Data Journeys as an approach for exploring the socio-cultural shaping of (big) data: the case of climate science in the United Kingdom

Jo Bates, University of Sheffield
Paula Goodale, University of Sheffield
Yuwei Lin, University for the Creative Arts

Abstract
The paper reports on a pilot study aimed at developing, and assessing the utility of, a *data journeys* approach for critically exploring the socio-cultural shaping of interconnected data infrastructures. At various points along the journey of a (metaphorical) datum - from production through processing, re-use and intersection with other *data journeys* - selected organisations and projects are brought into focus and empirical data about the socio-cultural values and practices shaping the life of data within that particular space are collected using a variety of qualitative data collection methods. These empirical data are then critically and thematically analysed in relation to the broader social context.

This paper outlines the rationale for the *data journeys* approach prior to presenting initial findings from The Secret Life of a Weather Datum research project which applies the approach to explore the socio-cultural values and practices interacting with weather and climate data as they move through a variety of data infrastructures. The initial findings presented in this paper focus specifically on weather data production and climate science at Weston Park Museum Weather Station in Sheffield, UK; the Met Office (the UK's national meteorological organisation); the Climatic Research Unit at the University of East Anglia, UK (a world leading climate research institute); and the Old Weather Project (a citizen science project involved in historical weather data recovery and rescue). Emerging themes of vulnerability and resilience within the weather and climate data infrastructure are presented.

These initial findings highlight the importance of situating our understanding of data infrastructures firmly within the social. Through drawing on the *data journeys* concept to guide and inform the selection of sites for data collection, we begin to demonstrate the utility of the approach for beginning to build a picture of the "contingent and contested" (Dalton and Thatcher, 2014) relations between people, interconnected in time and space through data infrastructures, that are core to the development and shaping of climate data and knowledge. We also begin to draw out the interrelations between local and global spaces and infrastructures; and to ground amorphous 'big' data infrastructures in local sites and cultures of production.

Keywords: Big Data, Critical Data Studies, Data Journeys, Climate Science, Weather Data, Data Infrastructures
Citation: Bates, J., Goodale, P., Lin, Y. (2015). Data Journeys as an approach for exploring the socio-cultural shaping of (big) data: the case of climate science in the United Kingdom. In *iConference 2015 Proceedings*.

Acknowledgements: The paper is based on research undertaken for The Secret Life of a Weather Datum project, funded by the UK's Arts and Humanities Research Council (Grant number AH/L009978/1). The authors would like to thank all participants for the time and effort they have committed to the project.
Contact: jo.bates@sheffield.ac.uk, p.goodale@sheffield.ac.uk, ylin21@ucreative.ac.uk

1 Introduction

In a recent paper calling for social science to engage critically with the emerging public discourse around (big) data, Dalton and Thatcher (2014) remind us that

> "All data…is always the result of contingent and contested social practices that afford and obfuscate specific understandings of the world [and]…The innovation, production, and popular use of a technology occurs within and reflects a social context shot through with power, economies, identities, and biases."

As Kitchin (2014) notes there is currently a scarcity of critical empirical work on developments in the field of (big) data. However, an emerging field of critical data studies that aims to tackle such issues is gaining traction across a range of disciplines including Media and Communications (boyd and Crawford, 2012; Gitelman, 2013), Geography (Kitchin and Lauriault, 2014; Dalton and Thatcher, 2014) and Sociology (Ruppert, n.d.).

One approach for exploring the "contingent and contested" socio-cultural values and practices that shape (big) data production, processing, distribution and re-use is through the "inversion" of information and knowledge infrastructures (Bowker, 1994). As Edwards (2013, p. 20) explains:

> "to understand the infrastructure, you have to invert it. You turn it upside down and look at the 'bottom' – the parts you don't normally think about precisely because they have become standard, routine, transparent, invisible."

Edwards adopted the "infrastructural inversion" approach to analyse the socio-technical processes underlying the historical development of weather and climate knowledge infrastructures. Through looking beneath the surface of information and knowledge infrastructures and recognising them as social and relational, scholars have explored some of the complex and often invisible political, cultural and ethical processes that shape their development (see Bowker *et al.*, 2010; Star, 1999). Gitelman (2013, p. 4) recognises the utility of such an approach for exploring (big) data. Drawing on these studies of infrastructure, she argues that in order to understand the social shaping of data, we must "look[] into data or, better, look[] *under* data to consider their root assumptions".

Whilst infrastructural inversion offers one way of engaging critically with questions around the social shaping of (big) data, there are limitations in focusing on bounded infrastructures when re-use of data *across* infrastructures is becoming an increasingly important aspect of data practices. An alternative suggestion is therefore to focus on *data journeys*, a concept which easily transcends infrastructural boundaries.

The idea of an "information journey" in information science is most readily visible in Blanford and Attfield's (2010) work in information retrieval. Their model focuses on the "information journey" of a user as they travel from the initial point of identifying an information need, on to finding, interpreting, evaluating and using the information they discover. The idea also has an underlying presence in work on 'paths' and 'trails' through interactive systems; both as learning aids and as a means of tracing where users have been (e.g. Furuta et al, 1997; Reich et al, 1999; Shipman et al, 2000; White and Huang, 2010). Research drawing, either explicitly or implicitly, on the idea of an information journey has therefore tended to be orientated towards understanding and improving information access and information interaction for individual users, rather than on understanding the wider socio-cultural context of data production, use and reuse, which may impact significantly on the user's information journey.

Sands et al's (2012) "Follow the Data interview protocol" begins to suggest an alternative way of conceptualising information journeys by focusing on the flow of data as it moves through particular spaces. Their interviews with astronomers focus on the flow of data leading into and out of research publications in order to explore the "people and infrastructure involved in the building, sustaining, and curation of large astronomy sky surveys" (p. 1). Whilst Sands et al's (2012) research is focused on a specific aspect of a data journey, it demonstrates how a focus on data objects as they move through time and space might be a valuable approach for capturing the "contingent and contested social practices" (Dalton and Thatcher, 2014) shaping data as they journey through and interconnect different sites of production, processing, distribution and re-use. As anthropologist Appadurai (1988, p. 5) argued, in order to understand socio-cultural value in relation to things, in our case data, we need to "follow the things themselves".

Such an approach is loosely applied by Hallam Stephen (2013) in his history of bioinformatics in which he traces scientific data as it moves from organisms, to sequencing machines, to software, databases, images and scientific publications. Similarly, the Socialising Big Data project led by sociologist Evelyn Ruppert in the UK is exploring the "social lives of data-objects" as they move through and are employed within the Office for National Statistics, urban waste management and genomics, with the aim of exploring the potential for (big) data to both illuminate and create "risks and vulnerabilities" (Ruppert, n.d.).

In this paper, we conceptualise a *data journey* as the journey of a data-object through interconnected data infrastructures, from its original 'raw' state, to being cleaned and otherwise processed, to becoming an input into various forms of derived data. Further, the notion of a *data journey* can be theorised more abstractly in relation to the spatial characteristics of data infrastructures, including the journeys and interrelations between the local and the global, and between small, singular sites of data production and big and multiple sites of data use and re-use. We can assume that there will be multiple intersecting *data journeys* to be explored across interconnected data infrastructures, and our aim is not to comprehensively map all data journeys within a given set of infrastructures.

There is significant scope for further research around this approach, particularly in relation to the development of more systematic methods for approaching *data journeys,* developing methodologies for understanding the socio-cultural values and practices and power relations that they are embedded in.

The aim of this paper is twofold. Firstly, it aims to begin to demonstrate, through the presentation of initial findings from The Secret Life of a Weather Datum project, the value of a *data journeys* approach for guiding and informing the building of a picture of how "contingent and contested" (Dalton and Thatcher, 2014) relations between people, interconnected over time, place and space, are core to the development and shaping of climate data infrastructures. Secondly, it aims to outline data journeys within the UK's weather and climate data infrastructure and to illuminate the different socio-cultural values and practices and power relations in circulation at various sites of weather and climate data production, processing and re-use in the UK.

2 Methodology

The research reported in this paper is based on the initial findings of a wider study (The Secret Life of a Weather Datum) which aims to develop and pilot a *data journey* approach for better understanding how socio-cultural values, practices and power relations shape the flow and transformation of weather data on its journey from production through to various contexts of processing, distribution, use and re-use as derived data and information. Four key areas of investigation were selected: initial production and processing; climate science; financial markets; and citizen science initiatives. These particular areas were selected because they incorporated a range of social contexts including institutions of state, science, and the market, as well as grassroots citizen participation.

The *data journeys* approach that is being developed is based upon following the journey of a (metaphorical) datum – in our case a weather temperature datum – and using this journey as a guide to inform the selection of interconnected sites for researching socio-cultural values and practices within data infrastructures. At various points along the datum's journey from its initial production at Weston Park Museum weather station in Sheffield, UK through to different contexts of re-use in climate science and financial markets, key organisations and events were selected in order to collect empirical data relating to the socio-cultural values and practices shaping the life of the datum - or data derived from it - within that particular space. The study also begins to explore intersecting *data journeys*, specifically in relation to data generated by citizen science projects, in order to begin to build a more complex picture of the data infrastructures our primary datum exists within.

Data collection methods at each of these sites of investigation include semi-structured interviews that adopt a partial oral history approach, observations including photographic observations, digital ethnography of spaces for online participation connected to the organization, event or project, and the collection of online information such as relevant policy documentation and website content.

An initial thematic analysis of part of the empirical data has been undertaken, focusing on explicit and implicit values and attitudes (e.g. integrity, trust, transparency), valued data practices (e.g. improving data quality, data sharing) and social relations (e.g. community, tensions). These themes are being analysed in relation to the broader social and political economic context that the participants and the data infrastructure are situated within. A more detailed picture of the data journey is also being extracted from the empirical data.

This paper reports on initial findings from data collected at sites of weather data production and climate science; specifically our initial findings from Weston Park Museum Weather Station in Sheffield, the Met Office (the UK's national meteorological organisation), the Climatic Research Unit at the University of East Anglia, UK (a world leading climate research institute), and the Old Weather Project (a citizen science project involved in historical weather data recovery and rescue).

3 Initial findings

3.1 **The journey of our temperature datum**

This section of the paper will outline the journey of our temperature datum from its production at Weston Park weather station, through to the UK's Met Office and production of national climate datasets, and then into global climate science and the production of global gridded temperature datasets. The intersection of this journey with data produced by citizen scientists engaged in the recovery of historical shipping data will then be described.

Our datum's journey begins at the Weston Park Museum weather station approximately 2 miles west of Sheffield city centre. Sheffield is a large city in the north of England well known for its industrial

heritage, particularly in the area of steel production. Data has travelled from this site to the Met Office since 1882; however since 2010 two sets of data have been recorded. The first is generated from the museum's equipment and is forwarded to the Met Office monthly to contribute to the climate record, and the second set is produced by new Met Office equipment which since 2010 has fed regular synoptic data direct to the Met Office.

As our datum leaves Sheffield Weston Park it is transmitted internationally via the Global Telecommunications System (GTS) of the World Meteorological Organisation which enables "rapid collection, exchange and distribution of observations and processed" weather data globally (WMO, n.d.). Within this system, our datum is integrated with observations from other national meteorological offices, before being downloaded by Met Office scientists working in a variety of fields within forecasting and climate science including the production of national and global gridded temperature datasets.

The next stage of our datum's journey focuses on the Met Office Public Weather Service. After being downloaded from the GTS, it undergoes a series of quality checks by Met Office staff in Edinburgh, before it enters the Met Office's MIDAS climatological database[1] hosted by the British Atmospheric Data Centre. From MIDAS, the datum is then processed by Met Office climate scientists contributing to the generation of a range of statistics and gridded datasets of the UK's climate for use by academics, policy makers, businesses, the media and the general public, and makes a significant contribution to national climate adaptation planning through the UKCP09 climate projections project[2].

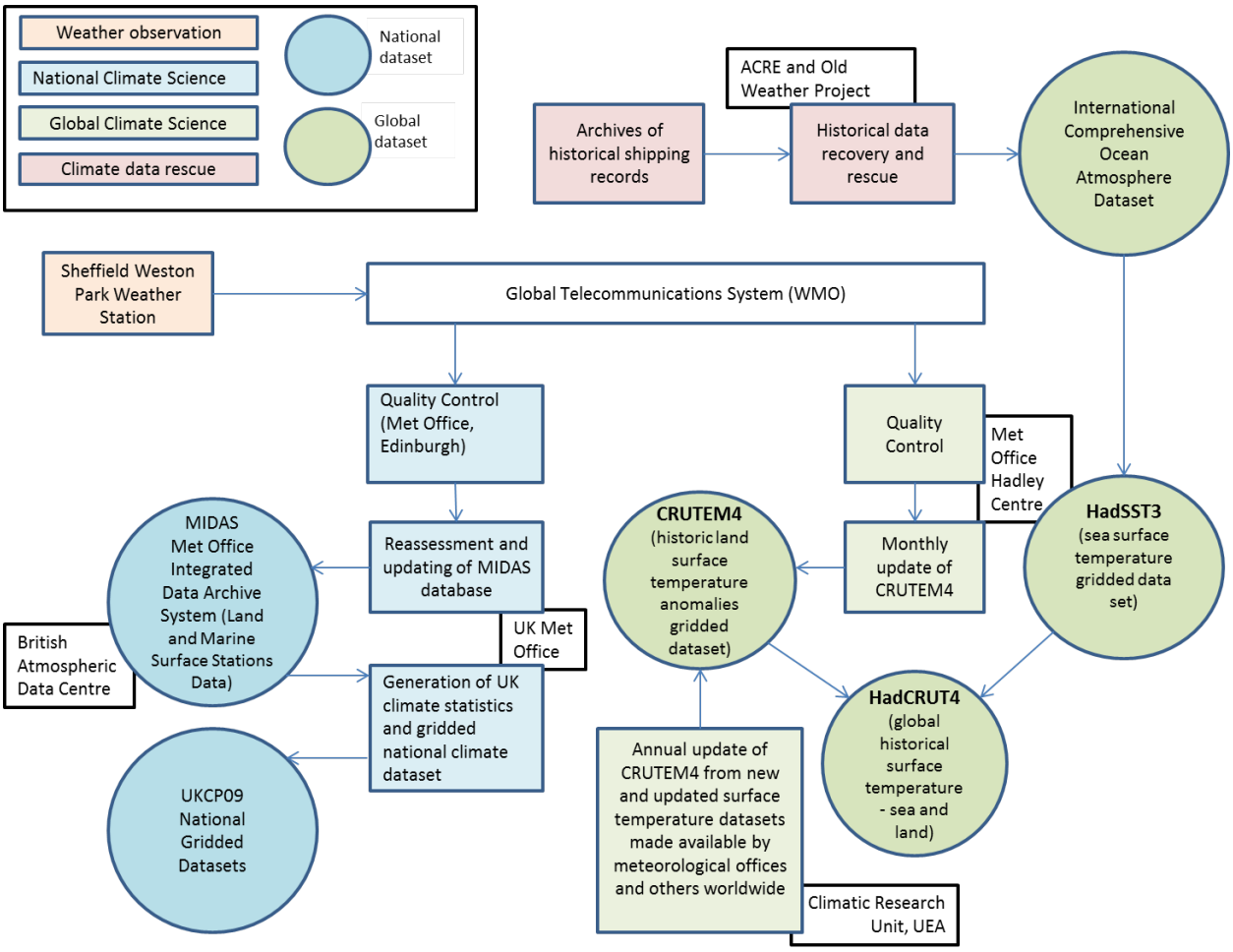

Figure 1 – Simplified (in progress) data journey map covering Weston Park, UK national climate science, global gridded dataset and historical data rescue.

[1] http://badc.nerc.ac.uk/data/ukmo-midas/
[2] http://ukclimateprojections.metoffice.gov.uk/

At the same time as the processing of national climate data occurs, our datum also enters the realm of global climate science. For this part of the journey we focus specifically on how data derived from our datum contributes to the production of the critically important global gridded land surface temperature dataset CRUTEM4. The monthly average temperature data from the Sheffield Weston Park station has long been an important data series for those developing the CRUTEM gridded dataset, and today it feeds into the monthly updates of the latest version CRUTEM4. CRUTEM was initially developed by the Climatic Research Unit (CRU), University of East Anglia, yet in recent years responsibility for these monthly updates has moved from CRU to the Met Office Hadley Centre for Climate Prediction and Research. However, once a year the CRU team updates the CRUTEM4 dataset with new climate data that has been made newly available or updated by meteorological agencies around the world.

The story of our datum presented in this paper will finish by exploring its intersection with the journey of data produced by citizen scientists and others who are engaged in recovering and rescuing temperature data from historical shipping records stored in archives around the world through projects such as ACRE and Old Weather. These data recovery efforts make a significant contribution to the historical climate record including datasets such as HadSST3 (Sea Surface Temperatures). Through combining together CRUTEM4 and HadSST3, climate scientists have produced HadCRUT4; a dataset which provides a picture of changes in the global surface temperature over time. Historical data recovery activities contribute to filling in gaps in the historic record, and make a significant contribution, alongside data from Sheffield Weston Park Station, to our understanding of temperature changes in the global climate since 1850.

The following sections will explore the socio-cultural values and practices encountered by our datum as it moves through these different environments, and intersects with the journey of data produced by shipping crews that is now being recovered from archives by citizen scientists. The story begins at Weston Park Weather Station in Sheffield, and then moves on through the Met Office and Climatic Research Unit, before exploring historical data recovery efforts of the Old Weather project.

3.2 **Weston Park Weather Station**

Our datum's journey begins at Sheffield's Weston Park Museum weather station. The station was founded in 1882 in an effort to understand better the link between weather conditions - specifically underground temperatures (figure 2) - and outbreaks of often fatal diarrhoea in the locality. The museum curator, Elijah Howarth, took on the job of setting up and running the station at the request of the local authorities, and since this time responsibility for the station has passed down through generations of curators. The station is one of the longest running and most complete climate datasets in the UK and thus makes a significant contribution to the UK's climatic record (Museums Sheffield, 2012).

As the current curator, Alastair McLean, describes, the weather station is part of the local fabric of Sheffield:

> "It means much more to us than it does to the Met Office I think. The Met Office may say differently. I do think they consider the dataset and the fact that it's still growing, important. But I think that that weather station out there means more to the people that work in the museum service and the people of Sheffield than it does to anyone else."

He reports how when a Campbell Stokes device - a valuable piece of observing equipment - was stolen from the roof of the museum in the 1990s, the people of Sheffield rallied to replace it, fundraising £3000 "in a matter of weeks".

This local meaning is a source of both pride and responsibility for the current curator, who exhibits a deep knowledge of the history of the station:

> "I'm quite proud of my involvement with the weather station, possibly more proud than I am of any other aspect of my curatorial work"

> "You just feel like you're the next link in the chain, which of course conveys a lot of pressure on to make sure you don't cock it up as well [laughs]. The weather station, during its life has been under threat on a couple of occasions. I think it's safe to say that, and you don't want you to be the last curator."

Ground temperature was originally observed at Sheffield Weston Park in order to understand outbreaks of fatal illness in the city. Readings continue to be taken in order to maintain the archive for posterity

Figure 2 – Alastair McLean (Curator of Natural Science, Museums Sheffield) with the ground temperature thermometer at Weston Park Weather Station.

This sense of rootedness in time and place has a profound impact on the data that is generated by the station. Significant pride is taken in ensuring the quality of the data:

> "Quality was very important, as it still is, because any data that we sent was rigorously checked by the Met Office…and any data that was considered wrong for whatever reason was always sent back with a red underline, and we felt bad that we'd made a mistake of some variety. It didn't happen very often."

Once it is generated, the data is understood by the curator to be "part of our heritage" and within "public ownership". The curator is dedicated to contributing to a wider local weather data ecosystem, answering over 200 queries a month from local people about weather conditions, making regular appearances in the local newspaper – the Sheffield Star ("They've even got a stock photograph of me now!"), providing datasets for local students and academics, providing a Twitter feed of observations for 600 followers, and even undertaking his own research project on the impact of weather on the emergence patterns of male black garden ants, which he plans to publish in the local natural history society's journal - *The Sorby Record* - "after a decade or so" of data collection.

Much of the data produced by the museum station flows relatively freely, however, barriers to free re-use of the museum's weather data are erected by the curator that reflect an understanding of the relationship between public ownership and private exploitation that challenges the increasingly dominant notion that public data should be an open infrastructure readily available for commercial exploitation:

> "I think that there probably is a risk that if you make it open for everybody, including people that are going to make an awful lot of money from it, you're ripping off the country to a certain extent."

The station therefore sustains itself through levying token charges (£15) for some uses of the data. Yet, these charges are applied in a way that aims to avoid exploiting the data unfairly:

> "What I tend to do with insurance claimants is check the data first to see if it's backing up what they're saying and then say, 'it's not, and we don't need to charge you for that reason.'"

As the curator describes, despite its importance to the local community, in recent years the Sheffield Weston Park museum weather station has been threatened by public sector austerity measures which have led to staffing cuts. In order to adapt to some of these pressures and ensure the continuity of the weather station, the previous curator allowed the Met Office to add its own weather observation equipment to the weather station compound in 2010. This new equipment feeds data directly to the Met Office and is part of the national synoptic network. As the curator describes, this development, whilst

likely necessary for the longer term preservation of the station, has impacted upon the power relations between the local museum weather station and the national Met Office:

> "It's more one sided I think, the relationship. In the past they needed us more than we needed them, whereas now it's not the case…it's much less of an equal relationship I think from that point of view."

The addition of the Met Office observation equipment thus threatens to disturb the embedded nature of the UK's climate data infrastructure within the local history and culture of Sheffield. Yet, for the time being the museum station remains resilient, defended by a curator who recognises the Met Office equipment as "secondary to our equipment", and with confidence that the people of Sheffield would "do their nut, not to put too fine a point on it" if the museum station was ever threatened with closure.

3.3 The Met Office: The Public Weather Service and national climate science

Synoptic data that flows from Sheffield Weston Park is transmitted to the WMO's Global Telecommunication System from where it is downloaded by Met Office climate scientists working within the Public Weather Service to understand better national climate change (Met Office, n.d.). The monthly climate data from the museum equipment is also integrated into this process.

What was to become the UK's Met Office was founded in 1854. Since the 1980s, the commercial value of the meteorological data collected and processed by the organisation has played a more important role. In 1990, it became a commercial body within the Ministry of Defence, prior to becoming a Trading Fund – a form of public sector body in the UK that must generate a significant proportion of its income from commercial activity – in 1996. Despite the increasing emphasis on its commercial activity, the Met Office's primary customer continues to be the UK public sector and most of its income it receives from the state for provision of the Public Weather Service. However, the organisation also generates a proportion of its income from the sale of data and other value added information products to commercial users. The UK academic community is able to use Met Office data free of charge through the British Atmospheric Data Centre. In 2011, the Met Office, along with a number of other data rich Trading Funds, became part of the new Public Data Group within the Department for Business, Innovation and Skills as part of the government's Open Data strategy, and the Met Office came under increasing pressure to make its unrefined data open for free commercial re-use. In response to this it launched the DataPoint API[3] which allows re-users access to some of this data under the UK's new Open Government Licence.

Whilst meteorological science remains at the heart of the UK's Met Office, over recent years the context that this science exists within, like many public institutions, has shifted towards an increasingly business orientation. A series of initiatives within the Met Office have aimed at generating a dialectical relationship between the science and business sides of the organisation in which the two "push and pull both ways" (Andy Brown, Director of Science in Met Office (2013)), to create a synthesis of the two known internally as the "Bow-tie". As one climate scientist described:

> "Well, it's service driven…but, underpinned by the integrity of the science and the methods that are applied"

This fusion of science, business and public service produces a somewhat ambiguous socio-cultural space. This uncertainty is represented most obviously in the organisation's at times conspicuous efforts to create a public identity for itself. The Met Office's corporate identity is central to its functioning, with efforts to generate and present a unified identity through a variety of mechanisms. A strong emphasis is placed by the organisation on brand identity, with Met Office colours and logos saturating public spaces and offices within the buildings (figures 3-5).

The change in name to *Met Office* in 2000 further represents the prioritisation given to brand identity by the organisation. The organisation's public identity is also shaped by the creation of a restricted discursive space for employees, whereby engagement with potential critics including the press and academics is heavily controlled by the organisation's Press Office, indicating a level of guardedness within the internal culture of the organisation.

This restricted discursive space is *in part* a product of the broader trends of neoliberal New Public Management that have been adopted by the UK's public sector over the last three decades (see Laing, 2003). A strong corporate identity has become increasingly perceived as necessary not just to leverage one's position in the commercial market place, but also to build reputational capital amongst domestic and international networks of policy makers, funders, and potential collaborators. In the case of the Met

[3] http://www.metoffice.gov.uk/datapoint

Office, the managerial impulse to control discursive space can also be understood as a response to the current vulnerability of the institution as it comes under threat of outsourcing by policy makers:

> "They haven't been subject to some of the disciplines that even the rest of the public service has been subject to…Privatisation is probably the poor model. A better model is…where…it's still a government owned asset… but they've been sort of hollowed out to a private sector management company to run it in the most efficient way possible. And that is something that is [?screaming] out to be done in the case of the Met Office" (Senior UK Policy Maker, recently retired).

Figure 3 – Operations Centre of the Met Office public weather service, Exeter, UK (1)

Figure 5 – Operations Centre of the Met Office public weather service, Exeter, UK (2)

Figure 4 – 'The Street' - lobby of the Met Office, Exeter, UK

Note the heavy use of the 'swoosh' logo on office surfaces and plant pots, and the use of banners to present a corporate identity.

Despite efforts to synthesise the two ends of the 'bow-tie', this socio-cultural context seems at odds with the science that is at the core of the institution. There are glimpses of the subtle impact of this discursive space on the way that science is imagined by some scientists working within the organisation. For one individual, the language used to discuss work was heavily shaped by a managerial discourse of "our remit… function" and the term "climate services" was enthusiastically embraced. Similarly, restrictions on sharing Met Office weather data – for example, the fact that free use of the Met Office climate database MIDAS is restricted to UK academics - were described but not questioned. Interestingly, when discussing the reduction of the rain gauge network by nearly 50% since the 1970s, the effectiveness and efficiency of the observing network were the primary considerations for this individual; the idea of a weather station being an important part of the local community as in Sheffield was not mentioned. This suggests a degree of alienation from cultures of data production, and the sense that as our datum enters this 'big' data infrastructure its socio-cultural context has shifted towards a more technocratic space that is less about people, and more about process, utility, functionality and economic reasoning. These examples suggest subtle ways in which a broader organisational culture can infiltrate into the scientific culture within it. Yet these findings should not be generalised, since, as will be explored below, scientific culture within the Met Office is multi-faceted, and in some cases may challenge some of the more restrictive elements of corporate identity described above.

3.4 Global climate science: CRUTEM4 and the Climatic Research Unit

At the same time as our datum is downloaded from GTS for eventual processing by those working on national climate science, monthly average data derived in part from our datum is downloaded by climate scientists working in the Met Office's Hadley Centre. The scientists are working in collaboration with the Climatic Research Unit (University of East Anglia, UK) to develop the CRUTEM4 and HadCRUT4 datasets that are aimed at better understanding the temperature of the global climate.

The Met Office Hadley Centre, which sits on the third floor of the Met Office, is one of the leading international climate research centres and here our datum enters a socio-cultural environment somewhat removed from the more corporate space on the lower floors of the Met Office. The absence of any branding within the working environment was particularly noticeable (see Figures 6 & 7).

Figure 6 - Offices of the Met Office Hadley Centre for Climate Prediction and Research, Exeter, UK (1)

Figure 7 – Offices of the Met Office Hadley Centre for Climate Prediction and Research, Exeter, UK (2)

Note the less corporate environment, particularly the absence of branding within the centre.

The CRUTEM gridded dataset was first produced in 1982 by researchers working at the University of East Anglia who were interested in developing a record of global temperature variability to complement their earlier work on pressure data:

> "We just thought at the [?time] it was an interesting thing to do… the first paper came out in '82 and…the temperatures had been cooling from the 1940s to the late 1970s. When we first did it, it wasn't to show warming, it was just to produce a dataset that we could look at."

Collating data from sources worldwide, including Sheffield Weston Park, the researchers generated the CRUTEM dataset which reports differences in the monthly average land surface temperature compared to the mean temperature for 1961-1990 in 5° by 5° grid-squares covering the land surface of the globe. Over the years this dataset has been updated and improved, and has transitioned through a number of versions, the latest being CRUTEM4 (Jones et al, 2012). Since the mid-2000s, CRU has worked in collaboration with the Met Office Hadley Centre in order to produce these new versions. The CRUTEM datasets have made a significant contribution to understanding global climate change over recent years.

The entrance to the Climatic Research Unit at the University of East Anglia is typically understated; with only a sign printed on A4 green paper taped to the door to indicate your arrival (figure 10). There is a sense of quiet pride within the Unit - recent publications, research posters, certificates celebrating achievements and team photographs are modestly displayed on the walls; the ostentatious marketing efforts that are becoming increasingly commonplace in Universities are noticeably absent.

The sense of being part of communities situated in place, space and time permeates the culture of CRU. The Unit itself is a place of deep seated community. Members find time for a daily crossword in the common space (figure 12), take turns to cook a group lunch once a week, and head off to a local café for lunch every Friday. Whilst Director of the Unit, Phil Jones, who joined CRU in 1976, is a less regular participant in the social life of the Unit, he is integrated into the community almost as a paternal figure. His main aim is to keep the CRU team together within the Unit; no mean feat given the precarious nature of a Unit income dependent upon short term research grant funding:

> "Our main aim is not a scientific one at all, it's to keep people like David and Harry funded….we are trying to do that [good science] as well, but we're also trying to keep the people we've got…trying to keep good people…to actually do the work on a day to day basis."

The community is similarly defended from efforts by University management to incorporate the Unit into another department and turn their common space into offices; a move that would undoubtedly weaken the community bonds. This protective familial structure is also witnessed in visual representations of the CRU team displayed in the Unit which resemble the structure of a family tree.

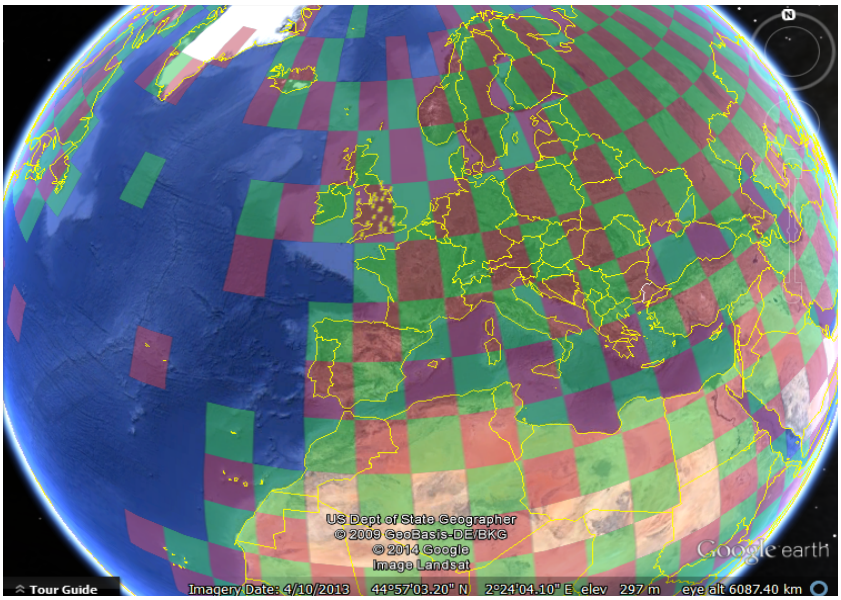

Figure 8 – CRUTEM4 on Google Earth (Osborn and Jones, 2014).

In order to make the CRUTEM4 data more accessible CRU have made a visualisation of the dataset in Google Earth[4]. Through Google Earth the user can explore temperature variability within different grids and at individual station level.

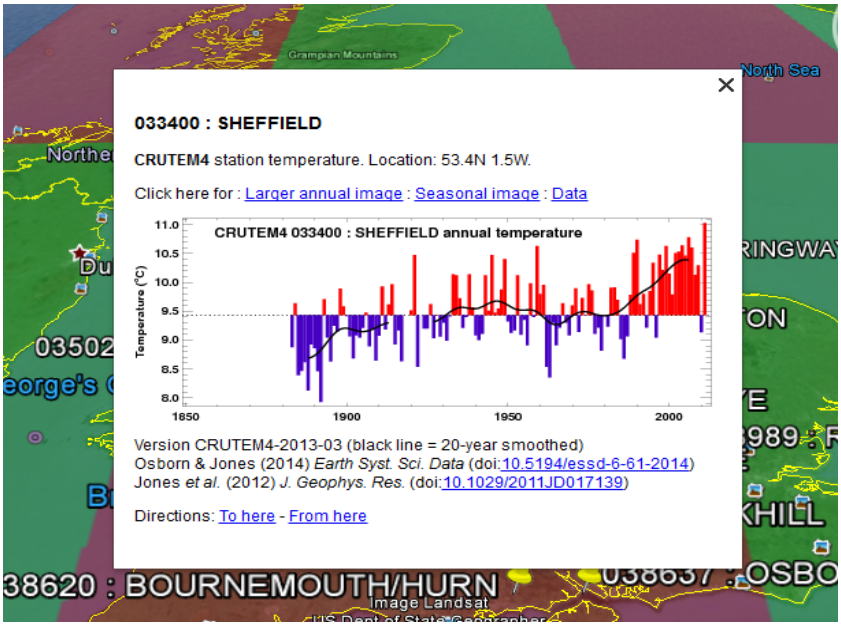

Figure 9 – Google Earth CRUTEM4 graph of the average air temperature near the ground at Sheffield Weston Park since 1883 (Osborn and Jones, 2014).

The horizontal line marks the average temperature for the period 1961-1990 at this location. Blue bars show where the annual temperature was cooler than this period; red bars show where it was warmer.

The graph demonstrates that an increase in the annual average temperature has been observed at Sheffield Weston Park since records began

[4] http://www.cru.uea.ac.uk/cru/data/crutem/ge/CRUTEM4-2013-03_gridboxes.kml

Figure 11 - Managing datasets for climate science, Ian Harris, Climatic Research Unit, University of East Anglia

Figure 10 - The entrance to the Climatic Research Unit, University of East Anglia

Integrity is at the heart of the CRU community and is keenly defended when it is brought into question by those outside the Unit, particularly by some climate 'sceptics' who are perceived to be unable to comprehend how a value system based on integrity can be the driving force behind the Unit's work.

> "A lot of people outside, especially the critics, think we're all sort of evangelistic environmentalists who are frantically trying to change society and everything else to try and save the planet…But for me, I just want to get to the truth, and it's reliable datasets, the best we can do at this point in time, that's my real interest… we have to be unbiased as scientists to do what we do."

> "[We've had] a little bit of money from the World Wildlife Fund for Nature, I've forgotten what it's called now, over the years to one or two things, but when you do work for them--, well, we've also had money from BP and Shell over the years, you give them what you find. You don't give them what they want, you give them what you find and these sceptics can't seem to understand that. It's the same with any government money or research council money, or EU, you're not giving them what they want, you're giving them what the data says."

For the researchers of CRU, the starting point in their work is, therefore, the data that has flowed into the Unit over the years from stations such as Sheffield Weston Park. Their current practice with regard CRUTEM4 involves gathering new data, performing quality checks, and updating the dataset on an annual basis, in co-operation with the Hadley Centre who perform the monthly updates. This data work is relatively routine, yet the output of this process – evidence of warming of the global climate - carries a burden. Humour plays a role in attempting to lighten this load, as seen in a local newspaper cartoon that hangs on the common room wall which features three unconscious CRU researchers and a newspaper headline reporting praise for the Unit alongside the caption "No wonder the poor beggars fainted – this is the first cheerful news that's ever come out of this department".

These outputs of the Unit also, of course, threaten many established interests, some of whom refuse to accept the findings and attempt to generate scepticism about the results within the political and public spheres:

> "The Koch Foundation. It's two brothers…They've been funding loads of right wing organisations all across America in various things and they have massive input into the Senate and the House of Representatives….They want a status quo, basically, they want to continue burning the fossil fuels and everything."

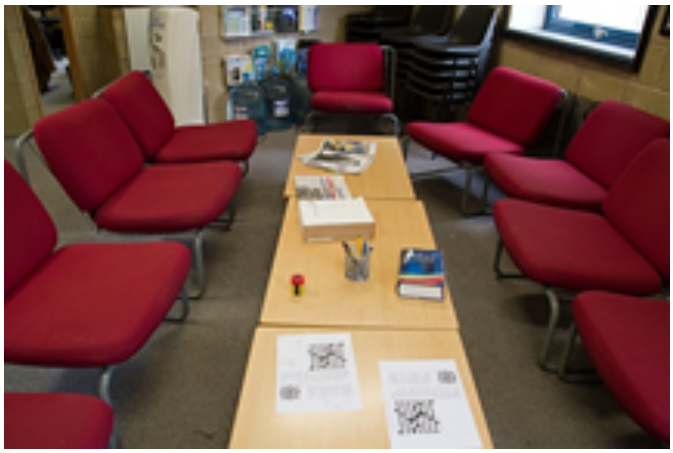

Figure 12 - The common space in the Climatic Research Unit where members regularly meet for crosswords, cooking and chat.

Most notably this threat from climate change 'scepticism' resulted in the hacking of the Unit's email system in November 2009, resulting in the Unit being publicly accused of "deliberately misrepresent[ing] the data, in order to produce results that fit its preconceived views about the anthropogenic warming of the climate" (House of Commons, 2010, p. 12). This accusation was found to be "patently" false by the House of Commons Science and Technology Committee (House of Commons, 2010, p. 48). After investigating the accusations the Committee argued that there was "no case to answer" on this accusation (p. 50). However, the Committee did point to issues around the transparent publishing of data and methods by CRU, yet recognised that practice was "in line with common practice in the climate science community" (p. 50), and a likely breach of the Freedom of Information Act, the responsibility for which they argued lay more with the University of East Anglia than CRU (p. 50).

This public attack on the integrity of the Unit was sorely felt: "people were trying to trash our reputation". Yet in what appears to be typical CRU fashion these events are made light of, being referred as "the pseudo scandal" and "the hoo-hah in 2009" by CRU members. Humour is also used to dilute the symbolic power of their aggressors: "their surname is K-o-c-h, so we might call them Koch [pronounced 'Cock' in UK], but apparently in America they're pronounced "Coke"".

The sense of community within CRU stretches beyond the protective walls of the Unit, expanding both spatially into the networks of global scientific collaboration, and temporally towards the founders of the Unit who provide "a strength and a foundation" and other historical observers of climate including British school teacher Guy Stewart Callendar whose climate records and diaries produced in the 1950s-60s were held for many years at CRU.

The current work of CRU on the CRUTEM4 datasets is based upon principles of trust and collaboration within the international scientific community:

> "A lot of it is on we're trusting their work";
>
> "Everyone's a link in a chain."

Data sharing is an important practice within this scientific community. The gridded datasets produced by the Unit have always been shared freely. Limitations on the sharing of CRUTEM input data have generally been the result of restrictions that were agreed with donor organisations many years ago, although these restrictions were not necessarily taken lightly by CRU members:

> "As someone who works in climate, you often despise these attitudes, but you've got to be able to see why people might not want them to be publicly available."

However, since the events of 2009-10 CRU has taken an active role in pushing for the publication of the data sources used to produce the CRUTEM4 dataset and allow more transparency around its development.

The tensions with climate sceptics and the institutional, managerial and funding pressures faced by CRU generate vulnerabilities for the Unit. Yet, similar to Weston Park weather station, a form of resilience grounded in community, integrity and humour has helped to protect members and the work of the Unit, ensuring for now their continuing contribution the development of the crucial CRUTEM4 dataset.

3.5 Intersecting with historical shipping data: The Old Weather project

Whilst historical gridded datasets such as CRUTEM4 provide scientists with excellent information about the temperature of the global climate since records began, there are many areas of the world where temperature has not been systematically observed for as long as at Sheffield Weston Park. One significant gap in the records is temperature observations over sea. For those climate scientists interested in climate variability over the last two hundred years, one place to beginning looking for missing records are the logs of historic ships that are now stored in archives around the world. Filling in these gaps is crucial for new methods of climate modelling that have developed over the last 5 years such as reanalysis. The journey of data from historic shipping logs, through archives and data recovery projects, finally intersects with the journey of data derived from our Weston Park datum in the HadCRUT4 dataset. This dataset combines the historical land temperature record of CRUTEM4 with the sea surface temperature dataset, HadSST3, which these historic shipping record observations feed into.

The approach that Met Office climate scientists are taking to this process of "fog removal" contrasts with the culture of the Met Office described above. As one scientist described, projects such as ACRE (Atmospheric Circulation Reconstructions over the Earth)[5] and Old Weather[6] are far more organic in structure and "work like Pirates of the Caribbean", as opposed to the institutions they are attached to that "work a bit like the Royal Navy". Facing a difficult funding environment which tends to value "sexy" modelling as opposed to data work, climate scientists in this field have pushed forward their work in historical data recovery through adopting an enterprising form of resilience that values organic and informal networked structures and leadership, serendipity as a method to find and take advantage of opportunities in unusual places, and an iterative development process that "chips away" at the problem of data recovery and rescue.

One project that has come out of this approach is the citizen science Old Weather project, which recruits volunteers to transcribe the weather (and other historic) details of old shipping records. Our initial analysis suggests that there are a multitude of socio-cultural values and practices that motivate people to engage in this project, however in terms of the sustainability and resilience of the project there are echoes of the values and practices observed at Weston Park and the Climatic Research Unit. In particular, the Old Weather project online forum acts as a community hub where members of the project congregate to discuss and find humour both in their Old Weather transcribing and their wider lives:

> "I think it's quite a remarkable place, and I think it's probably quite a large part of what keeps most of the hard core interested and whatever. You know because it's not just you sitting at your computer in isolation transcribing away, you know it's also actually relating to other people who are doing it"

> "It's really nice to know that there's somebody else out there who kind of cares for you as part of the party."

A strong temporal connection is also evident, whereby people are connected historically through the data infrastructure. Similar to the way in which the Weston Park curator experienced a connection to previous curators of the station and climate scientists at CRU discussed historical climate observers, some of the Old Weather volunteers experienced a strong sense of connection with the crews of the ships taking the observations they were transcribing:

> "The events page can be really stunning, actually, it can be breath taking…lovelorn sailors…waxing lyrical about the skies and the sunset…It's all the things that hang around that tiny little comment that I find interesting."

> "I mean on one of the ships I was on…the number of the sick list kept going up, and of course it was the influenza epidemic. And I remember realising that I was really quite anxious about this ship and this crew…And I was thinking this is silly, you know, this is all a very long time ago, whatever's happened has happened. But I realised I was really getting quite anxious about my

[5] www.met-acre.org/
[6] http://www.oldweather.org/

> crew, and you know, hoping that they were all going to, you know, having come through the war that they were actually going to come through the flu epidemic."

For some of the observers it was this historical context of the data they were transcribing that really drew their interest in relation to their daily work on the project. Yet, again, a sense of responsibility in relation both to contributing to society and the accuracy of their transcriptions came through strongly as volunteers spoke about their work:

> "I'm trying to do my bit [for climate change]… when I wasn't transcribing I did feel a bit guilty"

> "The handwriting varies enormously…that's definitely one of the main frustrations is just trying to decipher what it is, and trying to make sure, particularly with the weather records that you're as accurate as possible."

Similar to Weston Park and the Climatic Research Unit, there is, therefore, a strong sense community - both online and historical – that shapes the Old Weather project as a socio-cultural space. This strong community has in turn generated a level of resilience for the project when it has faced difficulties at specific points such as when the project shifted from transcribing Royal Navy ships to American vessels and lost a lot of volunteers for a variety of reasons. However, despite this resilience, the project and the broader data recovery efforts that it is part of do remain vulnerable unless sustainable forms of funding can be found to support their work.

4 Conclusion

These initial findings from The Secret Life of a Weather Datum study highlight the importance of situating our understanding of data infrastructures firmly within the social. Through developing the *data journeys* concept to guide and inform the selection of sites for data collection, the paper has demonstrated the utility of the approach for beginning to build a picture of the "contingent and contested" (Dalton and Thatcher, 2014) relations between people, interconnected in time and space through data infrastructures, that are core to the development and shaping of climate data and knowledge. Our research demonstrates that the values of data are not pre-defined, but emerge and change over time and space as they flow through interconnected but different socio-cultural contexts each with their own conceptual frameworks and value systems. It also begins to draw out the interrelations between local and global spaces and infrastructures; and to ground amorphous 'big data' in local sites and cultures of production.

Through surfacing these socio-cultural aspects, a narrative emerges of the vulnerability of parts of the weather and climate data infrastructure resulting from funding difficulties, neoliberal organisational cultures and policy aims, and tensions with climate change sceptics. Yet, also a deeper and more profound narrative of resilience within the infrastructure emerges based upon values and practices of community, integrity, innovation, and self-actualisation through data work.

In the next stages of the study we will be analysing and theorizing these initial findings in more depth, and also incorporating findings from other research sites, including organisations involved in the re-use of weather data in the financial services industry – specifically within weather derivative markets. Initial readings of our collected data for these sites suggest that similar themes of vulnerability and resilience are also present within this space, albeit in a repurposed form. It is expected that the incorporation of these findings into the analysis will further illustrate the value of the *data journey* approach for exploring the socio-cultural values, practices and power relations that shape (big) data production, processing, distribution and re-use across data infrastructures and contribute to the development of critical empirical work in this field.

References

Appadurai, A. (1986). *The Social Life of Things: commodities in cultural perspective.* Cambridge: Cambridge University Press.

Blanford A and Attfield, S. (2010). Interacting with Information. *Synthesis Lectures of Human-Centred Informatics.* Retrieved from http://www.morganclaypool.com/doi/abs/10.2200/S00227ED1V01Y200911HCI006

Bowker, G. C. (1994). *Science on the run: Information management and industrial geophysics at Schlumberger, 1920-1940*. Cambridge, Mass.: MIT Press.

Bowker, G. C., et al. (2010). Toward Information Infrastructure Studies: Ways of Knowing in a Networked Environment. In J. Hunsinger et al. (eds.), *International Handbook of Internet Research.* New York: Springer.

boyd, d. and Crawford, K. (2012). 'Critical questions for big data', *Information, Communication and Society*, 15(5), 662–79.

Dalton, C. and Thatcher, J. (2014) What does a critical data studies look like, and why do we care? Seven points for a critical approach to 'big data'. *Space and Society Open Site.* Retrieved from http://societyandspace.com/material/commentaries/craig-dalton-and-jim-thatcher-what-does-a-critical-data-studies-look-like-and-why-do-we-care-seven-points-for-a-critical-approach-to-big-data/

Edwards, P. N. (2013). *A Vast Machine: computer models, climate data and the politics of global warming*. Cambridge, Mass: MIT Press.

Furuta, R., Shipman, F.M., Marshall, C.C., Bremner, D. & Hsieh, H. (1997). Hypertext Paths and the World Wide Web: Experiences with Walden's Paths. In *Proc. Hypertext ' 97*, *Southampton, UK*. ACM: New York, NY. 167-176.

Gitelman, L. (2013). *'Raw Data' is an Oxymoron*. Cambridge, Mass.: MIT Press.

Hallam, S. (2013). *Life out of sequence: a data driven history of bioinformatics.* Chicago: University of Chicago Press.

House of Commons Science and Technology Committee (2010). *The disclosure of climate data from the Climatic Research Unit at the University of East Anglia*. London: The Stationery Office Limited. Retrieved from: http://www.publications.parliament.uk/pa/cm200910/cmselect/cmsctech/387/387i.pdf

Jones P.D., Lister D.H., Osborn T.J., Harpham C., Salmon M. and Morice C.P. (2012). Hemispheric and large-scale land surface air temperature variations: an extensive revision and an update to 2010. *Journal of Geophysical Research* 117, D05127.

Kitchin, R. (2014). *The Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequences*. London: SAGE.

Kitchin, R. and Lauriault, T. (2014). Towards Critical Data Studies: Charting and Unpacking Data Assemblages and Their Work. *The Programmable City Working Paper 2.* Retrieved from: *http://papers.ssrn.com/sol3/Papers.cfm?abstract_id=2474112*

Laing, A. (2003). Marketing in the Public Sector: Towards a Typology of Public Services. *Marketing Theory* 3(4). 427-445.

Met Office (2013). Science and technology across the board. *Barometer Online.* Retrieved from: http://www.metoffice.gov.uk/barometer/people/2013-07/science-and-technology-across-the-board on 29th August 2014.

Met Office (n.d.). *PWS CSA 09-14 Annex B1 Appendix A.1 Detailed Output Definitions - PWS Outputs.* Retrieved from: http://www.metoffice.gov.uk/media/pdf/o/s/PWSCG_CSA_SDA_B1_Appendix_A_1_v9.pdf on 29th August 2014.

Museums Sheffield (2012). *130 years of Weston Park Weather Station.* Retrieved from: http://www.museums-sheffield.org.uk/blog/2012/9/130-years-of-weston-park-weather-station.

Osborn T.J. and Jones P.D. (2014). The CRUTEM4 land-surface air temperature dataset: construction, previous versions and dissemination via Google Earth. *Earth System Science Data* 6, 61-68.

Reich, S., Carr, L., De Roure, D., and Hall, W. (1999). Where have you been from here? Trials in hypertext systems. *ACM Computing Surveys (CSUR),* 31(4es), 11.

Ruppert, (n.d.). Centre for the Study of Invention and Social Process Current Research, Goldsmiths University of London. Retrieved from http://www.gold.ac.uk/csisp/research/current-research on 29th August 2014.

Sands, A. et al. (2012). Follow the Data: How astronomers use and reuse data. In *Proc. ASIST 2012*, October 28-31, 2012, Baltimore, MD, USA

Shipman, F.M., Furuta, R., Brenner, D., Chung, C. and Hsieh, H. (2000). Guided Paths through Web-Based Collections: Design, Experiences, and Adaptations. *Journal of the American Society for Information Science*, 51(3), 260-272.

Star, S. L. (1999). The Ethnography of Infrastructure. *American Behavioural Scientist.* 43(3). pp. 377-391.

White, R. W. & Huang, J. (2010). Assessing the Scenic Route: Measuring the Value of Search Trails in Web Logs. In *Proceedings of the 33rd international ACM SIGIR conference on Research and development in information retrieval*, 587-594.

WMO (n.d.). The Global Telecommunication System (GTS). World Meteorological Association. Retrieved from http://www.wmo.int/pages/prog/www/TEM/GTS/index_en.html on 29th August 2014.

Table of Figures

Table of Tables

No table of figures entries found.