Mapping Data Journeys: Design for an interactive web site

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

Abstract
In this poster we present an overview of our approach to researching the information infrastructure for weather and climate data in the UK, which aims to map the various data journeys of an individual datum through this big data environment, and to uncover the socio-cultural values that shape the data and processes involved in data production, transformation, use and reuse. We then illustrate how we will disseminate our findings through the design of a forthcoming interactive web site, which presents the data journeys using a path/map metaphor, enabling the exploration of four interconnected case studies and several cross-cutting themes, in a way that is both flexible for the user and expandable, as the research progresses further.
Keywords: data journey; thematic analysis; socio-cultural values; research dissemination; interaction design.
Citation: Goodale, Paula; Bates, Jo; Lin, Yuwei. (2015). Mapping Data Journeys: Design for an interactive web site. 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.
Research Data: In case you want to publish research data please contact the editor.
Contact: p.goodale@sheffield.ac.uk, jo.bates@sheffield.ac.uk, ylin21@ucreative.ac.uk

1 Introduction

Information infrastructures for big data, such as the global network for weather and climate data, present several levels of complexity that can make them difficult to comprehend for those not directly involved in the core practices of data production, transformation, use and reuse. Several techniques have evolved to aid the understanding of these data environments. Infrastructural inversion (Bowker, 1994; Star, 1999; Edwards, 2013) takes a 'bottom-up' approach to understanding technologies, organisations and their associated socio-technical relations which enable knowledge infrastructures to function. Ethnographic research methods can generate rich pictures of the information infrastructure, giving insights into the socio-cultural, political and economic values that shape both the infrastructure and the data practices that take place within it (Star, 1999). A related approach is the 'Follow the Datum' method which begins with a knowledge output (e.g. a research paper) and unpacks how that research was carried out, starting with the original datasets, and identifying processes and analyses that have been undertaken in completing the work (Sands, et al, 2013).

The [project name removed for review] uses infrastructural inversion as a means of investigating weather and climate data in the UK, in particular four selected case studies focusing on institutions of state (data production and transformation), institutions of science (climate), institutions of the market (financial sector), and citizen engagement (crowd-sourced data contributions). Our goals in this research are three-fold:

- To map the journey of a single weather datum from its initial production through the systems and processes associated with the selected cases
- To uncover the socio-cultural values that shape the datum and the information practices within this information environment
- To disseminate these findings via an interactive web site, that allows users to 'follow the datum' through the four case studies and investigate core elements of the information infrastructure in more detail

In this poster we present an overview of our approach to carrying out this research, and illustrate the design of our forthcoming interactive web site, which maps out data journeys that are evident within the weather and climate data infrastructure.

2 Methods

Our primary method of investigation has been a series of semi-structured interviews with 20 participants across several organisations involved in the four case studies identified above. Interviews were tailored to some degree for each participant to reflect the different roles and processes in which they are each involved, in essence, providing pieces of a jigsaw puzzle that develops a picture of the overall infrastructure for weather and climate data. Interviews did however incorporate common themes, including: a description of the participant's role and data use; discussion of motivations and challenges in carrying out their work; exploration of issues such as economic and political factors affecting their work, (for example, government policy on open data); and, exploration of the nature of relationships with people and organisations related to their work.

Additional methods included observations of participants doing their work, and/or their working environment (comprising field notes and photographs), observation of a conference presenting results of a major review of climate science, digital ethnography based upon a content analysis of Twitter data and online discussion forums (especially for case 4 on citizen science), and documentary evidence of various kinds, including organizational web sites, policy documents, corporate publications, research reports and papers.

Primary data analysis is based upon a thematic analysis of the interview and observation data. Three core themes have emerged: attitudes, values and beliefs; valuable activities; and, social relations. From each of these several more specific sub-themes have also been identified, which provide a lens onto the important socio-cultural values inherent within this information environment. Additionally, we have mapped out the journey of a weather datum, originating at a local climatic weather station, and followed it through each of the cases as it is transformed, and used in different operational and research contexts.

3 Web Site Design

The act of following a data journey correlates well to the 'path' (or 'trail') metaphor that has been widely used in web and other hypertext systems, as a means of traversing inter-related content. A path can be described as a collection of information items (e.g. documents, web pages, images) that are organised in some way, connected, and contexualised with additional related information. The essential components of a 'path' are therefore 'nodes' (information items), 'connections' which represent the relationships between items and which may be directional, and 'annotations' that provide context and narrative.

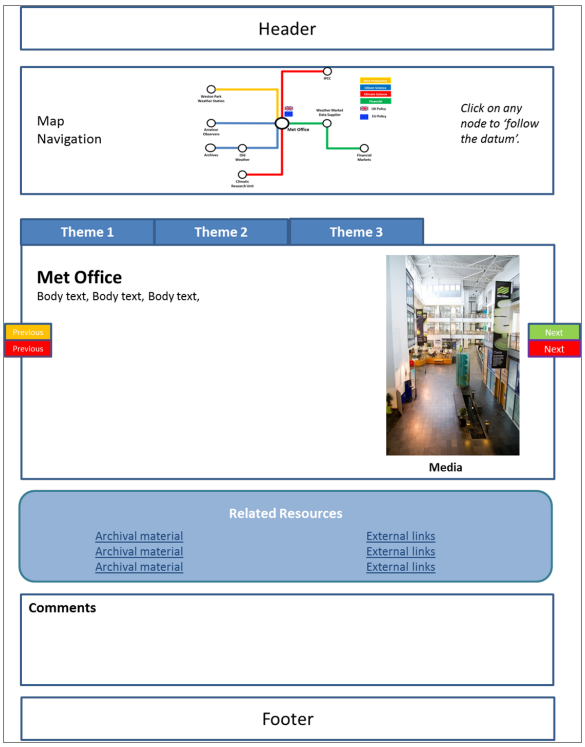
Several research projects have utilised the 'path' metaphor as a means of navigating web-based content. Walden's Paths was designed for teachers to create content based upon connected web pages, with annotations and directional indicators to aid students as they followed the thread (Shipman, et al, 2000). Projects that have incorporated 'paths' as a primary or secondary aid for learning, navigation or creativity in the cultural heritage domain have included StoryBank (Frohlich and Rachovides, 2008), Cultura (Agosti, et al, 2013), and PATHS (Hall, et al, 2012). Making connections between items has also been suggested as a means of surfacing narratives through the rigid database structures that often underlie online cultural collections, thereby supporting a more cohesive interaction experience. Opportunities to develop multiple narratives through the same information space may arise, allowing for different interpretations to co-exist (Manovich, 1999). The path metaphor therfore seems to be well-suited to meet our needs for following a data journey, understanding people and organisations involved in that journey, and exploring the multiple socio-cultural values and themes emerging from our analysis of a complex infrastructure.

For the [project] web site[1] (Figure 1), we have applied the path metaphor in the form of a familiar 'tube' (underground/metro) map (Figure 2). This allows us to illustrate the four cases, and the points at which they intersect, to provide a holistic view of the journey of our selected weather datum. Nodes (or stations) relate to organisations that the datum passes through on its journey, with each represented in detail in the node's web page (Figure 1 illustrates an outline of the Met Office node).

The node web page is initially accessed via a map navigation scheme, and from this first node, the user can follow the datum back and forth through the different 'lines', which represent the four cases, using the colour-coded 'previous' and 'next' navigation buttons, located on the left and right of the page. Each node page may contain a number of tabbed content sections, allowing for exploration of the node content by theme or case. Additional background information on the government policy areas relevant to each node is indicated and accessed via UK and EU flag symbols on the map navigation.

[1] http://www.lifeofdata.org.uk

Dotted lines on the navigation map represent potential extensions to the case study that have not yet been investigated in detail. In this way, the navigation can be expanded, as needed, with addition organizational nodes (stations), or even additional cases (new lines), providing a high degree of flexibility in future development of the web site.



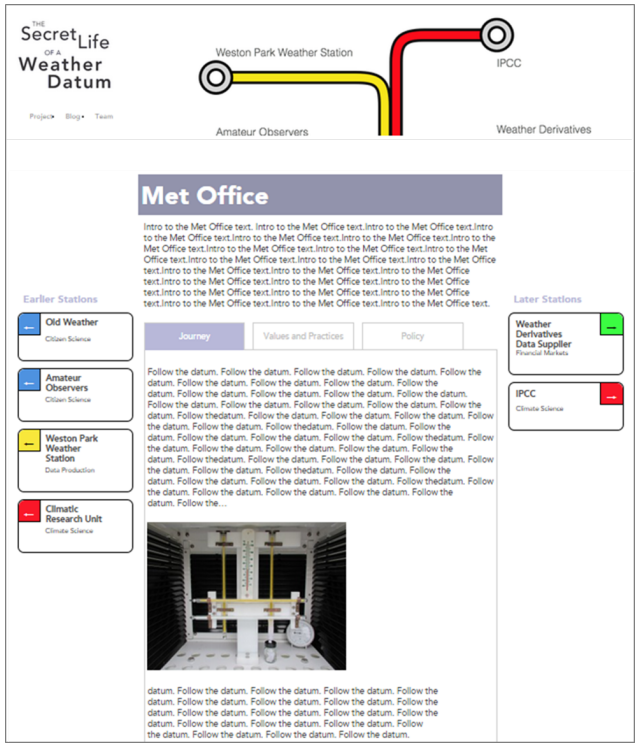


Figure 1. Outline design for The Secret Life of a Weather Datum interactive web site (left)

Figure 2. Initial screen-shot for The Secret Life of a Weather Datum interactive web site (right)

Content for each node comprises a summary description of key findings, illustrated with media such as photographs, videos and audio clips. This content is suitable for a general audience, and provides a narrative relating to the data journey, as well as an overview of key socio-cultural themes identified from the research data. Users can explore further by following the 'related resources' links to the full versions of the original research data, offered with a Creative Commons licence[2] via an online archive, and also to related content on external web sites. These links allow for much deeper exploration of content related to each node and their associated cases, including more comprehensive and scholarly research findings, background information from organizational web sites, and documentary evidence from relevant reports and other publications.

The web site is currently in development. An important stage of this development will be a card sort exercise with potential end users, to determine the optimal organization of content within the nodes and their associated tabs. A usability testing phase will also be undertaken before launch, in which users will be encouraged to undertake simple tasks (e.g. to follow the datum on its journey), and to feedback on the usability of the site, in particular, their perceptions of the path/map metaphor, ease of navigation, and understanding of the data journey.

4 Conclusion

We have shown how the findings of a research project based upon infrastructural inversion methods, can be presented and disseminated to a potentially wide and diverse audience through the means of an interactive web site incorporating a map-based navigation. The map enables multiple paths to be taken through the information environment, allowing for exploration of case studies, cross-cutting themes, and

[2] https://creativecommons.org/licenses/by-nc/3.0/

the policy context, as well as providing the flexibility to expand coverage, as further research is carried out. Our poster will show more detail from the web site, and a demo may also be given, if required.

References

Agosti, M., Conlan, O., Ferro, N., Hampson, C. & Munnelly, G. (2013). Interacting with digital cultural heritage collections via annotations: the CULTURA approach. In *Proceedings of the 2013 ACM symposium on Document Engineering,* pp. 13-22. ACM.

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

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

Frohlich, D. M. & Rachovides, D. (2008). Using Digital Stories for Local and Global Information Sharing. In *Proc. CHI'08*, Florence, Italy, 5-6th April 2008. New York: ACM.

Hall, M,. Goodale, P., Clough, P. & Stevenson, M. (2012). The PATHS System for Exploring Digital Cultural Heritage. In: Mills, C., Pidd, M. & Ward, E. (Eds.), *Proc. Digital Humanities Congress 2012*, *Studies in the Digital Humanities*. Sheffield: HRI Online Publications. Available: http://www.hrionline.ac.uk/openbook/chapter/dhc2012-hall4

Manovich, L. (1999). Database as Symbolic Form. *Convergence: The International Journal of Research into New Media Technologies*, 5(2), pp. 80-99.

Sands, A., Borgman, C. L., Wynholds, L. &. Traweek, S. (2012). Follow the Data: How astronomers use and reuse data. In *Proc. ASIST 2012*, October 28-31, 2012, Baltimore, MD, USA

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

Table of Figures