberg, A. (Eds.) *Modernity and technology*, MIT Press, Cambridge, Mass., pp. 185-225.

Faucher, K. X. (2014), “An information meta-state approach to documentation”, *Journal of Documentation,* Vol. 7- No. 4, pp. 503-525.

Gerolami, N. (2015), “The library assemblage: Creative institutions in an information society”, *Journal of Documentation,* Vol. 71 No. 1, pp.165-174.

Howe, C. et al (2016), “Paradoxical infrastructures: ruins, retrofit, and risk”, *Science, technology and Human Values*, Vol. 41 No.3, pp. 547-565.

Kitchin, R. (2014). *The Data Revolution: big data, open data, data infrastructures and their consequences*, SAGE, London.

Kitchin, R., and Lauriault, T. P. (2014), ‘Towards critical data studies: Charting and unpacking data assemblages and their work’. Retrieved from: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2474112

Larkin, B. (2013), “The politics and poetics of infrastructure”, *Annual Review of Anthropology*, Vol. 42, pp. 327-43.

Latour, B. (2005), *Reassembling the Social: an introduction to actor-network theory*, Oxford University Press, Oxford.

Law, J. (2009), ‘Actor-network Theory and material semiotics’, in Turner, B. (ed.) *The New Blackwell Companion to Social Theory*, Blackwell, Oxford, pp. 141-158.

Leonelli, S. (2013), “Global data for local science: Assessing the scale of data infrastructures in biological and biomedical research”, *BioSocieties* Vol. 8 No. 4, pp. 449–465.

Leonelli, S. (2016), *Data-Centric Biology: A Philosophical Study*, University of Chicago Press, *.* Chicago.

Li, T. (2014), “What is land? Assembling a resources for global investment”, *Transactions of the Institute of British Geographers,* Vol.39, pp. 589-602.

Merrick White, J. (2017), ‘Following data threads’, in Kitchin, R. Lauriault, T, and McArdle, G. (Eds.2017) *Data and the City*, Abingdon, Routledge, pp. 85–97.

Müller, M. and Schurr, C. (2016), “Assemblage thinking and actor ‐ network theory: conjunctions, disjunctions, cross ‐ fertilisations”, *Transactions of the Institute of British Geographers*, Vol. 41 No. 3, pp. 217-229.

Pickren, G. (2018),“’The global assemblage of digital flow’: Critical data studies and the infrastructures of computing”, *Progress in Human Geography*, Vol. 42 No. 2, pp. 225 - 243.

Robinson, L., and McGuire, M. (2010), “The rhizome and the tree: Changing metaphors for information organisation”, *Journal of Documentation,* Vol. 66 No. 4, pp. 604-613.

Ruppert, E., Harvey, P., Lury, C., Mackenzie, A., McNally, R., Baker, S. A., Kallianos, Y. and Lewis, C. (2015), *“Socialising Big Data: From concept to practice”*, working paper, CRESC, Available at: http://www.cresc.ac.uk/medialibrary/workingpapers/wp138.pdf (accessed 30 June 2016).

Ryan, G and Bernard, H. (2003)., “Techniques to Identify Themes”, *Field Methods,* Vol. 15 No. 1, pp. 85–109.

Shankar, K. (2004), “Recordkeeping in the Production of Scientific Knowledge: An Ethnographic Study”, *Archival Science*, Vol. 4, pp. 367-382.

Star, S. L., and Ruhleder, K. (1996), “Steps Toward Design an Ecology and Access of Infrastructure: for Large Information Spaces”, *Information Systems Research*, Vol. 7 No. 1, pp. 111–134.

Taylor, L. (2017), “What is data justice? The case for connecting digital rights and freedoms globally”, *Big Data and Society* (online Early)

Williamson, B (2017), *Big Data in Education: The Digital Future of Learning, Policy and Practice*, London, Sage.

Wilson, A. (2016), “The Infrastructure of Intimacy”, *Signs: Journal of Women in Culture and Society*, Vol. 41 No. 2, pp. 1–34.