IJDC | *Peer-Reviewed Paper*

Research Transparency: A Preliminary Study of Disciplinary Conceptualisation, Drivers, Tools and Support Services

Liz Lyon
School of Computing and Information
University of Pittsburgh

Wei Jeng
School of Computing and Information
University of Pittsburgh

Eleanor Mattern
School of Computing and Information
University of Pittsburgh

Abstract

This paper describes a preliminary study of research transparency, which draws on the findings from four focus group sessions with faculty in chemistry, law, urban and social studies, and civil and environmental engineering. The multi-faceted nature of transparency is highlighted by the broad ways in which the faculty conceptualised the concept (data sharing, ethics, replicability) and the vocabulary they used with common core terms identified (data, methods, full disclosure). The associated concepts of reproducibility and trust are noted. The research lifecycle stages are used as a foundation to identify the action verbs and software tools associated with transparency. A range of transparency drivers and motivations are listed. The role of libraries and data scientists is discussed in the context of the provision of transparency services for researchers.

Received 20 October 2016 ~ *Revision received* 23 February 2017 ~ *Accepted* 23 February 2017

Correspondence should be addressed to Professor Liz Lyon, School of Computing and Information, University of Pittsburgh, 135 North Bellefield Avenue, Pittsburgh, PA, USA 15260. Email: elyon@pitt.edu

An earlier version of this paper was presented at the 12th International Digital Curation Conference.

The *International Journal of Digital Curation* is an international journal committed to scholarly excellence and dedicated to the advancement of digital curation across a wide range of sectors. The IJDC is published by the University of Edinburgh on behalf of the Digital Curation Centre. ISSN: 1746-8256. URL: http://www.ijdc.net/

Introduction

Research transparency is gaining traction as a key objective for many stakeholders engaged in scientific endeavours. As a concept, however, transparency encompasses many different facets and dimensions. This paper is based on a preliminary study of researchers from different disciplines. It seeks to explore the research community's understanding of the concept of research transparency and begins to articulate the language, vocabulary and terminology associated with this concept. The study utilizes the lifecycle as a grounding framework or construct for exploring a theoretical conceptualisation and practical behaviours towards research transparency by faculty researchers across different disciplines. In particular we focus on identifying the critical action verbs aligned with and embedded within the various stages of the lifecycle, which, when considered as a whole, encompass the critical research practices required to assure research transparency in open science. This paper aims to inform more substantive work on research transparency. It begins with a brief contextual framing for the study, followed by a description of the methodology used, an exposition of the results, a discussion section and the identification of next steps.

Contextual Framing

The term 'transparency' has been applied in a range of contexts by diverse research stakeholders, who have articulated and framed the concept in a number of different ways. At the global and national level, transparency has been identified as *a principle* by thirty countries (OECD, 2007) and by the G8 countries in their Open Data Charter (Gov.UK, 2013), and as *an action* for departments and government agencies in the United States Memorandum on Transparency and Open Government (Holdren et al., 2009). Federal funding agencies have cited transparency in planning statements for more rigorous research (NIH, 2015). Transparency has been framed *in policy* by the Royal Society (2012) and Research Councils UK (2015). Transparency has been described as *a rationale* for open science and open data (OECD, 2015) and as *the bedrock* for "progress of science in the modern era" (ICSU, 2015). Professional organisations, such as the Federation of American Societies for Experimental Biology have published recommendations which position transparency as *a parameter* (FASEB, 2015), whilst the American Political Science Association has recommended higher transparency *standards* (APSA, 2012). A number of scholarly publishers have included transparency *statements* within their policies e.g., PLOS Competing Interests Policy[1] and the British Medical Journal (BMJ)[2]. Transparency has also been defined as *a value* by Etzioni (2010), who recognises regulatory requirements for disclosure, and by Vayena, Salathe, Madoff and Brownstein (2015), who discuss the ethical challenges of big data. In the UK Academy of Medical Sciences Symposium Report on Reproducibility and Reliability of Biomedical Research (2015), greater openness and transparency is listed as *a measure* for both methods and data. Lyon (2016) positions transparency as *a third dimension* of open science and notes the inter-dependency and

1 PLOS Competing Interests Policy: http://journals.plos.org/plosgenetics/s/competing-interests
2 British Medical Journal: http://www.bmj.com/about-bmj/resources-authors/forms-policies-and-checklists/transparency-policy

connectedness with other related concepts and terms such as reproducibility, which has been examined in some depth by Stodden et al. (2013). The 'confusion of terms' associated with reproducibility, repeatability and replicability has been raised by Kenett and Shmueli (2015), indicating the semantic complexities of this area.

Depending on the thematic area and discipline, the transparency concept has been further unpacked and interpreted in different ways. Miguel et al. (2014) describe transparency within three core practices in social science: in design (disclosure), in intentions (preregistration) and in analytics (open data and materials). Moravcsik (2014) describes transparency as *the cornerstone* of social science, with qualitative political science as his focus. He posits that there are three dimensions of research transparency: data transparency (access to the evidence or data); analytic transparency (access to evidence which supports a claim); and production transparency (access to information about methods). An ethics perspective on digital disease detection (DDD) is presented by Vayena et al. (2015), who identify three categories linked to transparency: context sensitivity (privacy laws), methodology (personal data and provenance) and legitimacy (monitoring bodies and policy). Conversely, Lyon (2016) has listed ten terms describing what 'transparency is not' and associates these terms with related concepts of 'clarity' and 'integrity'. Lyon (2016) goes on to define a 'transparency *action*', a 'transparency *agent*' and a 'transparency *tool*'.

Taking a practical perspective, various mechanisms have been proposed to facilitate enhanced transparency during research workflows. These include authors signing a publication *declaration* of transparency for each research article as part of every journal submission (Altman and Moher, 2013), a policy that is supported by the BMJ and the EQUATOR Network[3] in health research. Similarly, a transparency *appendix* has been proposed by Moravcsik (2014) for the field of qualitative political science, which includes linking an empirical citation to an annotated excerpt from the original source in a process that he calls active citation. Open Data and Open Materials *badges* have been adopted by the journal *Psychological Science,* signalling that the journal values transparency and that authors have met transparency *standards* for their research; the successful application of badges has been described by Kidwell et al. (2016). The Center for Open Science (COS)[4] has published the *Transparency and Openness Promotion (TOP) Guidelines* for journals which cover eight components (Nosek et al., 2015), and has developed the Open Science Framework as a software platform to support more transparent research practices. Goecks, Nekrutenko and Taylor (2010) note that "transparency has received less attention than accessibility and reproducibility, but it may be the most difficult to address". They propose the Galaxy platform for the life sciences as *a substrate* for addressing transparency. Other tools to improve transparency in neuroimaging research have been listed by Gorgolewski and Poldrack (2016) and include domain-specific platforms such as NeuroVault.org[5], a repository for un-thresholded statistical maps and atlases of the human brain. The importance of baking transparency into research design and research protocols has been emphasised by Wilbanks and Friend (2016), who describe a new informed consent procedure framed as *a contract* of data sharing "so that anyone can know how data are being used and by whom". The link between provenance and transparency has been articulated by Downs et al. (2015) and a Provenance and Context Content Standard (PCCS) *matrix* proposed by the Federation of Earth Science Information Partners (ESIP), which has been adopted by NASA. These authors claim that "data citation alone does not solve the

3 EQUATOR Network: http://www.equator-network.org/
4 Center for Open Science: https://cos.io/
5 NeuroVault.org: http://neurovault.org/

transparency issue; full documentation of dataset provenance and context is necessary." Further detailed recommendations for data models and workflows in bioinformatics are made by Gonzalez-Beltran et al. (2015), who advocate the use of Research Object, ISA and nano-publication *models* as mechanisms for assuring reproducibility and transparency in science.

There are challenges associated with assuring research transparency. Whilst journal publishers promote mechanisms to advocate transparency in submissions through declarations, policy statements, badges and *mandates* for data sharing, there is the issue of researcher compliance. Van Noorden (2014) describes a mixed landscape of compliance with the PLOS data sharing mandate and notes that the PLOS ONE editorial director believes that "a complete culture shift will be further down the line". This links to the need for education and training in good transparency practices. The Berkeley Initiative for Transparency in the Social Sciences (BITSS)[6] runs a Summer Institute and awards prizes for open science to academics and researchers. Lyon (2016) proposes that a librarian can act as a transparency *advocate*, by advising on transparent (open) scholarship, reproducible methods and validation approaches. The risks of data sharing and open science for early career scientists are described by Gewin (2016); the desire to be open without becoming scientifically vulnerable is noted, with "scary stories" of scooping emphasising the dilemma. Preparing data for sharing and re-use also has a time investment for researchers, and may lead to senior colleagues questioning researcher productivity. The costs of reproducibility (and transparency) are highlighted by Gonzalez-Beltran et al. (2015), and the Netherlands Organisation for Scientific Research (NWO)[7] is funding a significant Replication Studies pilot, aiming "to make a contribution to increasing the transparency of research," but recognising that such reproducibility efforts carry substantive costs. One possible cost-effective solution is to implement a Data Quality Review and Reproduction of Results Service, which is the approach adopted by the Cornell Institute for Social and Economic Research (Arguillas and Block, 2016). We also note that research transparency can be used to present contrasting political and ideological positions (Sarewitz, 2015) and may be viewed as a 'red flag area' which "can help to differentiate healthy debate, problematic research practices and campaigns that masquerade as scientific inquiry" (Lewandowsky and Bishop, 2016).

Methodology

In this context and to gain a better understanding of researcher perspectives on the concept of transparency, we explore the following research questions:

1. How do researchers conceptualize research transparency?
2. What are the drivers and motivations for transparency during the lifecycle?
3. What tools or services are desirable to support transparency in the lifecycle?
4. What is the perceived role of libraries and research data services?

6 Berkeley Initiative for Transparency in the Social Sciences (BITSS): http://www.bitss.org/

7 Netherlands Organisation for Scientific Research (NWO): http://www.nwo.nl/en/news-and-events/news/2016/nwo-makes-3-million-available-for-replication-studies-pilot.html

To address the research questions, we obtained IRB approval at the University of Pittsburgh, USA (PRO15040061) to conduct four disciplinary focus group sessions between October 2015 and October 2016. Inspired by related work taking advantaging of visual presentation and the use of sticky notes to facilitate discussion (e.g., Bowler, Mattern, and Knobel, 2014; Mattern et al., 2015), we conducted four focus group sessions with faculty.

In qualitative research, a focus group approach is used to stimulate discussion and encourage reluctant participants to contribute their ideas (Peterson and Barron, 2007). The data collection protocol (Table 1) was directly modified from a pilot study, reported in Lyon et al. (2016). In Phase I, participants, all academic researchers, were asked to write down simple words or phrases to conceptualise the term 'research transparency', writing these concepts on a sticky note. The participants were then asked to merge or cluster similar concepts, finding connections and themes among the concepts that they and their colleagues noted.

Table 1. Protocol phases in focus group sessions.

Stages	Description
Setting the stages	Review information and consent; Participants were asked to briefly introduce their research interests.
Phase I: Concept construction	Facilitators distributed sticky notes (a person can use multiple notes); Participants were asked to write down the meaning of the term "Research transparency" in their own words, followed by discussion. They then merge or cluster any similar concepts, followed by discussion.
Phase II: Researchers' current practices of research transparency	Facilitators drew a research lifecycle on a whiteboard; Facilitators asked participants to write down actions or tools related to their day-to-day practices regarding research transparency on sticky notes and to place the notes on the research lifecycle; Facilitators asked participants: "Why are you doing these actions? What are the drivers and motivations?"
Phase III: Researchers and services	Facilitators interviewed participants using questions: • "Can you think of any desired tools or services which would facilitate your actions toward research transparency?" • "Any suggestions for library services or research data services (RDS)?"
Debriefing	Research participants provided suggestions for the focus group protocol.

In Phase II, participants were presented with a research lifecycle (Figure 1) and asked to describe their actions associated with its stages that are related to research transparency. Sticky notes were again used, with participants placing them alongside the relevant research stage(s), thereby situating the actions within the larger research

workflow. A subsequent discussion about the drivers and motivations behind these actions followed. In Phase III, participants were asked about the tools they use which can facilitate research transparency and their suggestions for relevant library research data services. Our participants included 15 senior professors (associate professors or full rank professors) in four different broad disciplines: chemistry, law, social and urban studies, and civil and environmental engineering.

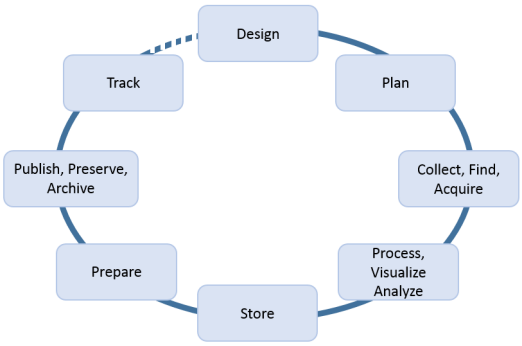

Figure 1. A Research Lifecycle Model, prepared by the University Library System Research Data Management Working Group in 2015.

Table 2 summarizes the number of participants, research disciplines, and the total number of sticky notes collected in each focus group phase (n=72 in Phase I; n=141 in Phase II) and in totality (N=213). The chemistry group was held at the University of Southampton, UK; the three other groups were conducted at the University of Pittsburgh, USA. Each focus group lasted between 50-65 minutes.

Table 2. Participants in focus group sessions.

Discipline	Number of participants	Number of notes in Phase I	Number of notes in Phase II	Total number of notes
Chemistry	3 (P01-P03)	18	46	64
Law	4 (P04-P07)	23	40	63
Social and Urban Research	4 (P08-P11)	19	34	53
Civil and Environmental Engineering	4 (P12-P15)	12	21	33
TOTAL	15	72	141	

The frequency of terms on the sticky notes was recorded in a spreadsheet file; neutral words ‘research’ or ‘study’ were not considered. The visual clustering of concepts during Phase I of the focus group was recorded as an image using an iPhone. Summary headings for each cluster were extracted from the terms written on the sticky notes, either directly by participants or indirectly by facilitators. The focus group discussion was recorded using an oral recording device and then transcribed.

Results

In Phase I, we received 72 sticky notes, ranging from 12 to 23 per group. The Law group contributed the most notes, while civil and environmental engineering the least. To gain sharper insight into the synergies and differences between how researchers construct their definitions of research transparency in their own words, we visualized terms from the participants' sticky notes into word clouds (Figure 2). The most mentioned word is 'data', which appeared 15 times in notes, then 'methods' and 'full' (both six times). The latter term was followed by 'disclose/disclosure', 'description', and 'accessibility'. Other highly mentioned terms were transparency (n=5), open (n=4), and disclosure (n=4). Based on the frequency analysis, it is apparent that researchers were connecting research transparency with data availability and data accessibility. We observed that the Social and Urban Studies group (hereafter: Urban) mentioned the term 'method' five times - a higher frequency compared with other disciplinary groups. The Law group exhibited an evenly distributed list of words, in which no term was mentioned more than three times. The term 'metadata' only appeared in the Civil and Environmental Engineering group (hereafter: Engineering). However, when we followed up with the two researchers who mentioned metadata, one of them explained the term in the following way: "If I publish something, then I should have a metadata, original data that I can give it over to whoever and they should come up with a similar conclusion." On the basis of this description, we believe that 'metadata' was more closely aligned to 'raw data' or 'full disclosure of data' for the Engineering group.

Figure 2. Disciplinary semantic trends associated with research transparency.

In Phase I of the protocol, participants were also asked to reflect on the collection of transparency concepts that they individually captured on sticky notes and to cluster similar concepts around themes. In doing so, the participants identified patterns in their understanding of research transparency. Table 3 presents the themes that participants saw emerging from the conceptualization that they shared as a focus group. We further clustered the themes that appeared across the four focus groups, by merging closely related or synonymous ideas e.g. joining 'ethics' (Engineering) with 'research integrity' (Law).

Table 3. Disciplinary concept themes associated with research transparency.

Transparency concept theme	Discipline(s) identifying theme
Data availability/sharing	All
Transparent methods/research process	All
Open access	Chemistry
Research integrity/ethics	Law, Engineering
Replicability	Law, Urban
Citation and attribution	Law, Urban
Disclosure (e.g. of conflict of interests, funding sources)	Law, Engineering
Metadata	Engineering
Authorship	Engineering

There were two predominant themes that cut across disciplinary understandings of research transparency. In each of the focus groups, participants associated the notion of 'research transparency' with the availability and sharing of data and with richly documented and reported research methods. In relatively disparate disciplines like law and engineering, there was a confluence of other core themes: both the legal scholars and the engineers identified research integrity and disclosure as research transparency themes.

In Phase II, researchers created a total of 141 sticky notes associated with research transparency through the lifecycle shown in Figure 3. The total numbers of sticky notes created by researchers varied across disciplines: Chemistry (n=46), Law (n=40), Urban (n=34) and Engineering (n=21), with disciplinary concentrations (defined as ≥ 10 notes) at particular lifecycle stages: Collect (Law), Process (Urban), Publish (Chemistry and Law). The lifecycle stages with the most transparency notes across all disciplines were: Publish (n=47), Collect (n=30), Process (n=15), Prepare (n=15), Design (n=13) and Store (n=12). The action verbs written by the researchers on the sticky notes were identified in a spreadsheet. From a total of 158 action verbs, *share, use, track, collaborate, collect, record, reference, write, attribute, check, cite, deposit, document, present, read, save, store and submit* were each used ≥ 3 times, however there was a very long tail vocabulary of other action verbs used only once or twice in notes.

 doi:10.2218/ijdc.v12i1.530

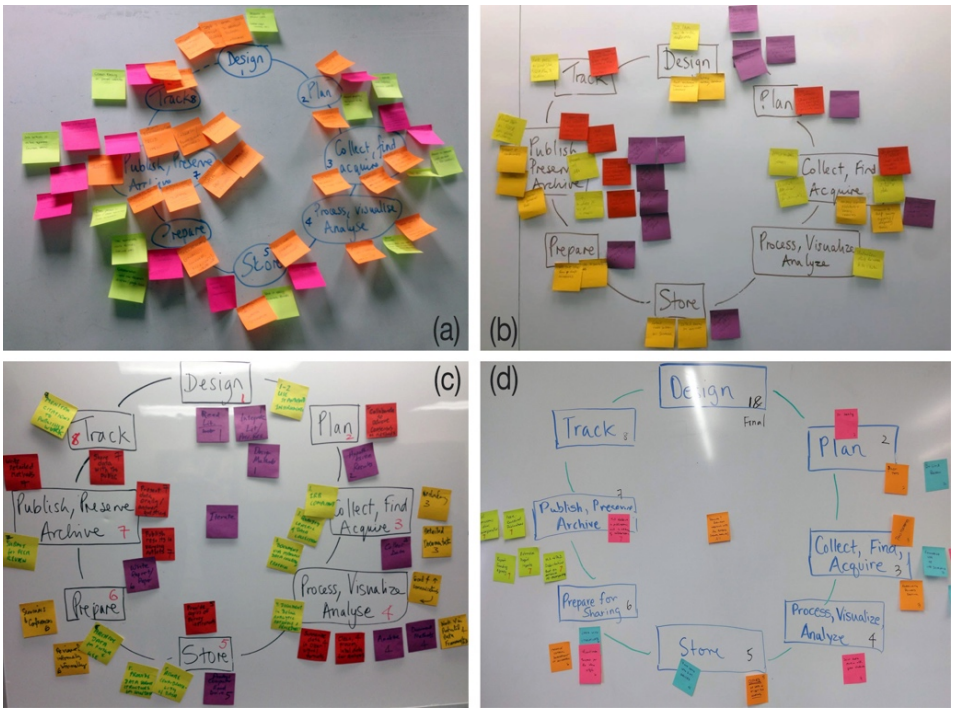

Figure 3. Disciplinary distribution of research transparency actions through the lifecycle in Chemistry (a), Law (b), Urban (c), and Engineering (d).

Table 4 summarizes the distribution of distinct action verbs across the lifecycle stages and illustrates that a range of action verbs are associated with each stage; certain stages (collect, process and publish, for example) having the most varied vocabulary. As the table indicates, there were terms that the participants associate with all stages of the research lifecycle, as well as terms that the participants associated with two stages (e.g. the Plan and Collection stages).

Table 4. Distribution of distinct action verbs associated with research transparency by lifecycle stage.

Lifecycle stage	Action verbs (distinct)
All	iterate, meet
Design	collaborate, develop, attribute, read, search, design, email, integrate, survey, test
Design-Plan	read, use, consider, identify
Plan	collaborate, search, discuss, do, hypothesize, outline, pilot, plan
Plan-Collect	record, reference, detail
Collect	collect, record, use, employ, attribute, share, check, document, disclose, follow, capture, control, encourage, find, keep, obtain, send, tag
Process	document, record, use, share, analy[s]e, clean, label, process, review, summarize, validate, work, work up
Store	save, store, share, deposit, file, assume, protect, provide
Prepare	check, write, use, attribute, collaborate, cite, edit, archive, get, re-examine, revise

Lifecycle stage	Action verbs (distinct)
Publish	share, reference, present, submit, write, cite, deposit, link, retain, use, edit, store, file, disclose, answer, append, associate, discard, distribute, place, post, preserve, publish, put, report, sort, tweet, update
Track	track, follow, monitor

There are many tools and resources that can assist researchers to ensure research transparency. Table 5 lists the tools that our focus group participants perceived to be helpful during the research lifecycle to support transparency; the majority are software and web applications; some of them are ostensibly more general tools to support research and scholarship. Human resources in libraries were also mentioned, with participants from three disciplines (Chemistry, Law, and Engineering) noting the importance of the liaison librarians.

Table 5. Disciplinary tools for research transparency.

Type of tools or resources	Tools to assist research transparency mentioned by participants	Research lifecycle stage	Discipline(s) identifying tool
Software	Standardized statistics software, e.g. SPSS	Process	Urban
	Plagiarism software, e.g. iThenticate	Publish	Engineering
Hybrid platform (desktop sync to web)	Reference managers, e.g. Mendeley, EndNote, Zotero	Publish	Chemistry, Engineering
	Cloud storage and collaboration platform, e.g. Box, Google Drive, Dropbox	Store	Chemistry, Law, Engineering
Web platforms or applications	Scholarly database, e.g. Westlaw and LexisNexis; PubMed	Collect	Law, Urban
	Academic social networking services e.g., ResearchGate	Publish	Chemistry
	Generic social web e.g.,Twitter	Publish	Chemistry
	Real-time collaboration interface e.g.,Google Docs	Plan/Publish	Chemistry
	Google Scholar Citations	Track	Law
	Scholar identifier e.g.,ORCID and ResearcherID	Track	Chemistry; Law; Engineering
	Discipline data repository e.g., ICPSR	Publish	Urban
	Nesstar	Process	Urban
	Regional data centre	Publish	Urban

Type of tools or resources	Tools to assist research transparency mentioned by participants	Research lifecycle stage	Discipline(s) identifying tool
Standards	Standard guidelines for reporting surveys and studies e.g., Manual of Operations	Process	Urban
	Measurement standards e.g., clinical indicators	Collect	Urban
Human resources	Liaison librarians		Law, Urban, Engineering

The research lifecycle mappings show that many tools listed in Phase II of the protocol are associated with the 'Publish' stage. Cloud technologies also play an important role in the researcher's toolbox and account for most of the storage tools. Moreover, reference managers, such as Mendeley and EndNote, are hybrid software tools which can synchronize between multiple devices via clouds. The Urban group reported more tools related to each research lifecycle stage, but this result may be due to discipline differences or individual researchers' preferences for specific tools.

A summary of the disciplinary drivers and motivations for research transparency extracted from the focus group discussion transcriptions is given in Table 6.

Table 6. Disciplinary drivers and motivations for research transparency.

Drivers and motivations	Identifying discipline(s)
Grant applications	Chemistry, Engineering
Documentation, records and re-use	Chemistry, Engineering
Professional status, track metrics, impact	Law
Future cross-disciplinary collaboration	Chemistry, Law
Policy, laws, influence decision-makers	Law, Urban
Disciplinary norms	Law
Publish your work in high-rated journals	Urban
Replicability and standards	Urban
Ethics, public good, honesty	Urban, Engineering
Societal and real-world impact	Urban, Engineering
Trust between student and supervisor	Engineering

Whilst there was some disciplinary variation, there were also common drivers e.g., securing research funding for the chemists and engineers. These same groups also cited documentation to enable the re-use of data as drivers. Participants from Law raised the issue of professional status in the context of enabling collaboration across disciplines; this was also a transparency driver for the chemists:

> 'There's a professional status motivation…the impact of your work is linked to making your work well-documented, accessible and public. We build mechanisms for transparency and access, it makes it easier for future collaborators to become part of the field… its especially true with….cross-disciplinary or inter-disciplinary engagement. If I want people outside of Law…using our methods, then I need to make my work and its processes as transparent as possible, so that they can interpret…and apply them from their own home disciplines' (P04 Law).

The social scientists cited policy, replicability and standards as transparency drivers, the latter was linked to securing publications:

> 'At the backend…to publish your work in highly-rated journals, you have to read all of these transparency standards' (P09, Urban).

The relationship between research transparency and research ethics was raised by the engineers:

> 'Research and transparency and research ethics to some extent may overlap' (P12, Engineering).

They also recognised the need for (full) disclosure:

> 'Research transparency is more about being honest about what the data you actually measure ... the data you are actually recording…there's also this disclosure of funding, who's supporting your work…' (P14, Engineering).

The engineers also offered an unanticipated driver for research transparency, one associated with the concept of trust:

> 'There is also the transparency issue that we're not addressing here, and that is between the student and the faculty member and the supervisor. That... is very critical because you have to be able to trust your student that actually collected the data point' (P14, Engineering).

Part III of the study protocol invited participants to identify existing and desirable University services that would support research transparency (Table 7). Focus group participants diverged from thinking about services in support of research transparency and moved to a discussion of general research and computing services. For example, Urban researchers discussed the value of having a more robust GIS service in the library, which has an unclear link to research transparency. Similarly, a legal faculty member discussed a library subscription to a journal currently inaccessible through the University. Such services and tools seen outside the scope of transparency, are omitted from Table 7. In this study, no strong patterns of valued services and tools emerged, either across disciplines or within disciplines, with participants tending to identify unique services or tools that would be useful to them. However, there was recognition by some participants that libraries, with their long-standing tradition of organization, documentation, and access, have a role to play in supporting research transparency and preserving, in the words of one chemist, "the research crown jewels".

Table 7. Services to support research transparency.

Services to support transparency	Identifying discipline(s)
Repositories for preserving and linking research output of the University	Chemistry
Organizing sources/Reference management	Law and Chemistry
Metadata and documentation	Urban
Plagiarism detection	Engineering

 doi:10.2218/ijdc.v12i1.530

Discussion

Whilst acknowledging that this was a small-scale study of researcher views on research transparency solely intended to begin to scope the field, the findings have highlighted some interesting perspectives. The wide range of terms and concepts collected is an indication of the complexity of the research transparency arena. Thematic complexity has been identified within the associated areas of trust (see Yoon's presentation in Curty, Yoon, Jeng, and Qin, 2016) and reproducibility (see Baker, 2016, who reports on surveys of researchers and finds "no consensus on what reproducibility is or should be"). Research transparency appears to have a similarly broad interpretation by researchers, albeit with some 'core' vocabulary and concepts that transcend disciplinary boundaries: data sharing/availability; richly documented methods/research process; full disclosure of funding sources and conflicts of interest. Even in relatively disparate disciplines (law and engineering), research integrity and ethics were raised as a key theme. These findings suggest that any future investigations of research transparency should include a special lens on these particular themes. There were notable intra-disciplinary concepts, including a strong focus on 'methods' from the Social and Urban participants. This group also listed replicability, perhaps reflecting awareness of recent reproducibility studies in this domain (Open Science Collaboration, 2015). Knowing the similarities and differences in meanings between disciplines, allows us to develop more effective data policies with nuanced language and targeted data curation practice guidelines to inform advocacy and training for researchers in each discipline.

We also observe that certain lifecycle stages (e.g., collect, process, publish) have comparatively more associated action verbs for research transparency than others. One possible interpretation is that research activities associated with these stages are implicitly link to the development of the researcher's professional profile, to tenure opportunities, or to career rewards. However, a further investigation is needed to determine the finding. For example, are these action verbs easy to execute? Are they supported by available software tools and institutional infrastructure? Conversely are these required actions from the researcher's perspective e.g. required for compliance with funder policy and therefore not optional? Action verbs have been used by the Australian National Data Service (ANDS) to identify the key functions which support the re-use of data (Burton and Treloar, 2009). Can a suite of action verbs have a similar role in promoting research transparency? Some stages (design, plan, store, prepare, track) have comparatively fewer actions associated with them. Are these real gaps in research transparency practice or gaps in researcher perception or understanding of research transparency good practice?

With regard to tools to facilitate transparency, participants pointed to software, Web applications and human resources. Different tools were identified for different stages of the lifecycle, with most tools associated with 'Publish', suggesting that researchers focus more on the transparency of 'research outputs' notionally at the end of the process, rather than 'research inputs' at the start of the process. For example, bibliometric tools (e.g. Mendeley), which offer document annotation and citation metadata management, were mentioned several times by multiple participants. These tools can help researchers maintain a full citation record during their literature review, thereby improving research transparency. Cloud-based platform tools, such as Google

Docs, were also mentioned frequently at the 'Store' stage; participants used them in regular research activities (e.g. co-authoring, sharing files with their teammates, or storing data). However, it is unclear whether such cloud-based tools have a strong relationship to research transparency and further content analysis is needed to reveal the context. In contrast, there were no tools associated by researchers with the 'Plan' stage; arguably tools such as DMPOnline and DMPTool can enhance transparency by documenting the early pre-award planning stages of research. There was also no mention of tools such as the Open Science Framework[8], which aims to provide transparency, open methods and to record all stages of the lifecycle. The implications for continuing researcher education and advocacy are clear. Some tools had an obvious disciplinary relevance, such as Nesstar developed for the social sciences (Urban group), whilst other tools were generic in nature and were cited by several disciplines e.g. ORCID identifiers. Once again, this suggests a need to target advocacy and training for transparency to particular disciplines.

Interestingly, participants in three disciplines identified liaison librarians as resources to facilitate transparency. Lyon (2016) has suggested that new data science roles, like data librarian, can act as 'transparency agents' to enable and catalyse research transparency. However, in this study it is not completely clear whether the focus group participants were suggesting that librarians act as channels or conduits for transparency, or as specific service providers. Once again, this point requires further investigation and has links to the services findings described below. The diversity of drivers and motivations cited by participants highlights the multi-faceted nature of research transparency. However, they can be divided into two distinct categories: a) political and community drivers, such as policy, laws, influence decision-makers, disciplinary norms, replicability and standards, ethics, public good, honesty, societal and real-world impact; and b) personal and professional motivations, such as grant applications, documentation, records and re-use, professional status, track metrics, impact, future cross-disciplinary collaboration, publishing your work in high-rated journals, and trust between student and supervisor. Whilst there is clearly some overlap (e.g. ethics and honesty), this division may inform the development of more effective advocacy messages to catalyse cultural change and to influence researcher practices and behaviours. The importance of 'trust' as an associated concept was also identified in this study and more work is required to unpack the relationships between research transparency and trust.

There are noticeable disciplinary differences in the nature of desirable and valued services to support research transparency. As an example, for the engineers, a primary area of desirable support focused on plagiarism detection and education. In emphasizing this need, the researchers appeared to equate transparency with academic honesty. They indicated that this need stems from collaborative writing with students and is rooted in a concern for both their and the University's academic reputation. Aspects of the research and scholarly communications culture in the disciplines studied here may have a bearing on identified needs. Co-authored publications are considerably more prevalent in engineering than in arts and humanities disciplines (see Sparks, 2005). It is unsurprising, then, that availability and support around a plagiarism detection tool were of interest to the engineers and not a recommendation from the focus group with legal scholars. There is more work required to tease out the types of services to facilitate research transparency and to identify the optimal providers of these services. However, given the services identified in this study, it would seem logical for libraries and information/data professionals to play a leading role.

8 Open Science Framework: https://osf.io/

This study has enabled us to refine the protocol for exploring research transparency. The research lifecycle proved to be a unifying foundation for these discussions and was helpful in mapping the use of specific transparency tools to different stages. However, the apparent disconnect between some cited services and research transparency suggests an adjustment to the protocol in Phase III. In future, we can ensure that the participants' thinking remains with transparency by asking them to place the service, tool, or resource alongside relevant action verbs previously captured on sticky notes; in so doing, we aim to gain a sharper picture on service, tool, and resource requirements, within and across disciplines.

Conclusion and Future Work

In conclusion, our preliminary study of research transparency has proved valuable in illustrating the multi-faceted nature of the area, identifying core concepts to investigate in more depth, and providing insights into the vocabulary and semantics used across different disciplines. We view this study as the first step towards building a 'lexicon' or 'taxonomy' for work in this critical field. The lifecycle motif has proved an effective foundation on which to explore transparency perspectives. Finally, we aim to carry out a scaled-up investigation into research transparency as the next stage in this research.

References

Academy of Medical Sciences. (2015). Reproducibility and reliability of biomedical research: improving research practice. Symposium Report. Retrieved from http://www.acmedsci.ac.uk/policy/policy-projects/reproducibility-and-reliability-of-biomedical-research/

Altman, D.G. & Moher, D. (2013). Declaration of transparency for each research article. *BMJ,* 347:f4796. Retrieved from http://www.bmj.com/content/347/bmj.f4796

American Political Science Association. (2012). A guide to professional ethics in political science (2nd ed.) Retrieved from http://www.apsanet.org/portals/54/Files/Publications/APSAEthicsGuide2012.pdf

Arguillas, F.O. & Block, W.C. (2016). R2: CISER's data quality review and reproduction of results service. In Proceedings of 11th International Digital Curation Conference (IDCC). Retrieved from http://www.dcc.ac.uk/sites/default/files/documents/IDCC16/54_Arguillas%20and%20Block%20-%20Poster%20IDCC%202016.pdf

Baker, M. (2016). 1,500 scientists lift the lid on reproducibility. *Nature, 533*(7604), 452-454. Retrieved from http://www.nature.com/news/1-500-scientists-lift-the-lid-on-reproducibility-1.19970

Bowler, L., Mattern E., & Knobel C. (2014). Developing design interventions for cyberbullying: A narrative-based participatory approach. In Proceedings of iConference 2014.

Burton, A. & Treloar, A. (2009). Designing for discovery and re-use: The ANDS data sharing verbs' approach to service decomposition. *International Journal of Digital Curation, 4*(3), 44-56. Retrieved from http://www.ijdc.net/index.php/ijdc/article/view/133

Curty, R., Yoon, A., Jeng, W., & Qin, J. (2016). Untangling data sharing and reuse in social sciences. In Proceedings of the 79th ASIS&T Annual Meeting.

Downs, R.R., Duerr, R., Hills, D.J., & Ramapriyan, H.K. (2015) Data stewardship in the earth sciences. *D-Lib Magazine, 21*(7/8). Retrieved from http://www.dlib.org/dlib/july15/downs/07downs.html

Etzioni, A. (2010). Is transparency the best disinfectant? *Journal of Political Philosophy 18(*4), 1-16. Retrieved from http://www.gwu.edu/~ccps/etzioni/documents/295Transparency.pdf

FASEB. (2015). Enhancing research reproducibility: Recommendations from the Federation of American Societies for Experimental Biology. Retrieved from http://www.faseb.org/Portals/2/PDFs/opa/2016/FASEB_Enhancing%20Research%20Reproducibility.pdf

Gewin, V. (2016). Data sharing: An open mind on open data. *Nature, 529*, 117-119. Retrieved from http://www.nature.com/naturejobs/science/articles/10.1038/nj7584-117a

Goecks, J., Nekrutenko, A., & Taylor, J. (2010). Galaxy: A comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. *Genome Biology, 11*:R86. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/20738864

Gonzalez-Beltran, A., Li, P., Zhao, J., Avila-Garcia, M.S., Roos, M., Thompson, M, van der Horst, E., Kaliyaperumal, R., Luo, R., Lee, T-L, Lam, T-W., Edmonds, S.C., Sansone, S-A., & Rocca-Serra, P. (2015). From peer-reviewed to peer-reproduced in scholarly publishing: The complementary roles of data models and workflows in bioinformatics. *PLOS ONE.* Retrieved from http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0127612

Gorgolewski, K.J. & Poldrack, R.A. (2016). A practical guide for improving transparency and reproducibility in neuroimaging research. *PLOS Biology.* Retrieved from http://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.1002506

Gov.UK. (2013). G8 open data charter and technical annex. Retrieved from https://www.gov.uk/government/publications/open-data-charter/g8-open-data-charter-and-technical-annex

Holdren, J.P., Orszag, P., & Prouty, P.F. (2009). President's memorandum on transparency and open government – Interagency collaboration. Retrieved from https://obamawhitehouse.archives.gov/the-press-office/transparency-and-open-government

ICSU. (2015). Open data in a big data world. An international accord. Retrieved from https://www.icsu.org/cms/2017/04/open-data-in-big-data-world_long.pdf

Kenett, R.S. & Shmueli, G. (2015). Clarifying the terminology that describes scientific reproducibility. *Nature Methods, 12*(8), 699. Retrieved from http://www.nature.com/nmeth/journal/v12/n8/full/nmeth.3489.html

Kidwell. M.C., Lazarevic, L.B., Baranski, E., Hardwicke, T.E., Piechowski, S., Falkenberg, L-S, Kennett, C., Slowik, A., Sonnleitner, C., Hess-Holden, C., Errington, T.M., Fiedler, S., & Nosek, B.A. (2016). Badges to acknowledge open practices: A simple low-cost, effective method for increasing transparency. *PLOS Biology.* Retrieved from http://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.1002456

Lewandowsky, S. & Bishop, D. (2016). Research integrity: Don't let transparency damage science. *Nature, 529*(7587), 459-461. Retrieved from http://www.nature.com/news/research-integrity-don-t-let-transparency-damage-science-1.19219

Lyon, L. (2016). Transparency: The emerging third dimension of open science and open data. *Liber Quarterly, 25*(4), 153-171. Retrieved from https://www.liberquarterly.eu/article/10.18352/lq.10113/

Lyon, L., Mattern, E., Jeng, W., & He, D. (2016). Investigating perceptions and support for transparency and openness in research: Using card sorting in a pilot study with academic librarians. In Proceedings of the 79th ASIS&T Annual Meeting.

Mattern, E., Jeng, W., He, D., Lyon, L., & Brenner, A. (2015). Using participatory design and visual narrative inquiry to investigate researchers' data challenges and recommendations for library research data services. *Program 49*(4), 408-423. doi:10.1108/PROG-01-2015-0012

Miguel, E., Camerer, C., Casey, K., Cohen, J., Esterling, K.M., Gerber, A., Glennerster, R., Green, D.P., Humphreys, M., Imbens, G., Laitin, D., Madon, T., Nelson, L., Nosek, B.A., Petersen, M., Sedlmayr, R., Simmons, J.P., Simonsohn, U., & Van der Laan, M. (2014). Promoting transparency in social science research. *Science, 343*(6166), 30-31. Retrieved from http://science.sciencemag.org/content/343/6166/30

Moravcsik, A. (2014). Transparency: The revolution in qualitative research. *PS Political Science & Politics, 47*(1), 48-53. Retrieved from https://www.princeton.edu/~amoravcs/library/transparency.pdf

NIH. (2015). Enhancing reproducibility through rigor and transparency. Notice Number NOT-OD-15-103. Retrieved from http://grants.nih.gov/grants/guide/notice-files/NOT-OD-15-103.html

Nosek, B.A. et al. (2015). Promoting an open research culture. *Science, 348*(6242), 1422-1425. Retrieved from http://science.sciencemag.org/content/348/6242/1422

OECD. (2007). OECD principles and guidelines for access to research data from public funding. Retrieved from http://www.oecd.org/sti/sci-tech/38500813.pdf

OECD. (2015). Making open science a reality. Retrieved from https://www.innovationpolicyplatform.org/content/open-science

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science, 349*(6251), Retrieved from http://science.sciencemag.org/content/349/6251/aac4716

Peterson, E.R., & Barron, K.A. (2007). How to get focus groups talking: New ideas that will stick. *International Journal of Qualitative Methods, 6*(3), 140-144.

Royal Society. (2012). Science as an open enterprise: Open data for open science. Retrieved from https://royalsociety.org/~/media/Royal_Society_Content/policy/projects/sape/2012-06-20-SAOE.pdf

RCUK. (2015). RCUK common principles on data policy. Retrieved from http://www.rcuk.ac.uk/research/DataPolicy/

Sarewitz, D. (2015). Reproducibility will not cure what ails science. *Nature, 525*(7568). Retrieved from http://www.nature.com/news/reproducibility-will-not-cure-what-ails-science-1.18339

Sparks, S. (2005). JISC disciplinary differences report. JISC Scholarly Communications Group. Retrieved from https://www.jisc.ac.uk/news/new-reports-give-further-insights-into-open-access-publishing-07-sep-2005

Stodden, V., Bailey, D.H., Borwein, R.J., LeVeque, W.R., Rider, W., & Stein, W. (2013). Setting the default to reproducible: Reproducibility in computational and experimental mathematics. ICERM Workshop. Retrieved from http://stodden.net/icerm_report.pdf

Van Noorden, R. (2014). Confusion over publisher's pioneering open-data rules. *Nature, 515*(7528). Retrieved from http://www.nature.com/news/confusion-over-publisher-s-pioneering-open-data-rules-1.16409

Vayena, E., Salathe, M., Madoff, L.C. & Brownstein, J.S. (2015). Ethical challenges of Big Data in Public Health. *PLOS Computational Biology.* doi:10.1371/journal.pcbi.1003904

Wilbanks, J. & Friend, S.H. (2016). First design for data sharing. *Nature Biotechnology, 34*(4), 377-379. doi:10.1038/nbt.3516