The Disciplinary Shaping of Research Data Management Practices

Winnie Tam[1], Jenny Fry[1] and Steve Probets[1]
[1] Loughborough University

Abstract

Aim

The aim of this research is to focus on geography and its sub-disciplines with the intention of exploring how the nature of discipline shapes current and potential research data management practices and, in turn, how disciplines themselves are being reshaped by changes in data creation, use and management.

Design

The research is in two sequential phases. Phase 1 consists of a scoping stage which includes a web-based study using different techniques, such as link analysis and bibliometrics, and interviews with data management experts. Phase 2 is primarily based on a series of interviews with researchers, investigating researchers' research practices and attitudes of data management.

Findings

By the time of the conference preliminary results from phase 1 of the research will be available for reporting.

Value

This research will provide a better understanding of how the complexity of research data, and so the challenges for improving research data management, are grounded in the underlying nature of disciplines/sub-disciplines. As a by-product of this understanding, it will enhance the conceptual theory of disciplinarity through detailed analysis of changes around research data management in geography and its sub-disciplines.

Keywords: research data management, data practices, disciplinary culture, data creation, data use, data sharing

Citation: Tam, W., Fry, J., & Probets, S. (2014). The Disciplinary Shaping of Research Data Management Practices. In *iConference 2014 Proceedings* (p. 721–728). doi:10.9776/14338

Copyright: Copyright is held by the authors.

Contact: w.w.t.tam@lboro.ac.uk, j.fry@lboro.ac.uk, s.g.probets@lboro.ac.uk

1   Context

Research data management has become an increasingly important topic in information science (Borgman, 2012; Corrall, 2012). Data is a crucial part of research. Today many research councils and funders worldwide, such as the funding councils and the Wellcome Trust in the UK, the National Science Foundation and National Institutes of Health in the US, and the Australian Research Council require data to be managed according to best practices and standards. As a consequence, Higher Education Institutions and researchers are under pressure to manage their data better. In the UK, the British government has invested £10 million to create and support the Open Data Institute which aims to promote open data culture and drive innovation and growth in the UK. It is evident that policy-makers and research funders believe that effective data management is the foundation for good research and that opening up data will benefit society. In addition, a variety of benefits to researchers that arise from effective management of research data have been identified, e.g. verification of research, stimulating new collaborations, transferring knowledge to industry, sharing and reuse of data, reducing future preservation costs etc. (Fry et al, 2009; DCC, 2013). However, not all these benefits are relevant to all disciplines and all types of data.

One of the key challenges in managing research data is the complexity and diverse nature of data, coupled with diverse data practices across disciplines. The types of data produced within disciplinary

communities can vary greatly, from static numeric and text-based data to dynamic multimedia data. Consequently, even the concept of data is difficult to define (Borgman, 2012).

There are an increasing number of studies of research data management in the disciplinary context. For example, in the UK, the Research Information Network (RIN) and Digital Curation Centre (DCC) SCARP project have investigated disciplinary attitudes and approaches to data curation (Lyon et al, 2010; RIN, 2008). All projects suggest there are not only variations in kinds, file formats, value and long term viability of research data in different disciplines but also significant differences in data practices and culture of sharing data. For example, according to the RIN report (2008), re-use of data and data sharing is a norm in astronomy, while in climate science, the culture of data sharing and publication is not strong. It is evident that practices of data creation, use and sharing vary widely across subject disciplines and their sub-disciplines. However, these differences are actually shaped by something more fundamental, the nature and the complexity of disciplines. Thus, this research will attempt to provide insights into the process and the practices of knowledge communities and explore how these shape research data management practices.

The nature and use of research data is strongly related to disciplines (Pryor, 2009; Cragin et al. 2010). Effective support of research data management, therefore, is dependent on a detailed understanding of the characteristics of disciplines and data creation and use within disciplinary communities. Terms such as, 'specialization', 'fragmentation', 'hybridity' and 'fluidity' have been used to describe the dynamic nature of disciplinarity (Klein, 1996; Dogan & Pahre, 1990). As knowledge becomes increasingly interdisciplinary, some disciplines borrow concepts and methods from another discipline and become hybrid disciplines or sub-disciplines. Due to the dynamic nature of disciplines, data are highly specialized, fragmented and diverse. As knowledge production becomes more and more interdisciplinary in nature the hybridity and fluidity of disciplines also create complexity in relation to research data management.

Disciplines provide basic structures for organising knowledge and research in Higher Education Institutions. Becher and Trowler (2001) developed a matrix of disciplinary groupings, namely hard-pure, soft-pure, hard-applied and soft-applied.
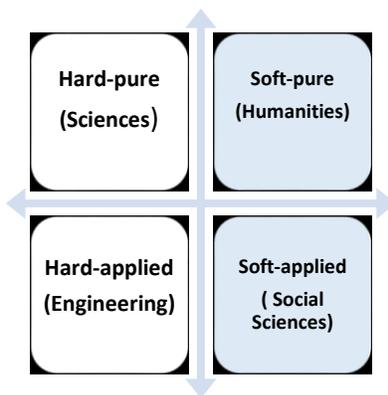

Figure 1: Becher's matrix of disciplinary cultures (1987)

Becher and Trowler (2001) use the terms hard/ soft, pure/ applied to refer to intellectual differences in disciplines, e.g. research problems, research objects, and methods. They also use terms such as convergent/divergent, rural/urban to refer to differences in social structures, e.g. community culture, communication patterns, and reward systems. For instance, Becher and Trowler (2001) classify history as a soft-pure rural discipline, which typically uses interpretative methods and in which researchers tend to work independently; while high energy physics is classified as a hard-pure urban discipline, which typically favours quantitative methods, and large scale collaborations. It appears that history is more likely to generate small scale data, e.g. interview data, than the large computational datasets often generated in disciplines such as high energy physics. In general, computational datasets are usually highly structured

and anonymised, whereas interview data is more personalised and not easily anonymise (Borgman, 2012). Thus the character of disciplines is likely to affect approaches to managing data and attitudes towards data sharing.

It seems likely, therefore, that the nature of disciplines, e.g. research methodologies, research problems, processes and practices within and across disciplines/ sub-disciplines, shape the nature of the data created, ways in which data are manipulated and stored, the possibility of data sharing and reuse, and what constitutes effective data management practices. It is essential, therefore, to understand how the nature of disciplines shapes research data management practices. In addition, having a deeper understanding of the changing nature of disciplines, may also reveal how the research data management agenda may bring potential changes to disciplines. This understanding is essential if appropriate measures can be developed to increase awareness, encourage best practices and further develop appropriate strategy, services and infrastructures for research data management.

In order to develop an understanding from a disciplinary perspective, one of the common approaches is to compare across a number of disciplines. However, in this research, comparisons within a single discipline are given more weight. Studies of discipline at a broad level may falsely represent them as unified. As mentioned earlier, the number of studies of research data management in the disciplinary context has increased. Most of the studies that have been done are focused on the Sciences, such as, astronomy, systems biology, genomics, with a few on Arts, such as, classics or artistic research. It appears that relatively less research has been conducted on social sciences. Most of the results of these studies show that much diversity exists even within a single disciplines e.g. at the sub-discipline level (Key Perspectives, 2010; RIN, 2008). Thus, this research focuses on sub-disciplines in one single social science discipline, which is geography.

Geography is a well-established discipline in universities. Today, geography departments can be found in most parts of the world, for example, the US, Australia, South America, China, Japan, India (International Geographical Union, 2013). Geography is chosen not only because there is limited research of research data management in this subject, with a few exceptions, e.g. there are some studies of awareness/activities of curating geospatial data (McGarva et al, 2009; Bose & Reitsma, 2005), but also because of its internal complexities.

Geography has a clear division between the two main branches in geography: physical geography and human geography. Physical geography concerns the sciences of physical landscapes and environmental processes. It belongs to the 'hard' side of Becher's (1987) matrix of disciplinary cultures (Fig. 1). In contrast, human geography belongs to the 'soft' side of Becher's (1987) matrix. It concerns human activities on the Earth, including cultures, societies and economies from a spatial perception. Geography is a dynamic discipline. Both physical geography and human geography consist of a wide range of sub-disciplines, reflecting diverse research problems in the discipline (fig.2 & 3). In addition, Geography has had a close relation to other disciplines and has developed some well-established interdisciplines, such as quaternary science and geo-archaeology etc.
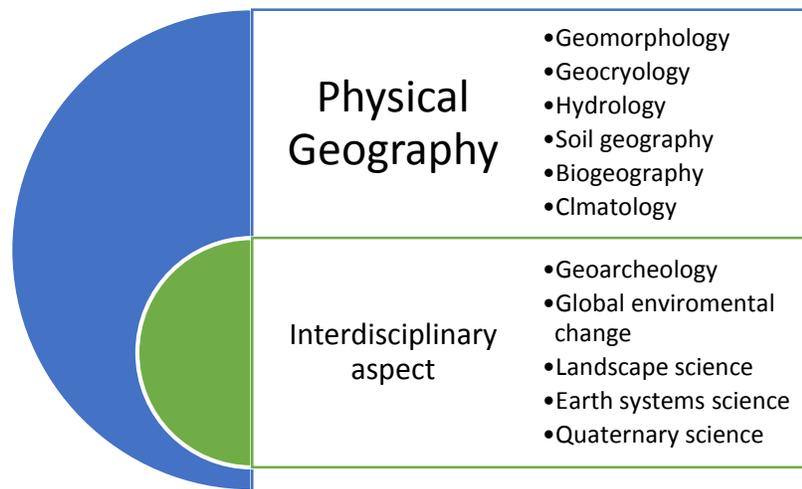
Figure 2: Sub-disciplines in physical geography (adapted from Matthews & Herbert, 2008)
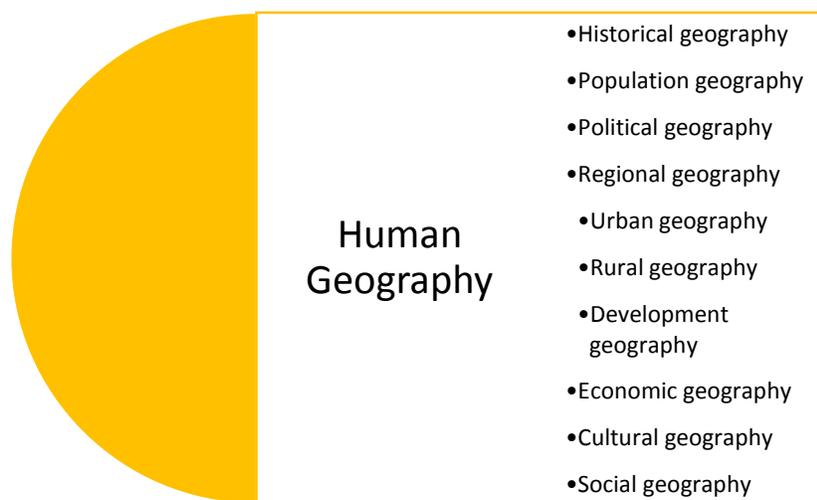

Figure 3: Sub-disciplines in Human geography

Today, geographers employ a wide range of methods for data collection, such as, secondary data, questionnaire surveys, interviews, visual images, observations and measurements in the field, numerical modelling, laboratory methods, spatial modelling and remotely sensed images (Clifford et al, 2010). It is noted that some cultural and historical geographers collect written text and arts images as their data, whereas soil geographers collect remote sensing data (Clifford et al, 2010; de Paul Obade & Lal, 2013). Broadly one could claim that human geographers collect qualitative data, while physical geographers collect quantitative data. However, the nature of a discipline is dynamics and changing. Research practices in geography are being reshaped by technological innovations. Today, contemporary human geography not only collects qualitative data but also quantitative data. Economic geographers analyse quantitative spatial data in addition to, feminist geography has adopted GIS technology for qualitative studies (Elwood & Cope, 2009; Kwan, 2002). It is evident that the nature of geography as a discipline has shaped and is shaped by the very nature of the data it uses.

In general, geography has a clear division between hard and soft research tradition, it also has strong interdisciplinary links with other academic fields. Thus, by examining this single discipline, it will

be possible to compare research practices at the sub-disciplinary level and gain understandings of how cultural characteristics of a discipline shape, and are shaped by research data management.

2   Research Aim

The overall aim of this PhD research is to focus on geography and its sub-disciplines with the intention of exploring how the nature of discipline shapes current and potential research data management practices and, in turn, how disciplines themselves are being reshaped by changes in data creation, use and management. The outcome of this understanding will be insights into how to achieve understand how effective data management that is sensitive to cultural identities.

3   Research Questions

1. What changes are happening in data practices and research data management policy in Geography (eg. increasing scale of digital data, open data, sharing data, big data, institutional policy)?
2. How are responses to changes in data shaped by the nature of (geography as a) discipline? What disciplinary factors influence research data management practices?
3. How are disciplines themselves (or how is geography) being reshaped by changes in the very nature of data, data creation, use and management?
4. How can effective data management policies sensitive to cultural identities be achieved?

4   Methodology

The research is in two sequential phases. Phase 1 consists of a scoping study which includes 1) a web-based study using different techniques, such as link analysis and bibliometrics, and 2) interviews with data management experts. Phase 2 is primarily based on a series of interviews with researchers.

4.1   Phase 1

The purpose of phase 1 is to understand the nature of geography and its-sub disciplines (e.g. to identify its research problems and methods) and its data (e.g. data types); to identify current activities, key issues in research data management in geography; and to identify suitable participants for interviews in phase 2.

The following activities take place:

*   a web-based study of geography departmental websites, supplemented with a link analysis study of disciplinary / interdisciplinary interpersonal network (e.g. based on links between researchers and researcher groups) and a small scale bibliometric study of geography (topic analysis)
*   Five interviews with data management experts (UK-based), e.g. staff from the UK Data Archive, to discover the key issues affecting research data management.

4.1.1   Phase 1: Web- based study

The web-based study examined research group leaders' profiles on departmental websites and their most cited publication in the last five years. This part of the study provides an opportunity to explore the relationship between research data and sub-disciplines of geography and sub-disciplinary differences in research data,

The study examines the following:

a) Researcher's profile, e.g.
    *   the research / methods interest of the researcher
    *   types of publication (e.g. books, journals)
b) and the most cited publication in the last 5 years, e.g.
    *   the topic of the research

- the research method employed
- the data in the research, including the creation, use and representation of data (e.g. images, histograms)

The sample consists of research group leaders from 15 geography departments, i.e. five top-ranking departments in the 'geography and environmental studies' unit of assessment in the Research Assessment Exercise 2008, five middle-ranking departments, and five low-ranking departments. Research group leaders are chosen to be examined because they usually have relevant research experience to the focus of the research groups. In addition, the sampling approach can cover a range of universities with diverse research strength, such as, research intensives universities vs. teaching universities.

As part of the web-based study, a link analysis is used to investigate the relationship between research groups, departments and interdisciplinary interpersonal networks in geography. The nature of this part of the study is very exploratory. Different sets of seed URLs, such as, a list of geography department home pages URLs or a list of individual researcher webpages URLs, are used to address different objectives. For instance, in order to explore the relationship between research groups, a list of research group' webpages URLs are used as the seeds URL.

A small bibliometric study will also be used as part of the web-based study. Bibliometrics has been defined as the quantitative study of published literature (Hood & Wilson, 2001). Bibliometrics have been used extensively in the study of scholarly communications, e.g. to explore collaboration patterns in academic networks (Velden et al, 2010; Melin & Persson, 1996; Subramanyam,1983) and emerging research topics of disciplines (Glanzel, 2012; Van den Besselaar & Heimeriks, 2006). In this research, Web of Science will be used as the data source, focusing on identifying relevant sub-disciplines and exploring relevant scholarly practices in geography more broadly.

4.1.2   Phase 1: Interviews with data management experts

A small number of interviews will also be conducted to complement the web-based study in order to gain a broader understanding of current trends and issues in research data management. Five interviews with UK based data management experts will be conducted to discover their views on current activities, key issues in research data management in geography.

4.2   Phase 2

The purpose of phase 2 is to investigate researchers' research practices, understandings and attitudes towards data management using interviews. A total of 32 interviews (8 researchers from 4 sub-disciplines = 32) with academic researchers in geography will be conducted in this phase.

By the time of the conference, preliminary results from phase 1 of the research will be available for reporting.

5   Expected contribution to current understanding

This research will provide a better understanding of how the complexity of research data, and so the challenges for improving research data management, are grounded in the underlying nature of disciplines/sub-disciplines. As a by-product of this understanding, it will enhance the conceptual theory of disciplinarity through detailed analysis of changes around research data management in geography and its sub-disciplines. In addition, it will contribute to the practical implementation of good research data management practices by deepening our understanding of the nature of disciplines.

6   References

Becher, T. (1987). The disciplinary shaping of the profession. In B. R. Clark (Ed.), *The academic profession.* Berkeley: University of California Press.

Becher, T., & Trowler., P. R. (2001). *Academic tribes and territories: intellectual enquiry and the culture of disciplines.* Philadelphia, PA: Open University Press.

Borgman, C. L. (2012). The conundrum of sharing research data. *Journal of the American Society for Information Science and Technology*, *63*(6), 1059–1078. doi:10.1002/asi.22634

Bose, R., & Reitsma, F. (2005). *Advancing Geospatial Data Curation.* Institute of Geography. The School of Geosciences.The University of Edinburgh. Retrieved from https://www.era.lib.ed.ac.uk/handle/1842/1074

Clifford, N., French, S., & Valentine, G. (2010). *Key methods in geography.* London: Sage Publications.

Corrall, S. (2012). Roles and responsibilities: libraries, librarians and data. In *Managing research data* (pp. 105–133). London: Facet.

Cragin, Melissa H, Palmer, C. L., Carlson, J. R., & Witt, M. (2010). Data sharing, small science and institutional repositories. *Philosophical transactions. Series A, Mathematical, physical, and engineering sciences*, *368*(1926), 4023–38. doi:10.1098/rsta.2010.0165

De Paul Obade, V., & Lal, R. (2013). Assessing land cover and soil quality by remote sensing and geographical information systems (GIS). *CATENA*, *104*(0), 77–92. doi: http://dx.doi.org/10.1016/j.catena.2012.10.014

Digital Curation Centre. (2013). Digital Curation Centre. Retrieved from http://www.dcc.ac.uk/

Dogan, M., & Pahre, R. (1990). *Creative margznality: Innovation at the intersections of social sciences.* Boulder, CO: Westview Press. 8

Elwood, S., & Cope, M. (2009). Introduction: Qualitative GIS: Forging mixed methods through representations, analytical innovations, and conceptual engagements. In *Qualitative GIS: A mixed methods approach* (pp. 1–13). London: Sage Publications.

Fry, J., Lockyer, S., Oppenheim, C., Houghton, J., & Rasmussen, B. (2009). *Identifying benefits arising from the curation and open sharing of research data produced by UK Higher Education and research institutes.*

International Geographical Union. (2013). Retrieved from http://www.igu-online.org/

Key Perspectives. (2010). *Data dimensions: disciplinary differences in research data sharing, reuse and long term viability. SCARP Synthesis Study.* Digital Curation Centre. Retrieved from http://www.dcc.ac.uk/projects/scarp

Klein, J. T. (1996). *Crossing boundaries: knowledge, disciplinarities, and interdisciplinarities.* Charlottesville, Va.: University of Virginia Press.

Kwan, M. P. (2002). Feminist Visualization: Re-Envisioning GIS as a Method in Feminist Geographic Research. *Annals of the Association of American Geographers*, *92*(4), 645–661. doi:10.2307/1515293

Lyon, L., Rusbridge, C., Neilson, C., & Whyte, A. (2010). *DCC SCARP: disciplinary approaches to sharing, curation, reuse and preservation - final report.* Retrieved from http://www.dcc.ac.uk/sites/default/files/documents/scarp/SCARP-FinalReport-Final-SENT.pdf

McGarva, G., Morris, S., & Janée, G. (2009). *Technology watch report: preserving geospatial data.* Digital Preservation Coalition.

Pryor, G. (2009). Multi-scale data sharing in the life sciences: some lessons for policy makers. *The International Journal of Digital Curation*, *4*(3), 71–82.

Research Information Network (RIN). (2008). *To share or not to share: publication and quality assurance of research data outputs - Report commissioned by the Research Information Network.* Retrieved from http://www.rin.ac.uk/our-work/data-management-and-curation/share-or-not-share-research-data-outputs

7   Table of Figures