

Identifying Design Requirements of a User-Centered Research Data Management System

Maryam Bugaje¹[0000-0002-6743-1943] and Gobinda Chowdhury¹[0000-0002-2122-9547]

¹ Department of Computer & Information Sciences, Northumbria University, Newcastle, UK
{maryam.bugaje,gobinda.chowdhury}@northumbria.ac.uk

Abstract. Research data repositories perform many useful functions, the key ones being the storage of research datasets, and making the same discoverable for potential reuse. Over the years, various criteria for assessing the user-centeredness of information systems have been developed and standards have gradually been improved. However, there has been less development in case of research data management (RDM) systems. By means of a combination of user-focused research methods viz. questionnaire surveys, face-to-face interviews, a systematic appraisal of existing services and a technical experiment, we have sought to understand the meaning of user-centeredness pertaining to research data repositories, and identify some key indicators of it. We have furthermore translated our findings into design requirements based on which we propose to develop and test a prototype of a user-centered RDM system. This paper reports on how we identified the design requirements that would make the RDM systems more user-centered.

Keywords: User-Centered Design, Research Data Management, Information Retrieval, Metadata, Research Data Repositories, Scientific Data.

1 Introduction

Research data repositories are an integral component of the RDM ecosystem, that combines all the essential functions of data management throughout the data lifecycle [1-4]. In addition to storing and retrieving data, and ensuring that it is discoverable and accessible, the burden of preservation and curation ultimately falls into the hands of repositories [3]. Recent data sharing mandates from research publishers, funding agencies, and governments have been the driving force behind the emergence of a host of research data repositories to meet the growing demand. However, many of which repositories are barely in a position to fulfill the comprehensive role of data disseminator and curator adequately enough to enable the realization of the benefits which were primary motivation of the mandates.

Research data repositories are necessarily end-user systems; the end-users being, among others (a) researchers in the roles of data producer and data consumer, (b) research funders who are keen to know the impact of their grants, (c) practitioners, (d) data scientists, and (e) other interested parties, including the general public. Making data publicly available, and making them discoverable and usable are different issues,

and in order to fully realize the benefits of data sharing [5-10], repositories must provide users with more than merely open access to research data. The importance of user-centeredness has generally been acknowledged as being central to the success of RDM systems [1,11,12], although as yet there aren't any clearly-defined criteria, established standards or guidelines for assessing specifically their usability or user-centeredness. Notwithstanding this lacuna, however, various efforts among institutions and research centers towards developing RDM products have resulted in user-needs studies prior to system design and made evident attempts to accommodate those needs [1,13-15]. But, though notable, the designs of these systems still leave much room for improvement [16,17].

In the next section we define the key concepts that form the foundation upon which we may proceed. Following this are the sections on methodology and requirements analyses of a user-centered RDM, in which we detail all the preliminary work which will go into, and culminate in, the system design section, where we present the conceptual design of the alpha version of a user-centered RDM system which we hope, in a future work, to develop into a prototype. Within the scope of this paper our research aims are to:

1. Gather sufficient practical information about researchers' data-seeking needs, practices, strategies and difficulties to enable us to make appropriate design choices as well as to identify areas for improvement;
2. Based on the above, prioritize user requirements and system features for a user-friendly RDM system; and
3. Determine resource requirements and the best allocation of the same for our system.

2 Background

Research Data Management is "the organization of data, from its entry to the research cycle through to the dissemination and archiving of valuable results" [18], and Research Data Management Systems are "the technical framework to collect, describe, and provide research data" [2]. Of the numerous problems and challenges facing RDM [5,12,19-21], our work focuses on those that can be mitigated through better engagement with the user and increased attentiveness to user needs in designing RDM systems. This being the case, we distinguish between simply RDM systems and specifically user-centered RDM systems; and, expanding upon our original definition of the former, we define the latter as in addition conforming in some degree to the formal guidelines set forth by research communities and authoritative bodies, such as the UK Research & Innovation Council (UKRI) [22], The Future of Research Communication and e-Scholarship (FORCE11) [23], and others [24] to improve the reuse potential and manageability of research data products. The guiding principles recommended by UKRI and FORCE11's FAIR principles are summarized in Table 1.

2.1 User-Centered Design

Data repository users can be identified in terms of their relationship to the system (i.e. primary users, e.g. researchers; secondary users, e.g. funding bodies; and tertiary users, e.g. search engines [25]) or the role(s) that they play in it (e.g. data creators, data consumers, and data administrators [26]). To be useful, the design of an RDM system must show due consideration to the needs of its potential users. For this, it is necessary to not only understand the different user types and groups but also to have a thorough appreciation of the tasks that each wish to accomplish through the system. In user-centered design, the interest of the end user is at the core of every design decision from the inception to the implementation of the system [26].

Research shows that users' information needs, and by inference, data needs, tend to be ambiguous, not definitely articulated [27], and often only recognized at sight [28]. Concerning this, it is worth noting that:

1. Users' knowledge of systems may range from very naive to highly skilled and sophisticated [28], and systems must be designed in such a way as to sufficiently enable the less sophisticated user to efficiently search for and find data; and
2. Users' data-seeking needs, as we observed from our face-to-face interviews with researchers, may go beyond a simplistic search for datasets on a single topic (e.g. data on climate change) and may entail more complex conditions such as associative relationships (e.g. climate change data related to ozone depletion) or comparative relationships (e.g. climate change data in which ozone depletion is compared with rise in sea levels) between multiple topics. There should be a matching capability in the system to enable the highly skilled user to satisfy his or her more complex data-seeking needs.

Data discovery is largely dependent upon good metadata [29] and although data creators are the primary providers of contextual metadata [8] and other complementary information about research data, they are not necessarily skilled in data management or knowledgeable as to its technicalities [8]. This being the case, it becomes the responsibility of the system to facilitate users (data creators) in the fulfilment of this important function to enhance the reuse potential of data [30].

Table 1. RDM guidelines by UKRI and FORCE11.

Principle	UKRI Guidelines	FAIR Principles
Findability (or Discoverability)	✓	✓
Accessibility	✓	✓
Intelligibility	✓	
Assessability	✓	
Usability (or Reusability)	✓	✓
Interoperability		✓

2.2 Aims of the system

Our system aims to at least partially comply with UKRI guidelines and the FAIR principles to address thereby some of the currently existing user-related issues of RDM systems hitherto alluded to, as well as others which will subsequently be considered in more detail in ensuing sections. We highlight our principal aims as follows:

1. To help data consumers (e.g. researchers, data scientists, practitioners) efficiently *discover* data, and provide them with necessary information to *access* and *use* the data;
2. To help data creators and proprietors expose their data effectively for *discovery*;
3. To add value to data by linking it with associated publications, data, or related or similar output and by enabling user annotation; and
4. To use system design & development best practices to enable the system to *interoperate* and future-proof it in the case of advancements in the field.

In the next section we describe our methodology and requirements-gathering process.

3 Methodology

Our design is the product of a variety of research methods, each addressing a specific part of the whole body of our research aims as set forth in Section 1, and each respectively reported in full in previously published (all except the one reporting on the face-to-face interviews, which is pending publication) research papers and conferences. Due to constraints of space which prevent our repeating the same in this paper, we refer the reader to the publications, listed below, for the full accounts:

1. Online questionnaire survey; addresses aims 1, 2 & 3, and is reported in [31]
2. Face-to-face interviews; address aims 1 & 2, and is reported in [32]
3. Market appraisal and review; addresses aim 1, and is reported in [17, 26]
4. A technical experiment comparing data retrieval (DR) with traditional information retrieval (IR); addresses aims 2 & 3, and is reported in [16]

Methods 1 and 2 above (questionnaire and interview) were conducted sequentially, the purpose of the latter being chiefly to probe further into and expand upon some of the findings and hints from the former. We briefly present each of the four in turn.

3.1 Online Questionnaire Survey

Data was collected via questionnaire surveys conducted at universities UK-wide between the summer and winter terms of the 2016/2017 session. The survey garnered a total of 201 (complete) responses from researchers at various stages in their academic careers. The questions asked were mainly with a view to understanding the following:

1. The type, volume, and variety of data used and created by researchers;
2. Researchers' common practices with respect to data storage;

3. Researchers' familiarity with standards, metadata, and their university data policy;
4. Requirements and opportunities for training & support;
5. Views, perceptions, and practices pertaining to data sharing and open access; and
6. Researchers' previous experiences on, and impressions about, using research data repositories.

Details of the survey and findings have been reported in [31].

3.2 Face-to-Face Interviews

We conducted semi-structured, face-to-face, interviews with 18 researchers; 6 each from the departments of History, Solar Physics, and Information Science at a British University. The disciplines were selected on the basis of the contrasting nature of their datasets with regards to size, conformity to metadata standards, methods of data collection/generation, and data formats among others [33]. The goal was to ensure diversity by studying representatives from two polar ends of the disciplinary spectrum (History and Solar Physics) with respect to data sharing practices and the use of technology [33], as well as a middle-ground (Information Science); with a view to learning the similarities that unify all, while exploring differences that make each unique. Key questions asked include amongst others the following:

1. Where and how do you obtain data for your research? Do you employ any strategy or have a standard workflow for this?
2. What are some of the problems you've faced before in finding, using, or accessing research data, if any?
3. What data repositories have you used before or do you currently use? What motivates you to use a particular repository rather than another?
4. Have you ever uploaded your own data in an online repository? Why or why not?
5. What are your thoughts on research data sharing and open access?
6. Do you or your research group follow any metadata formats for tagging research data? What are some of the issues you've faced in this regard, if any?

3.3 Market Appraisal and Review

To facilitate our review of research data repositories, we consulted the re3data.org directory to create the following non-mutually exclusive groups into which we organized the repositories: disciplinary, institutional, publisher-service, location-based, dedicated content-type, and commercial/general purpose repositories. For each group, we hand-picked a few representative examples, based on recommendation by Nature¹, for evaluation against the following performance yard-sticks:

1. *Use of metadata*. The degree to which metadata appears to be exploited to provide features for browsing, searching/querying, filtering and result presentation;

¹ <https://www.nature.com/sdata/policies/repositories>

2. *Querying facility*. The level of expressiveness allowed in searching/querying the repository; and
3. *Result filtering*. The availability of options for filtering down search results, and the furthest granularity to which this is possible.

This study has been fully reported in [17].

3.4 A Technical Experiment Comparing DR with Traditional IR

We carried out a controlled experiment to demonstrate some fundamental differences between retrieval of data and text, and their corresponding implications on user interface design and on network and computing resource requirements. To do this, we consulted Wikipedia's broad classifications of academic disciplines and obtained 5 random keywords/phrases each of the domains of Arts & Humanities, Social Sciences, Natural Sciences, and Applied Sciences, the last of which we had represented by its sub-domain of Computer & Information Science. A search was conducted on each of the 20 total keywords/phrases for both data retrieval (using Dryad, UK Data Service, or DataOne) and text retrieval (using Thomson Reuters Web of Science database). The top 10 results of each search were noted for file size and file format(s). In computing file sizes, we considered for text retrieval (research publications) only full research papers; and for data retrieval (research datasets), both the dataset itself and any documentation(s) it comes with. Detailed findings and discussions may be found in [16,17].

4 Study Findings and User Requirements Analyses

Key findings from our studies are summarized in Tables 2 (questionnaire & interview), 3 (market appraisal and review) and 4 (experiment comparing DR and IR) below.

Table 2. Summary of findings from questionnaire survey and interview.

Key findings
Incomplete documentation or its lack altogether often prevents datasets of interest from being reusable
Many users are unskilled information seekers and are unsure as to what search terms to use to find data
Researchers commonly follow non-standard, ad hoc methods for tagging or annotating their data with metadata
Tools for creating metadata are found to be too hard to use, and very few researchers have received any degree of formal training on metadata or data management
Data file sizes in the megabyte range are the most commonly used and produced, closely followed by files in the gigabyte range. File sizes in the terabyte range are rare in most disciplines
Google is frequently used for data search, though often with unsatisfactory results
Most researchers create new primary data rather than reuse existing data. The main reason(s) given for this is lack of knowledge about or access to existing data
The process of obtaining access to data may be particularly tedious in some disciplines (e.g. History)

There is a general reluctance among researchers to upload data online before the maximum number of papers have been published on it
Standard office documents (e.g. text, spreadsheets) are the most common file formats used and produced by researchers. Next are images, structured scientific and statistical data, and web-based data (e.g. social media data)
Many researchers felt that some way of visualizing datasets would be useful in helping them understand and decide on the usefulness of data
Researchers are generally reluctant to voluntarily spend long hours tagging data to upload online unless doing so is a requirement

Table 3. Summary of findings from market appraisal & review, with their corresponding implication(s) on user-experience

Key findings	Comments	Implications
Limited user interactivity	e.g. No feature(s) for previewing dataset content on the web browser before download. This unnecessarily increases the rate of download, making each session highly resource intensive	<ul style="list-style-type: none"> – Downloading data that ends up unused unduly strains network resources – Poor use of storage space – Renders download count unreliable as a measure of dataset relevance/usefulness/impact
Insufficient or unavailable metadata	The lack of use of standard metadata to sufficiently contextualize data for discovery [34, 35] & re-use [36] is a major challenge. Deficiency in metadata quality or quantity, along with the fact that using generic metadata for greater inclusivity directly translates into loss of nuanced features, represents a delicate problem.	<ul style="list-style-type: none"> – Complex or precise queries cannot be supported – Loosely matching search results – Tedious manual browsing of results – Unproductive use of researchers' time – Threatens the discoverability and, consequently, reuse rate of research datasets
Quality of data questionable or not assured	Researchers tend to reuse the datasets of others whom they trust [37]. Many services do not have mechanisms to ensure the quality of user-uploaded datasets; nor are there any standard criteria for measuring the quality of research data.	<ul style="list-style-type: none"> – Skepticism, which may stunt the rate of data reuse – Time which could be used more productively in active research spent on making inquiries about data.

Table 4. Summary of findings from technical experiment comparing DR and IR, with resource implications of each.

Key findings	Implications
Text can be read online, while data often requires downloading prior to being “read” or used	More network (in terms of bandwidth) and storage resources are required for data retrieval.

A single data item record may constitute several composite files (as many as 524 have been noted in this experiment)	A system, e.g. metadata schema, for efficiently identifying and linking associated files is imperative
Texts (research publications) often come as a single, self-sufficient file. Data is nearly always accompanied with separate documentation files	A system, e.g. metadata schema, for efficiently identifying and linking associated files is imperative
Unlike texts (research publications), the same dataset may come in many different file types or formats	This places additional burden on computing resources (e.g. more storage is required for the same dataset) and also human resources (e.g. in terms of data preservation/curation requirements).
The average retrieved file size of datasets is typically several times larger than that of text (research publication).	More network (in terms of bandwidth) and storage resources are required for data retrieval.

4.1 User Requirements

We now proceed to extract from our findings that part of it which broadly or loosely spells out user requirements. For this, we adopt a structured approach by identifying the main themes around which the requirements are clustered, as shown in Table 5, and relating them to the UKRI guidelines and FAIR principles from Table 1 to which they contribute to conforming.

Table 5. Summary of user requirements and how they meet formal guidelines in the literature.

Broad theme	List of requirements	Guideline or Principle
Search and discovery	Linking data to associated research publications and similar datasets	Findability
	Query expansion features to help users who are unsure of the precise search terms to use	Findability
	Optimizing repositories for search engines to index, for the benefit of those who use traditional search engines like Google for data search	Findability, Interoperability
	Require the provision of some metadata with each uploaded dataset	Intelligibility, Assessability, Reusability
	Provide support for granular search, e.g. through enabling Boolean queries	Findability
	Use reliable methods for ranking research datasets based on different criteria useful to the user in making decisions	Assessability
Access	Provide clear and full information on how to obtain access to data	Accessibility
User interface	Recognize and adequately cater to the different types of users and their roles (e.g. the typical researcher vs a research funding agent)	Findability
	Follow standards and best practices in designing and laying out the interface	Findability, Interoperability

Other special features	Provide data visualization plugins	Intelligibility, Assessability
	Provide features that allow some interaction with datasets on the web-browser before download, at least for commonly used file formats.	Assessability
	Enable users to be able to make their data private or invisible to the general public during peer review (embargo period)	Accessibility
	Ensure there are clear statements about the relationship of each associated file(s) to the dataset in question	Intelligibility
	Offer automated or semi-automated methods for extracting or tagging metadata	Findability

5 Functional Requirements and Prototype Design

Focusing on the fourth aim of this study, the design proposed in this section was informed by the user requirements gathered from our studies and listed in Table 5. A simplified diagram of how the main components of our design are related and work together is given in Fig. 1 and explained in further detail in this section.

Persistent identification. Our system will use a combination of Digital Object Identifiers (DOI) and the ORCID(s) of the data creator(s) as a means to uniquely identify data objects in the database.

Metadata. We collected the metadata elements of over 18 different data upload templates from institutional and non-institutional data repositories, as well as those recommended in the literature [38, 39]. From these, we selected those elements common across all templates (Table 6) for use in our proposed system. The uploading of disciplinary metadata will also be encouraged and supported where available.

Ontological schema. Aside from the metadata schema, which describes the data object alone, we shall additionally use an RDF/XML ontological schema to establish the relationship between data objects to enhance discovery by linking and recommendation.

Table 6. Common metadata elements across all templates.

Mandatory elements	Optional elements
Title	Funder
Depositor Name	Data License Type (e.g. creative commons)
Data Publisher	Date
Discipline	Keywords
	Data Description
	Related Publication URI (or URL)
	Publication Title

Searching facility. Aside from supporting advanced search by variables culled from available metadata, dictionary look-up will be used for query expansion. All of the findability features identified in Table 5 will also be supported.

Search results ranking facility. There will be options for users to rank returned search results by number of views or downloads and number of publications linked to it.

Search results filtering facility. Users will be given the option to filter search results based on variables culled from available metadata.

Test collection. To test our prototype, we have built a small test collection consisting of over 200 datasets carefully described with at least the mandatory metadata elements specified in Table 6 above.

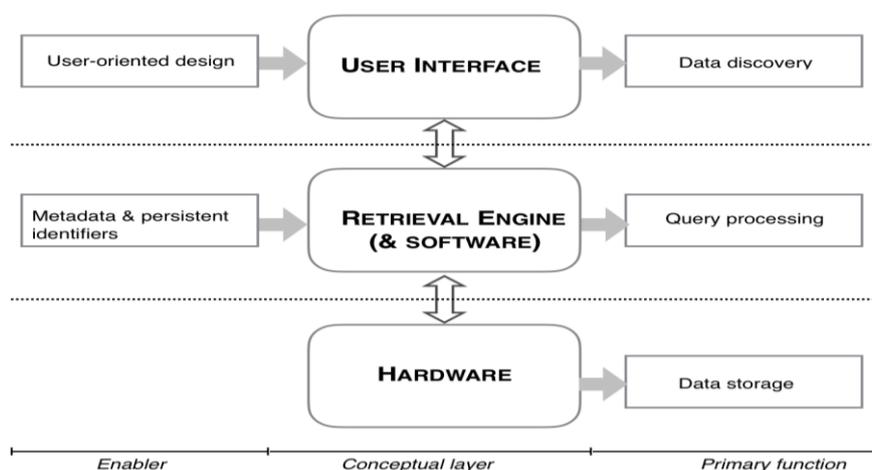


Fig. 1. A simplified conceptual diagram showing the workings of our system

6 Conclusion

The paper presents various thoughts and principles that went into the design of our user-oriented RDM system. It shows how the selection of our metadata elements [38, 39], and user requirements informing our design (Table 5), additionally comply with guidelines recommended in literature. Moreover, our metadata selection, being collated from the elements common to over 18 of the major institutional and non-institutional research data repositories, will enable interoperability. Some functionalities are more difficult to implement than others. However, we hope in a future work to demonstrate an actual prototype and present results from user evaluation studies to be conducted.

References

1. Arend D, Lange M, Chen J et al. e!DAL - a framework to store, share and publish research data. BMC Bioinformatics. 2014;15(1):214. doi:10.1186/1471-2105-15-214.

2. Curdt C, Hoffmeister D. Research data management services for a multidisciplinary, collaborative research project. *Program*. 2015;49(4):494-512. doi:10.1108/prog-02-2015-0016.
3. Cox A, Pinfield S. Research data management and libraries: Current activities and future priorities. *Journal of Librarian*
4. Amorim R, Castro J, Rocha da Silva J, Ribeiro C. A comparison of research data management platforms: architecture, flexible metadata and interoperability. *Universal Access in the Information Society*. 2016;16(4):851-862. doi:10.1007/s10209-016-0475-y.
5. Patel, D. (2016). Research data management: a conceptual framework. *Library Review*, 65(4/5), pp.226-241.
6. Perrino, T., Howe, G., Sperling, A., Beardslee, W., Sandler, I., Shern, D., Pantin, H., Kaupert, S., Cano, N., Cruden, G., Bandiera, F. and Brown, C. (2013). Advancing Science Through Collaborative Data Sharing and Synthesis. *Perspectives on Psychological Science*, 8(4), pp.433-444.
7. The Royal Society (2012). Science as an open enterprise. [online] Available at: <https://royalsociety.org/~media/policy/projects/sape/2012-06-20-saoe.pdf> [Accessed 11 Jun. 2018].
8. Borgman, C. (2011). The Conundrum of Sharing Research Data. *SSRN Electronic Journal*.
9. Costello, M. (2009). Motivating Online Publication of Data. *BioScience*, 59(5), pp.418-427.
10. Faniel, I. and Jacobsen, T. (2010). Reusing Scientific Data: How Earthquake Engineering Researchers Assess the Reusability of Colleagues' Data. *Computer Supported Cooperative Work (CSCW)*, 19(3-4), pp.355-375.
11. Carlson J. Demystifying the data interview. *Reference Services Review*. 2012;40(1):7-23. doi:10.1108/00907321211203603.
12. Mückschel, C., J. Nieschulze, C. Weist, B. Sloboda & W. Köhler (2007): Herausforderungen, Probleme und Lösungsansätze im Datenmanagement von Sonderforschungsbereichen. In: eZAI (elektronische Zeitschrift für Agrarinformatik) 2, 1-16.
13. Curdt C, Hoffmeister D, Waldhoff G, Jekel C, Bareth G. Scientific Research Data Management for Soil-Vegetation-Atmosphere Data – The TR32DB. *International Journal of Digital Curation*. 2012;7(2):68-80. doi:10.2218/ijdc.v7i2.208.
14. bioCADDIE | biomedical and healthCAre Data Discovery and Indexing Ecosystem. Available at: <https://biocaddie.org>. Accessed June 12, 2018.
15. Research Data Discovery Service - Laying the firm foundations for a Jisc UK Research Data Discovery Service. Available at: <https://rdds.jiscinvolve.org/wp/>. Accessed June 12, 2018.
16. Bugaje, M., Chowdhury, G.: Is data retrieval different from text retrieval? An exploratory study. In: Choemprayong, S., Crestani, F., Cunningham, S.J. (eds.) ICADL 2017. LNCS, vol. 10647, pp. 97–103. Springer, (2017). https://doi.org/10.1007/978-3-319-70232-2_8
17. Bugaje, M., Chowdhury, G.: Data Retrieval = Text Retrieval? In: Chowdhury, G., McLeod, J., Willett, P., Gillet, V. (eds.) iConference 2018. LNCS, vol. 10766, pp. 253–262. Springer (2018). https://doi.org/10.1007/978-3-319-78105-1_29
18. Whyte A, Tedds J. Making the Case for Research Data Management. Digital Curation Centre. dccacuk. 2011. Available at: <http://www.dcc.ac.uk/resources/briefing-papers/making-case-rdm>. Accessed June 12, 2018.
19. Santos C, Blake J, States D. Supplementary data need to be kept in public repositories. *Nature*. 2005;438(7069):738-738. doi:10.1038/438738a.
20. Sallans A, Lake S. Data management assessment and planning tools. *Research Data Management: Practical Strategies for Information Professionals*. West Lafayette, Indiana: Purdue University Press; 2014:87-107.
21. Dumontier M, Gray A, Marshall M et al. The health care and life sciences community profile for dataset descriptions. *PeerJ*. 2016;4:e2331. doi:10.7717/peerj.2331.

22. UK Research and Innovation. Concordat on Open Research Data.; 2016. Available at: <https://www.ukri.org/files/legacy/documents/concordatonopenresearchdata-pdf/>. Accessed June 13, 2018.
23. Boeckhout M, Zielhuis G, Bredenoord A. The FAIR guiding principles for data stewardship: fair enough? *European Journal of Human Genetics*. 2018;26(7):931-936. doi:10.1038/s41431-018-0160-0.
24. Starr J, Castro E, Crosas M et al. Achieving human and machine accessibility of cited data in scholarly publications. *PeerJ Computer Science*. 2015;1:e1. doi:10.7717/peerj-cs.1.
25. Alsos, O. A., Svanæs, D.: Designing for the Secondary User Experience. In: Campos, P., Graham, N., Jorge, J., Nunes, N., Palanque, P., Winckler, M. (eds.) *Human-Computer Interaction - INTERACT 2011*. LNCS, vol. 6949, pp. 84–91. Springer, Berlin, Heidelberg (2011). https://doi.org/10.1007/978-3-642-23768-3_7
26. Bugaje, M., Chowdhury, G.: Towards a More User-Centered Design of Research Data Management (RDM) Systems [abstract]. In: *Information: Interactions and Impact (i3)*.; 27-30 June 2017; Aberdeen; 2017:53-55.
27. Taylor R. Question-Negotiation and Information Seeking in Libraries. *College & Research Libraries*. 2015;76(3):251-267. doi:10.5860/crl.76.3.251.
28. Morris R. Toward a user-centered information service. *Journal of the American Society for Information Science*. 1994;45(1):20-30. doi:10.1002/(sici)1097-4571(199401)45:1<20::aid-asi3>3.0.co;2-n.
29. Willis C, Greenberg J, White H. Analysis and synthesis of metadata goals for scientific data. *Journal of the American Society for Information Science and Technology*. 2012;63(8):1505-1520. doi:10.1002/asi.22683.
30. Van Noorden R. Data-sharing: Everything on display. *Nature*. 2013;500(7461):243-245. doi:10.1038/nj7461-243a.
31. Chowdhury G, Walton G, Bugaje M.: Research Data Management: Practices, Skills and Training Needs of University Researchers in the UK. In: Špiranec S, Bartol T, Stopar K, Boh Podgornik B. (eds.) *The Fifth European Conference On Information Literacy (ECIL) 2017*. Saint-Malo: Information Literacy Association (InLitAs); 2017:30
32. Bugaje, M., Chowdhury, G: Disciplinary Contexts in Research Data Management: A Case-Study of Three Disciplines. (accepted contribution). *The Fifth European Conference On Information Literacy (ECIL) 2018*. Finland.
33. Borgman C. *Big Data, Little Data, No Data*. 1st ed. Cambridge, Massachusetts: The MIT Press; 2015:81-161.
34. Boru, D., Kliazovich, D., Granelli, F., Bouvry, P., Zomaya, A.Y.: Energy-efficient data replication in cloud computing datacenters. *Cluster Computing*, 18(1), 385-402 (2015)
35. Chowdhury, G.G.: *Sustainability of scholarly information*. Facet Publishing, London, (2014).
36. Weber, A.; Piesche, C. Requirements on Long-Term Accessibility and Preservation of Research Results with Particular Regard to Their Provenance. *ISPRS Int. J. Geo-Inf*. 2016,5,49.
37. Research Information Network (RIN). *To Share or Not To Share: Publication and Quality Assurance Of Research Data Outputs*.; 2008:48. Available at: <http://www.rin.ac.uk/system/files/attachments/To-share-data-outputs-report.pdf>. Accessed June 25, 2018.
38. Rumsey, S. and Jefferies, N. (2013) ‘Challenges in building an institutional research data catalogue’, *International Journal of Digital Curation*, 8(2), pp. 205–214. doi: 10.2218/ijdc.v8i2.284.
39. Weibel S. The Dublin Core: A Simple Content Description Model for Electronic Resources. *Bulletin of the American Society for Information Science and Technology*. 2005;24(1):9-11. doi:10.1002/bult.70.