



## Transparency: The Emerging Third Dimension of Open Science and Open Data

Liz Lyon

Visiting Professor, School of Information Sciences, University of Pittsburgh,  
USA  
elyon@pitt.edu

### Abstract

This paper presents an exploration of the concept of research transparency. The policy context is described and situated within the broader arena of open science. This is followed by commentary on transparency within the research process, which includes a brief overview of the related concept of reproducibility and the associated elements of research integrity, fraud and retractions. A two-dimensional model or continuum of open science is considered and the paper builds on this foundation by presenting a three-dimensional model, which includes the additional axis of 'transparency'. The concept is further unpacked and preliminary definitions of key terms are introduced: transparency, transparency action, transparency agent and transparency tool. An important linkage is made to the research lifecycle as a setting for potential transparency interventions by libraries. Four areas are highlighted as foci for enhanced engagement with transparency goals: Leadership and Policy, Advocacy and Training, Research Infrastructures and Workforce Development.

**Key Words:** research transparency; open science; research data services; LIS education

## **1. Introduction**

This paper presents an exploration of the concept of research transparency as a precursor to a new program of qualitative research at the School of Information Sciences (iSchool) University of Pittsburgh, which is investigating the perceptions of research transparency amongst librarians in universities and research institutions. The motivations for exploring this concept (and the associated concept of reproducibility), are multi-faceted. Firstly at a national level, transparency has been highlighted by government leaders as a characteristic of open government (Holdren, Orszag, & Prouty, 2009). Secondly, federal funding agencies with significant research portfolios, such as the US National Institutes of Health (NIH), have articulated their policy and plans for more rigorous research (Lauer, 2015; NIH, 2015a) supported by revised grant application guidance (NIH, 2015b). Thirdly, leaders of professional societies such as the National Academy of Sciences, have voiced their concerns about the current reporting of science (Cicerone, 2015), and other organizations such as the Federation of American Societies for Experimental Biology (FASEB), have published Recommendations for enhancing research reproducibility, which include transparency parameters (FASEB, 2016). Finally some publishers such as PLOS Biology, are taking a proactive approach to encourage reproducibility efforts (Denker, 2016).

In laying the foundation, we first examine ‘transparency’ within the policy context, noting the broader policy arena of open science, which has been articulated by research funding agencies, national governments and other interested parties. This is followed by commentary on transparency within the research process, which includes a brief overview of the related concept of reproducibility and introduces the elements of research integrity, fraud and retractions. In the later sections, an existing two-dimensional model or continuum, is revisited and the paper builds on this framework by presenting a new three-dimensional model, which includes the additional axis of ‘transparency’. In order to acquire a better understanding of its relevance to libraries, the concept of transparency is further unpacked in terms of definitions and vocabulary and selected key terms are introduced. We review opportunities for potential transparency interventions and situate these within the research lifecycle. The final section considers the practical implications for library and information services. The emergence of a range of new research data services is attracting much professional debate as a key area of development for academic and research libraries. In this context, four areas

are highlighted as foci for enhanced engagement with transparency goals: Leadership and Policy, Advocacy and Training, Research Infrastructures and Workforce Development.

## 2. Two Perspectives on Transparency

### 2.1. Transparency in the Policy Context

Governments and policy-setting bodies in North America, Europe and Australasia, as well as in other global nations, have set a clear agenda for open science and open data, which is supported by research funding agency requirements for Data Management or Data Sharing Plans as a component of submitted research proposals e.g. NSF and NIH, UK Research Councils and the European Commission. Research transparency is an identified concept, principle or value articulated within many of these policy statements. The thirty countries of the OECD identified transparency as one of the principles in the OECD Guidelines for access to research data from public funding (OECD, 2007). These guidelines highlight four factors to consider in ensuring transparency: *“documentation on available datasets and conditions of use should be easy to find”*, *“research agencies should actively disseminate information on research data policies”*, *“members of the various research communities should assist in establishing agreements on standards for cataloguing data”* and *“Information on data management and access conditions should be communicated among data archives and data producing institutions”*. In the United States, the Obama Administration released a memorandum on Transparency and Open Government (Holdren et al., 2009) which set out three specific actions for departments and government agencies: Transparency, Participation and Collaboration. Transparency as a value has been discussed in ideological terms by Etzioni (2010), who describes the strong variant relating to regulatory contexts and disclosure. The Royal Society Report (2012) *“Science as an Open Enterprise”* makes reference to *“transparent policies for custodianship, data quality and access”* in outlining a set of principles of stewardship which should be shared by custodians of scientific work. In the UK, the Research Councils have published a set of Common Principles on Data Policy (originally published in 2011 and revised: RCUK, 2015) which state: *“Making research data available to users is a core part of the Research Councils’ remit and is undertaken in a variety of ways. We are committed to transparency and to a coherent approach across the research base”*. The G8 countries Open Data Charter (Gov.UK,

2013) states that “Open data can increase transparency about what government and business are doing” and Principle 4 states that “We will be transparent about our own data collection, standards, and publishing processes, by documenting all of these related processes online”. Furthermore, transparency was an integral element of the European Commission Horizon 2020 program calls, which highlighted transparency elements within innovation pilots for open government. In 2015, the OECD published a substantive report on open science, which identifies “increasing transparency and quality in the research validation process” as a rationale for open science and open data (OECD, 2015); also in 2015, four major global science organizations (the International Council for Science, the InterAcademy Partnership, the World Academy of Sciences and the International Social Science Council) published an international accord, which includes the assertion that “Openness and transparency have formed the bedrock on which the progress of science in the modern era has been based” (ICSU, 2015).

## 2.2. Transparency in the Research Process

Moving from the policy perspective to look more widely at the literature describing the practice of research, there has been a gradual increase in attention given to transparency concepts and reproducible science protocols, processes and products. At the disciplinary level, there is considerable diversity in open practices with certain domains having well-established norms for data release (e.g. astronomy and genomics), whilst in other disciplines, notably across the humanities and social sciences, open data practices are less common. Peng (2011), for example, speaks to these diversity differences. Addressing reproducibility in computational science, Peng observes disciplinary differences in the ‘culture of replication’ and suggests that replication may be hindered by the size of datasets and the amount of computing power, time, and money necessary to reproduce the study. Gezelter (n.d.) expands on the importance of verifiability in good science and in particular notes that we need “*verifiability in practice as well as verifiability in principle*”. Collins and Tabak (2014) call for enhanced reproducibility in research funded by the US National Institutes of Health, and describe positive steps associated with publishing practice and scholarly communications.

Examples of uncertain outcomes and flawed research due to a lack of transparency are highlighted in two articles in *The Economist* (2010, 2013), a *New York Times* (Carey, 2011) exposé on fraud in psychological research, and

a discussion of reproducibility issues in the biomedical/clinical trials domain (Ince, 2011). The related controversy around peer review failures and the current “retraction epidemic” has been described by Fang, Steen, and Casadevall (2013), where many retractions (43% of the sample) were due to fabrication or falsification. Many of these retractions have led to significant adverse reputational impacts on both institutions and individuals.

In parallel with these high-profile cases of research malpractice, a number of practical approaches towards achieving greater research transparency have emerged. These include institutional data policies that recommend data deposit and data sharing (e.g. University of Bath, 2014), open data repositories such as Dryad, open software tools for sharing workflows such as Taverna, and electronic laboratory notebooks like RSpace. The Reproducibility Initiative and the associated Reproducibility Project in Cancer Biology, led by the Science Exchange<sup>1</sup> and its independent validation service, have taken transparency and trust goals a stage further, by seeking to reproduce the results in fifty high-impact cancer articles published in 2010–2012 (Errington et al., 2014). A similar initiative was taken in psychology, where a collaborative group of researchers have attempted to replicate studies published in three psychological journals in 2008 (Nosek, 2012; Open Science Collaboration, 2015). This group has introduced the publication format of Registered Replication Reports (Nosek & Lakens, 2014), where the research methodology undergoes peer-review before the experimental procedures are executed, data collected and results published.

There is also ongoing work to explore different approaches to open data peer review processes (Mayernik, Callaghan, Leigh, Tedds, & Worley, 2014), and efforts to incentivize peer review exemplified by the GigaScience partnership with Publons.<sup>2</sup> Scholarly publishers and professional societies have an influential role in the research landscape, with organizations articulating their open data policies and expectations from researchers (American Meteorological Society, 2013). However the perceived lack of transparency in the scholarly publishing process has led to a proposal for a new Transparency Index which could contain details of editorial boards, data requirements and procedures for dealing with retractions (Marcus & Oransky, 2012). Furthermore, whilst the different stakeholders in the scholarly communications arena are exploring ways to increase the rigor of the research process, libraries are considering the conceptual framing and scope of open science, and delivering innovative research data services.

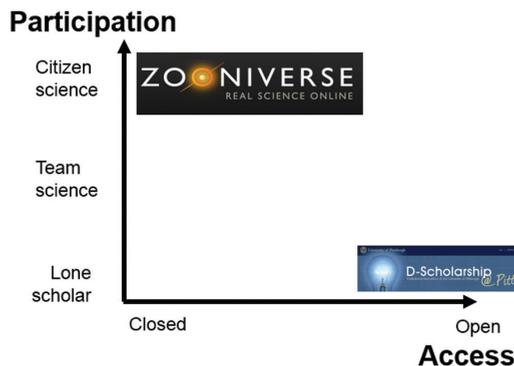
### 3. Open Science Models

#### 3.1. Open Science in Two Dimensions

Open research has been characterized by a commitment and adherence to accessibility, sharing, transparency, and inclusivity (Borgman, 2015). Previously, a Continuum of Openness was described by Lyon (2009) which had two orthogonal axes: 'Access' and 'Participation' or inclusivity. The 'Access' dimension addressed the ability to (freely) locate and retrieve articles, ranging from 'closed' sources, such as a peer-reviewed journal positioned behind a subscription paywall, to an open access institutional repository such as D-Scholarship (university name removed for review). The 'Participation' dimension encompassed the degree of collaboration in research ranging from a single lone scholar, through collaborative professional research teams (team science) to citizen science, where members of the public are partners in the design, implementation and publication of research, such as the Zooniverse projects. Thus a 'continuum' of openness can be described, with organisations, projects and infrastructure platforms positioned on the axes, according to their degree of openness. This two-dimensional model was framed as a Continuum of Openness and is shown in Figure 1.

The first of these two dimensions ('Access') was also explored in more detail by Corral and Pinfield (2014), who constructed a typology of "open" (open content, open development, open infrastructure) and examined convergence

Fig. 1: 2D-model of Open Science (Based on the Continuum of Openness in Lyon, 2009).



and coherence amongst initiatives in higher education and research. The second dimension ('participation') was examined in the context of libraries by Lyon and Beaton (2015), who reviewed citizen science initiatives, education and skills development and opportunities for academic libraries, public libraries and Library Schools/iSchools, recognizing a need for greater awareness and education of librarians to the opportunities in this field.

### 3.2. Introducing a 3D-Model of Open Science

The concept of transparency and the associated term 'reproducibility', have become increasingly important in the current interdisciplinary research environment which is exemplified by greater data volumes, burgeoning numbers of research publications and a critical requirement for accountability for the expenditure of and impact derived from research supported by public funds, particularly in these times of economic constraint. The Transparency Principle has been presented as "*Information on research data and data-producing organization, documentation on the data and conditions attached to the use of the data should be internationally available in a transparent way, ideally through the Internet*" (OECD, 2007). Further interpretations are "*full transparency in reporting experimental details so that others may reproduce and extend the findings*" (NIH, 2016) and "*the reporting of experimental materials and methods in a manner that provides enough information for others to independently assess and/or reproduce experimental findings*" (FASEB, 2016).

Whilst definitions of reproducibility and repeatability were published over twenty years ago by Taylor and Kuyatt (1994) in a NIST Technical Note, reproducibility concepts have been examined again from both computational and legal perspectives by Stodden (2009). Useful definitions of open science and reproducible research were proposed: *Open or reproducible research is auditable research made openly available. Furthermore: Auditable research is where sufficient records (including data and software) have been archived so that the research can be defended later if necessary or differences between independent confirmations resolved. The archive might be private, as with traditional laboratory notebooks.* (Stodden et al., 2013).

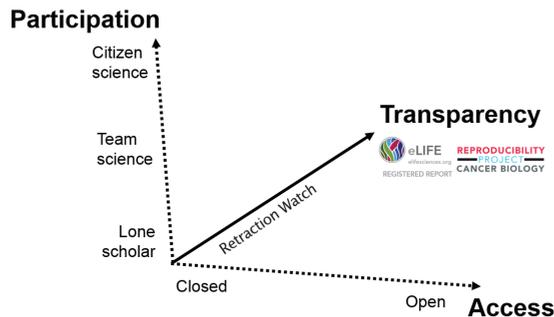
The use of the terms 'archiving' and 'records' in this definition, serve to emphasise the potentially significant role of libraries and information services in supporting open research in the long term. Three distinct categories

of reproducibility: computational (access to code, data and other implementation details), empirical (non-computational empirical scientific experiments) and statistical (analyses and assessments) were described by Stodden, Leisch, and Peng (2014). Easterbrook (2014) used a Venn Diagram to illustrate the relationship between repeatability and reproducibility based on open source code, which contributes towards the goal of verifying the computational research which has been undertaken and subsequently published. In this environment of accountability, the original 2D-model (or continuum) of open science has been extended to include a third dimension: “Transparency”, which is shown in Figure 2. This third dimension projects the efforts of the Retraction Watch (a blog which tracks retractions), towards the Reproducibility Initiatives in Cancer Biology (an initiative to replicate the results of high-profile cancer studies) and the eLife journal, which has published the outcomes of these reproducibility studies.

#### 4. Further Unpacking the Transparency Concept

What exactly is meant by transparency in the context of research? As a start it may be helpful to consider *what transparency is not*; various terms and phrases capture different aspects of this assertion and some are listed in Table 1. Note that certain transparency terms also align towards Participation/Inclusivity or towards Access, indicating the high degree of inter-dependency and connectedness between the three axes in the model. However the two sub-categories of ‘Clarity’ and ‘Integrity’ are unique to the

Fig. 2: 3D-Model of Open Science.



*Table 1: What Transparency is Not in Twenty Terms.*

<b>Clarity</b>	<b>Integrity</b>
Confusing	Not verified
Gray/grey	Not validated
Vague	Not auditable
Unclear	Not supported
Opaque	Not described
Ambiguous	Not documented
Obscured	Not recorded
Implicit	Not versioned
Hidden	Not tracked
Secret	No provenance

Transparency concept and have particular resonance within research practice, with implications for data, workflows and scholarly publications.

Building on earlier definitions and terminology associated with Reproducibility, and drawing on current computational and organizational vocabulary, some foundational terms and articulations for Transparency concepts in the context of Open Science, are proposed and summarized in Table 2. These may be developed, augmented and extended to build a more comprehensive vocabulary of transparency-related terms.

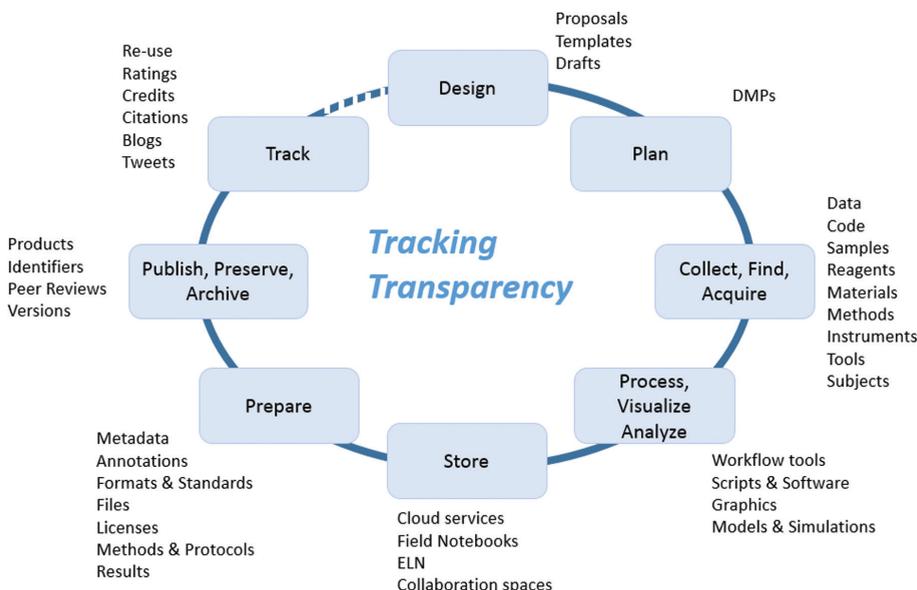
The fundamental concept of Transparency can be considered as an outcome from a combination of different behaviours and practices associated with reproducibility which are implemented by the various actors and stakeholders in the research process. Transparency is generally viewed positively, in particular within the settings of institutional or organizational audits, external scrutiny for research malpractice and demonstrating accountability to funding bodies. Transparent research practices and processes also serve to demonstrate more rigorous methodologies or experimental protocols and to strengthen public perceptions of research quality, integrity and trust in the results, claims, conclusions and assertions derived from research activities. The research lifecycle forms a rich foundation or substrate for grounding thinking about transparency. Figure 3 shows a research lifecycle developed by the University Library System Research Data Management Working

Table 2: Foundational Terms for Transparency in Open Science.

Term	Exposition
Transparency	The outcome from a suite of behaviours which characterize Reproducible Research
Transparency	Facilitates and enhances Research Quality, Research Integrity and Trust
Transparency Action	Describes a specific intervention which is a component of the processes, protocols and practices within the Research Lifecycle
Transparency Agent	Exemplified by the Data Science roles e.g. Data Librarian. These are key components of the Data Fabric (RDA) and supporting Infrastructure; they promote and demonstrate specific behaviours and practices which lead to culture change towards Open Science
Transparency Tool	The software and model frameworks which support Open Science practice

Group at the University of Pittsburgh led by Dr Nora Mattern, with some additional 'Transparency Tracking' points highlighted. (Note that in Figure 3, ELN=Electronic Laboratory Notebook; DMP=Data management Plan).

Fig. 3: Tracking Transparency within the Research Lifecycle.



Understanding the specific tasks, actions and transactions associated with the component stages, workflows, objects and infrastructure within the research lifecycle, will help to illustrate the complexity and proliferation of intervention points where greater transparency can be achieved. Research workflows contain tasks, sub-tasks and actions which are executed either by the researcher or by another physical (or software) ‘agent’ in the process. These proactive interventions can be characterized by Transparency Actions (or verbs) and examples include ‘describe’, ‘identify’ and ‘share’. Transparency Agents can be characterized as defined roles or named individuals or organisations, who execute a specific action or intervention. They are exemplified by six new Data Science roles described by Lyon and Brenner (2015) e.g. Data Librarian, Data Archivist, Data Steward. Transparency Agents also advocate, promote and demonstrate particular behaviours and good practices, which over time will lead to culture change towards a more Open Science environment. The Data Science roles or positions are key human infrastructure components of the Data Fabric articulated by the Research Data Alliance<sup>3</sup> (RDA) and complement the supporting technical infrastructure such as institutional repository platforms and software tools like the Open Science Framework<sup>4</sup>. These types of research lifecycle component can be designated as Transparency Tools, and their use and application within research workflows is desirable.

A presentation of transparency terms as “transparency standards” has been developed by the Center for Open Science as a part of the Open Science Framework and as a modular approach to their TOP Guidelines for Transparency and Openness promotion in Journal Policies and Practices (Center for Open Science, 2015). These transparency standards include 1. Citation, 2. Data transparency, 3. Analytic methods (code) transparency, 4. Research materials transparency, 5. Design and analysis transparency, 6. Preregistration of studies, 7. Preregistration of analysis plans and 8. Replication. Templates are provided for three levels for each standard, to assist with the common expression of research practices for journal publications.

## **5. Implications for Library and Information Services**

Academic library and information services (LIS) are currently tackling the diverse challenges of data curation, research data management and the

provision of more extensive research support services. Recent international studies of research data services have identified a range of issues associated with these developments (Corrall, Kennan, & Afzal, 2013; Si, Xing, Zhuang, Hua, & Zhou, 2015). LIS have a key role in engaging and contributing to different stages in the research lifecycle and there are also critical implications for LIS education. In this paper, four broad areas of opportunity for LIS engagement and action on transparency in open science are highlighted. Questions such as 'How does this trend impact on library and institutional policy?', 'What new library services might be developed?' and 'How should iSchools augment their educational offerings to encompass transparency concepts?' will be explored.

### **5.1. Institutional Research Policy and Library Leadership Opportunities**

Many academic institutions have a Research Policy or Research Code of Practice which states the principles, ethical foundations and expectations of researcher behaviour within that institution. These types of document may cover aspects of Open Science which correspond to the dimensions of the 3D Model. For example, a Research Policy may describe scholarly publication channels and include commentary on Open Access (OA) journals and institutional OA funds. The Policy may have some narrative regarding participation, inclusivity and academic inter-relationships with the public; it may explicitly support citizen science collaborations. Furthermore, the Research Policy may have clauses relating to Research Quality and Research Integrity in broad terms. Building on this point, LIS senior managers can highlight transparency and reproducibility issues and ensure that institutional policy developments in OA and research data management, reflect the third dimension of open science, through requirements for transparent science processes, methodologies and peer review. In this way, LIS can lead on the inclusion of transparency principles as part of institutional policy.

### **5.2. Advocacy and Training for Researchers**

The increasing data volumes generated from high-throughput devices such as sequencers, computational analysis, large-scale simulations and expanding collections of observational and environmental sensor data have led to the emergence of a new field of Data Science. There are a range of new

roles associated with this field which encompass data analysis, stewardship, software engineering, journalism, managing data archives and data librarianship (Lyon & Brenner, 2015). Some academic libraries are extending their existing Research Support Services to include Research Data Management as a key component with advocacy and advisory services (Cox & Pinfield, 2014). Librarians can act as a transparency advocate with faculty by advising on open (transparent) scholarship, reproducible methods and validation approaches. Raising the awareness of new-entrant researchers, providing transparency and reproducibility information, tools and training, are opportunities for libraries to further demonstrate their value and reach.

### 5.3. Research Infrastructures

National and academic libraries are making significant progress towards sustainable digital stewardship. This goal involves supporting all of the stages of the research lifecycle: designing research protocols and project planning, developing data management or sharing plans; creating, collecting or locating data including metadata descriptions and the use of logs/records and electronic laboratory notebooks; processing data, including cleaning and integrity checks; analysing data, including statistical analysis and visualization; storing and publishing data including through deposit in a repository for long-term preservation; provenance and version control; data peer review and linking to journal articles; managing access to the data with licenses and rights documentation; re-using data via persistent identifiers; and data citation and data attribution metrics. A range of data tools are appearing that address particular aspects of the research data lifecycle such as the DMPTool<sup>5</sup> for data management planning and ImpactStory<sup>6</sup> for collecting impact evidence and metrics. The Open Science Center has launched a tool called the Open Science Framework (OSF), which is positioned as an open and collaborative project management tool. OSF aims to integrate with other data workflow and research infrastructure components (e.g. data repositories such as figshare<sup>7</sup>), and thereby increase transparency in the practice of science. Other examples of research infrastructures which support transparency are open source code hubs, open workflow tools, open repositories for data and textual publications, open lab notebooks and open discussions spaces and forums. Libraries can adopt infrastructure which supports open protocols and processes, create new library services around this open infrastructure and ensure that library curation workflows support transparency.

#### 5.4. Workforce Development

The need to re-engineer LIS education to deliver a curriculum suited to the new data science roles has been noted (Lyon & Brenner, 2015). iSchool programs and courses require real-world relevance in order to produce work-ready graduates who can assume one of these new data science positions. An analysis of the educational requirements, skills, knowledge, and competencies from recent job descriptions for data librarians, data archivists and data stewards has identified the range of themes and topics in scope (Lyon, Mattern, Acker, & Langmead, n.d.). Many iSchools now have data curation or digital stewardship courses, but are transparency and reproducibility concepts embedded in the curriculum? The School of Information Sciences (iSchool) University of Pittsburgh MLIS Program is adopting an innovative translational data science approach (*‘the transition of data skills, software tools and research intelligence from the iSchool to the marketplace’* defined in Lyon and Brenner, 2015), which mirrors the established terminology of translational medicine. New Masters courses in Research Data Management and Research Data Infrastructures address transparency, reproducibility and validation concepts. The aim is to produce transparency-savvy LIS graduates and to upskill current LIS staff for these new data science roles.

### 6. Summary and Future Work

This short paper has begun to explore the emerging narrative associated with research transparency and has particularly focused on the implications and opportunities for libraries. The motivations for addressing transparency as a concept have been articulated and the development of open science policy which embraces transparency principles, has been described. The importance of transparency within research practice and associated scholarly communications, has been highlighted. Within this area, the current “retraction epidemic” has been noted together with new initiatives to assess reproducibility and replicability of key studies in certain domains. Prior models and expositions of open science have been described and in particular, a two-dimensional continuum approach has been cited as a basis for the 3D-model of Open Science presented in this paper, which has a transparency dimension as an additional axis. A foundational series of terms associated with transparency has been proposed and situated within a research lifecycle. The

implications for libraries and Information services have been explored and four broad areas of potential opportunity have been identified.

However, there is much work still to be carried out to realise the full promise of open science. At the practice level, a more detailed analysis and exposition of the ways in which transparency can be achieved, in terms of specific actions or interventions by particular transparency agents throughout the research lifecycle, is needed. The TOP Guidelines go some way towards this goal, but the role of libraries and librarians has not been addressed. As a potentially critical 'Transparency Agent', a data librarian may be able to advocate, train, guide and support researchers in following recommended transparency standards.

The question of how will libraries and librarians react to this new policy and practice objective of 'transparency' in open science and open data, remains to be investigated. It is acknowledged that there is a need for further research into transparency perspectives and perceptions, as well as into curriculum development and graduate education in this context. To this end, a new strand of qualitative research at the iSchool, University of Pittsburgh, is investigating the attitudes, awareness and activities of academic librarians towards research transparency and open science. It is hoped that the findings will inform future directions for institutional data policy, research data services and educational curriculum development.

## 7. Acknowledgement

This paper was accepted and presented at the LIBER 2015 conference in London.

## References

American Meteorological Society. (2013). *Full and open access to data*. Retrieved January 19, 2015, from <https://www2.ametsoc.org/ams/index.cfm/about-ams/ams-statements/statements-of-the-ams-in-force/full-and-open-access-to-data/>.

Borgman, C.L. (2015). *Big data, little data, no data: Scholarship in the networked world*. Cambridge, MA: MIT Press.

Carey, B. (2011, November 2). Fraud case seen as red flag for psychology research. *The New York Times*. Retrieved September 10, 2015, from <http://www.nytimes.com/2011/11/03/health/research/noted-dutch-psychologist-stapel-accused-of-research-fraud.html? r=0>.

Center for Open Science. (2015). *Guidelines for transparency and openness promotion (TOP) in journal policies and practices "The TOP guidelines"*. Retrieved January 19, 2016, from <https://osf.io/ud578/>.

Cicerone, R.J. (2015). *Research reproducibility, replicability, reliability* (Speech). Retrieved January 19, 2016, from <http://www.nasonline.org/about-nas/leadership/president/2015-presidents-address.pdf>.

Collins, F.S., & Tabak, L.A. (2014). NIH plans to enhance reproducibility. *Nature*, 505, 612–613. Retrieved September 10, 2015, from [http://www.nature.com/polopoly\\_fs/1.14586!/menu/main/topColumns/topLeftColumn/pdf/505612a.pdf](http://www.nature.com/polopoly_fs/1.14586!/menu/main/topColumns/topLeftColumn/pdf/505612a.pdf).

Corrall, S., & Pinfield, S. (2014) Coherence of “open” initiatives in higher education and research: Framing a policy agenda. In *iConference 2014 Proceedings* (pp. 293–313). doi:10.9776/14085. Retrieved September 10, 2015, from [https://www.ideals.illinois.edu/bitstream/handle/2142/47316/085\\_corrVersion.pdf?sequence=5](https://www.ideals.illinois.edu/bitstream/handle/2142/47316/085_corrVersion.pdf?sequence=5).

Corrall, S., Kennan, M.A., & Afzal, W. (2013). Bibliometrics and research data management services: Emerging trends in library support for research. *Library Trends*, 61(3), 636–674. Retrieved September 10, 2015, from [http://d-scholarship.pitt.edu/18948/1/08\\_61\\_3\\_corrall\\_636-674.pdf](http://d-scholarship.pitt.edu/18948/1/08_61_3_corrall_636-674.pdf).

Cox, A.M., & Pinfield, S. (2014). Research data management and libraries: Current activities and future priorities. *Journal of Librarianship and Information Science*, 46(4), 299–316. doi:10.1177/0961000613492542.

Denker, S.P. (2016, January 6). *A proactive approach to reproducibility with evidence-based research on research*. PLOS Blog post January 6 2016. Retrieved 19 January, 2016, from <http://blogs.plos.org/plos/2016/01/a-proactive-approach-to-reproducibility-with-evidence-based-research-on-research/>.

Easterbrook, S.M. (2014). Open code for open science? *Nature Geoscience*, 7, 779–781. doi:10.1038/ngeo2283. Retrieved September 10, 2015, from <http://www.nature.com/ngeo/journal/v7/n11/pdf/ngeo2283.pdf>.

Errington, T.M., Iorns, E., Gunn, W., Tan, F.E., Lomax, J., & Nosek, B.A. (2014). An open investigation of the reproducibility of cancer biology research. *eLife* 2014, 3, e04333. doi:10.7554/eLife.04333. Retrieved September 10, 2015, from <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4270077/>.

Etzioni, A. (2010). Is transparency the best disinfectant? *Journal of Political Philosophy*, 18(4), 1–16. doi:10.1111/j.1467-9760.2010.00366.x. Retrieved September 10, 2015, from <http://www.gwu.edu/~ccps/etzioni/documents/295Transparency.pdf>.

Fang, F.C., Steen R.G., & Casadevall, A. (2013). Misconduct accounts for the majority of retracted scientific publications. *Proceedings of the National Academy of Sciences of the United States of America*, 109(42), 17028–17033. Retrieved September 10, 2015, from doi:10.1073/pnas.1212247109.

FASEB. (2016, January 14). *Enhancing research reproducibility: Recommendations from the Federation of American Societies for Experimental Biology*. Retrieved January 19, 2016, from: [http://www.faseb.org/Portals/2/PDFs/opa/2016/FASEB\\_Enhancing%20Research%20Reproducibility.pdf](http://www.faseb.org/Portals/2/PDFs/opa/2016/FASEB_Enhancing%20Research%20Reproducibility.pdf).

Gezelter, J.D. (n.d.). *Open science and verifiability*. Retrieved September 10, 2015, from <http://web.stanford.edu/~vcs/Nov21/dg-OpenScienceandVerifiability.pdf>.

Gov.UK. (2013). *G8 Open data charter and technical annex*. Retrieved January 19, 2016, 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, February 24). *President's memorandum on transparency and open government – Interagency collaboration*. Retrieved September 10, 2015, from [https://www.whitehouse.gov/sites/default/files/omb/assets/memoranda\\_fy2009/m09-12.pdf](https://www.whitehouse.gov/sites/default/files/omb/assets/memoranda_fy2009/m09-12.pdf).

ICSU. (2015). *Open data in a big data world. An international accord*. Retrieved January 19, 2016, from <http://www.icsu.org/science-international/accord/open-data-in-a-big-data-world-long>.

Ince, D. (2011) The Duke University scandal – what can be done? *Significance*, 8(3), 113–115. Retrieved September 10, 2015, from doi:10.1111/j.1740-9713.2011.00505.x.

Lauer, M. (2015, October 30). *Bolstering trust in science through rigorous standards*. NIH Open Mike Blog post. Retrieved January 19, 2016, from <https://nexus.od.nih.gov/all/2015/10/30/bolstering-trust-in-science-through-rigorous-standards/>.

Lyon, L. (2009). *Open science at web-scale: Optimising participation and predictive potential consultative report*. Retrieved September 10, 2015, from <http://opus.bath.ac.uk/30056/1/open-science-report-6nov09-final-sentojisc.pdf>.

Lyon, L., & Beaton, B. (2015). Citizen science, open access, open data, and research inclusivity. In *2015 ALISE Annual Conference*, Chicago. Abstract retrieved September 10, 2015, from <https://alise2015.sched.org/event/a881a3eeb8aac9e003eb3831916bb92a#.VfGyBZfWyo>.

Lyon, L., & Brenner, A. (2015). Bridging the data talent gap: Positioning the iSchool as an agent for change. *International Journal of Digital Curation* 10(1), 111–122. doi:10.2218/ijdc.v10i1.349. Retrieved September 10, 2015, from <http://www.ijdc.net/index.php/ijdc/article/viewFile/10.1.111/384>.

Lyon, L., Mattern, E., Acker, A., & Langmead, A. (n.d.) *Applying translational principles to data science curriculum development*. In: iPREs 2015 Conference, Chapel Hill, North Carolina. (To be published in the Phaidra Repository).

- Marcus, A., & Oransky, I. (2012, August 1). Bring on the Transparency Index. *The Scientist*, 24. Retrieved September 10, 2015, from <http://www.the-scientist.com/?articles.view/articleNo/32427/title/Bring-On-the-Transparency-Index/>.
- Mayernik, M.S., Callaghan, S., Leigh, R., Tedds, J., & Worley, S. (2014). Peer review of datasets: When, why and how. *Bulletin of the American Meteorological Society*, 96,191–201. doi:10.1175/BAMS-D-13-00083.1. Retrieved September 10, 2015, from <http://journals.ametsoc.org/doi/pdf/10.1175/BAMS-D-13-00083.1>.
- NIH. (2015a, June 9). *Enhancing reproducibility through rigor and transparency*. Notice Number NOT-OD-15-103. Retrieved January 19, 2016, from <http://grants.nih.gov/grants/guide/notice-files/NOT-OD-15-103.html>.
- NIH. (2015b, November 2) *Rigor and reproducibility*. Retrieved January 19, 2016, from <http://grants.nih.gov/reproducibility/index.htm>.
- NIH. (2016, January 28). *Scientific rigor in NIH grant applications*. Blog post by Mike Lauer. Retrieved January 19, 2016, from <http://nexus.od.nih.gov/all/2016/01/28/scientific-rigor-in-nih-grant-applications/>.
- Nosek, B.A. (2012). An open, large-scale collaborative effort to estimate the reproducibility of psychological science. *Perspectives on Psychological Science*, 7(6), 657–660. doi: 10.1177/1745691612462588. Retrieved September 10, 2015, from <http://pps.sagepub.com/content/7/6/657.full.pdf+html>.
- Nosek, B.A., & Lakens, D. (2014). Registered reports: A method to increase the credibility of published results [Editorial]. *Social Psychology*, 45(3), 137–141. Retrieved September 10, 2015, from doi:10.1027/1864-9335/a000192.
- OECD. (2007). *OECD principles and guidelines for access to research data from public funding*. Retrieved September 10, 2015, from <http://www.oecd.org/sti/sci-tech/38500813.pdf>.
- OECD. (2015). *Making open science a reality*. Retrieved January 19, 2016, from <https://www.innovationpolicyplatform.org/content/open-science>.
- Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, 349(6251), aac4716. doi:10.1126/science.aac4716. Retrieved January 19, 2016, from <http://science.sciencemag.org/content/349/6251/aac4716.full>.
- Peng, R.D. (2011). Reproducible research in computational science. *Science*, 334(6060), 1226–1227. doi:10.1126/science.1213847. Retrieved September 10, 2015, from <http://www.sciencemag.org/content/334/6060/1226.full.pdf>.
- RCUK. (2015). RCUK Common Principles on Data Policy. Retrieved September 10, 2015, from <http://www.rcuk.ac.uk/research/DataPolicy/>.
- Royal Society Report. (2012). *Science as an open enterprise: open data for open science*. Retrieved September 10, 2015, from [https://royalsociety.org/~media/Royal\\_Society\\_Content/policy/projects/sape/2012-06-20-SAOE.pdf](https://royalsociety.org/~media/Royal_Society_Content/policy/projects/sape/2012-06-20-SAOE.pdf).

- Si, L., Xing, W., Zhuang, X., Hua, X., & Zhou, L. (2015) Investigation and analysis of research data services in university libraries. *The Electronic Library*, 33(3), 417–449. doi:10.1108/EL-07-2013-0130.
- Stodden, V. (2009). *Enabling reproducible research: open licensing for scientific innovation*. Social Science Research Network. Retrieved September 10, 2015, from [http://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=1362040](http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1362040).
- 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 December 10–14, 2012, Providence. Retrieved September 10, 2015, from [http://stodden.net/icerm\\_report.pdf](http://stodden.net/icerm_report.pdf).
- Stodden, V., Leisch F., & Peng, R.D. (2014). *Implementing reproducible research*. Boca Raton, FL: Chapman & Hall/CRC Press. Available as an eBook.
- Taylor, B.N., & Kuyatt, C.E. (1994). *Guidelines for evaluating and expressing the uncertainty of NIST measurement results*. NIST Technical Note 1297. Retrieved September 10, 2015, from <http://physics.nist.gov/Pubs/guidelines/TN1297/tn1297s.pdf>.
- The Economist. (2010, May 18). *The science of climate change: The clouds of unknowing*. Retrieved September 10, 2015, from <http://www.economist.com/node/15719298>.
- The Economist. (2013, October 17). *Unreliable research: Trouble at the lab*. Retrieved September 10, 2015, from <http://www.economist.com/news/briefing/21588057-scientists-think-science-self-correcting-alarming-degree-it-not-trouble>.
- University of Bath. (2014). *Research Data Policy*. Retrieved September 10, 2015, from <http://www.bath.ac.uk/research/data/policy/research-data-policy.html>.

## Notes

- 
- <sup>1</sup> Science Exchange Validation Service <http://validation.scienceexchange.com/#/>.
- <sup>2</sup> <https://publons.com/journal/530/gigascience>.
- <sup>3</sup> Research Data Alliance <https://rd-alliance.org/>.
- <sup>4</sup> Open Science Framework <https://osf.io/>.
- <sup>5</sup> DMPTool <https://dmptool.org/>.
- <sup>6</sup> ImpactStory <https://impactstory.org/>.
- <sup>7</sup> Figshare <http://figshare.com/>.