Journal of Universal Computer Science, vol. 23, no. 11 (2017), 1098-1108
submitted: 1/9/16, accepted: 26/11/17, appeared: 28/11/17 © *J.UCS*

(Big) Data in Library and Information Science: A Brief Overview of Some Important Problem Areas

Koraljka Golub
(Department of Library and Information Science, School of Cultural Sciences
Linnaeus University, Växjö, Sweden
koraljka.golub@lnu.se)

Joacim Hansson
(Department of Library and Information Science, School of Cultural Sciences
Linnaeus University, Växjö, Sweden
joacim.hansson@lnu.se)

Abstract: Libraries hold a long history of a multidimensional focus on collecting, storing, organizing, preserving and providing access to information resources for various types of users. Data is nothing new to Library and Information Science (LIS) and Big Data presents a quantitative expansion of an already well-known object of study. Scholarly communication, data sharing and data curation are three areas related to data in LIS and are discussed in this paper in the light of current developments as well as from the perspective of attaining the research area relevance in the discipline over time. Big Data, new technologies and networked research environments will continue to increase both in numbers and size. LIS is rapidly developing tools to meet the opportunities arising – through educational initiatives and the development of new research areas such as data curation and altmetrics. Since social and political demands for open data grow, these issues are pressing.

Key Words: Big Data, Library and Information Science, Scholarly Communication, Bibliometrics, Data Sharing, Data Curation
Category: E.0. Data – General

1 Introduction

The ever growing amount of data and both scientific and political interests in making them openly available has sparked major interest in many parts of the research community, not just those that are traditionally related to data production and curation. Some disciplines have been concerned with problems of data curation and organization for long, although as part of an overall interest in the social and scientific organization and dissemination of information. Library and Information Science (LIS) is such a discipline. The purpose of this paper is to illustrate LIS developments and challenges particular to the world of (Big) Data to readers from computer science and related disciplines, with a specific focus on defining a general problem picture, a contextualization within which data research finds its place among other subfields of LIS[1]. Otherwise, whole books have been written on different aspects of this deserving

[1] This paper is part of a focused topic section of J.UCS devoted to the development of Big Data research at Linnaeus University in Sweden.

topic, and interested readers are referred to them for further elaboration; see, for example, [Borgman 2015], on digital scholarship, [Oliver, Harvey 2016] on digital curation; [Farmer, Safer, 2016] for improving library itself based on data analytics. Our mission here is much more humble and aims to contribute to the discourse of Big Data as it evolves in the development specific to LIS, but also bringing in the perspective of the discipline into a computer science context where usually other issues concerning data curation and organisation lie at its prime focus.

The paper is structured as follows: section 2 sketches LIS as a scientific discipline and intellectual environment for Big Data studies; section 3 discusses Big Data in three subfields of LIS; bibliometrics, data sharing and data curation; the paper ends with some concluding remarks (section 4).

2 The Intellectual Environment of Library and Information Science

Libraries hold a long history of a multidimensional focus on collecting, storing, organizing, preserving and providing access to information resources for various types of users [Lancaster 2003]. To this purpose, information retrieval systems, knowledge organization systems and metadata standards, among others, are developed through applied research to support the various aspects of information resource management ([Svenonius 2000], [Ingwersen, Järvelin 2006], [Markey 2015], [Zeng, Quin 2016]). Today the *library science* is more and more becoming the *information science* (also referred to as *information studies* or *library and information science* to use a combinatory term) in the sense that it goes beyond the traditional library resources to embrace the challenges of managing data, information, knowledge of all kinds, in different environments and for a variety of users and uses in the local and global, private and public information environments ([Chowdhury, Chowdhury 2007], [Abbas 2010], [Golub 2014]). However, academic discussions on the library science versus information science, or merging of the two have been ongoing and a consensus has not been reached (see, e.g., [Miksa 1992], [Hansson 2004], [Saracevic 2009]).

LIS being placed at the intersection of technology, social sciences and humanities, it often focuses on qualitative aspects of mentioned applications through, for instance:

- Subject analysis (human and automated classification and indexing, social tagging);
- Bibliographic analysis (cataloguing, metadata structuring);
- Social analysis (analysis of social and institutional structures for information, documentation and knowledge such as libraries and information practices in various communities, both online and in the physical world); and,
- Cognition related analysis (information behaviour); however, a note worthwhile making here is that not all information behaviour research is considered to be rooted in a cognitive paradigm, as seen, e.g., in [Pettigrew, Fidel, Bruce 2001].

The width of research problems and theoretical propositions that emerge from this rich disciplinary structure have caused a continuous discussion on the emphasis and general position of the discipline within the scholarly landscape. Approaches have varied. In [Nolin, Åström 2010] it is argued that, although the disparate problem areas and theoretical positions that naturally arise inform the discipline that is simultaneously at once so narrow and so diverse, may seem weak, they can also be seen as token of strength in the light of current social and technological advances. They point to three main advantages in relation to traditionally strong disciplines with limited interdisciplinarity:

1. LIS has a long tradition of interdisciplinary study of information. Increased complexity in information structures and research favours research that is not 'locked up' in its own theoretical and methodological traditions.
2. LIS is often strongly and explicitly related to the study of information in other disciplines such as computer science, pedagogy and psychology, to name a few. This enable LIS researchers to contribute to advances in various fields.
3. LIS has a strong tradition of self reflection. In Kuhnian terms [Kuhn 1996], this would be seen as a weakness, but in contemporary society it must be considered a strength as it enables the discipline to address its main problem areas in a strategically fruitful way.

Another example of this self-reflective tradition is [Buckland 2012], who, however, takes it into a somewhat different direction than Nolin and Åström, raising the question "what kind of science *can* information science be" in the very title of his paper. From a perspective that includes the study of documentation as equally central as that of information, Buckland concludes that LIS is first and foremost concerned with cultural engagement: "Formal and quantitative approaches are extremely valuable, but the field in itself is incorrigibly cultural" [Buckland 2012, p. 6]. As formal and quantitative approaches are here meant, for instance, bibliometrics and information retrieval, point being that as much as LIS researchers happily engage in such areas, it is rarely the technology itself that stands in the spotlight, but rather use and value of various solutions, systems and applications. Buckland's position relates to a line of thought in LIS giving priority to the materiality of information, primarily through documents and documentation ([Day 2001], [Day 2006], [Frohmann 2004]). And, where there is documentation there is also data, for instance, in that both have complex historical relations to 'facts'. The distinction between data, documents and information is complex; specifically, the difference between documents and data as carriers of evidence (and/or information) has proven problematic [Furner 2016].

Data are thus nothing new to LIS. One thing that distinguishes this discipline from most others is, however, that data is seen not just as the result of empirical work or the raw material for statistical analysis, but as a research object in its own right. In [Rowley 2007] it is placed in a data-information-knowledge-wisdom hierarchy, where information is defined as data structured in a way that is helpful for analysis and understanding of a certain phenomenon. This placement of data resembles that of [Buckland 1991] who, when defining the following three aspects of information, 1) information-as-knowledge, 2) information-as-thing, 3) information-as-process, places data in the second grouping, 'information-as-thing', together with 'documents' and 'recorded knowledge'. In this context, data is defined as 'information-as-thing' which has been processed in some way for use (ibid., p. 45).

3 Big Data in Library and Information Science

In an intellectual environment such as this, 'Big Data' basically becomes nothing more than a quantitative expansion of an already well-known object of study. Big Data is simply data, but a lot of it. One of the main conceptual characteristics of the Big Data trope is that a sufficiently clear definition of Big Data is hard to find. In [De Mauro, Greco, Grimaldi 2016)] a synthesised definition is provided, based on 15 formal definitions categorised into 4 groups of essential features: 1) information, 2) technology, 3) methods, and, 4) impact. Drawing from these features they offer the following formal definition:

> "Big Data is the Information asset characterized by such High Volume, Velocity and Variety to require specific Technology and Analytical Methods for its transformation into Value" (ibid., p. 131).

The issue of creating value can be said to be at the very core of the analysis of Big Data and we will here highlight three broad areas of inquiry which are either influenced by or immediately depending on Big Data for their development: 1) bibliometrics, 2) data sharing, and 3) data curation.

3.1 Bibliometrics

The study of scholarly communication takes place at the intersection of sociology of knowledge, history and sociology of science, and LIS. Within the latter, the most developed and important subfields are the study of publication patterns within different fields of science, and bibliometrics, the statistical study of citation patterns and scientific output. As both political norms and technical developments have influenced scientific practices, new problems have arisen with demands for sometimes imminent solutions. Recent developments have meant that increasingly Big Data sets have influenced both research in bibliometrics, and bibliometrics as a research assessment tool [Moed 2012].

As networked research environments have emerged, the field has developed and is now divided into closely related subfields such as webometrics, informetrics and altmetrics, analysing statistical patterns between digital documents and sets of data (see, for example, [Thelwall 2008] for a discussion on these developments). Bibliometrics as a scientific field appeared in 1960s with the birth of E. Garfield's Science Citation Index, a database of scientific references across articles and authors. Webometrics uses similar principles of citation analysis, word frequency analysis and co-word analysis from bibliometrics, but is applied to the web, especially at the level at which the web can be considered to be a big database of academic documents. Of special interest here is altmetrics which has developed from a sense of insufficiency with the traditional bibliographical basis for statistical analytics, references and citations. In a networked environment, where the boundaries between publications and open data are blurred by digitally-based availability through, for instance, social media and blogs, providing what [Priem 2014, p. 263] describes as 'hidden impact', the need for new mapping methods and activity analysis has proven necessary.

The practical challenges facing the research community will increase as the amount and diversity of data gets bigger. In [Borgman 2007, p. 6] it is stated: "data and information have always been both input and output of research. What is new is the scale of data and information involved. Information management is notoriously subject to problems of scale". From an LIS perspective, primarily two problems stand out as specifically interesting: the problem of data sharing and the problem of managing and organising the data.

3.2 Data Sharing

Data sharing is not a new field of study: the influence of growing amounts of networked data has been a part of the discussion on scholarly communication since at least the mid 1960s ([De Solla Price 1963], [Meadows, O'Connor 1971]. At this time the significance of data was embedded in prioritised problems within LIS to develop scientific methods for organizing the network of scholarly communication which is displayed in the references of scientific publications. The interest for data as such has then emerged gradually from the bibliometric exclusive focus on publications. A more explicit interest for data in various sizes and velocities has been prevalent since the mid 1980's when the first major anthology on data sharing appeared in the form of a US government report on the issue [Fienberg, Martin, Straf 1985].

The increasingly less apparent differences between data and publications as analytical entities in the networked scholarly environment of today have led to questions concerning *data*; same questions that have been traditionally designated to the curation and measurement of *publication* dissemination, citation patterns, scientific impact and research evaluation: "the widespread availability of electronic data not only made network analysis of large datasets a possibility, but also made network analysis a cornerstone of bibliometric research" [West, Vilhena 2014, p. 152].

Arriving at a networked 'Big Data' situation in more contexts than traditional Big Data disciplines such as astronomy, genomics and particle physics, [Borgman 2012] analyses the reasons for, and problems arising from, sharing (or not) scientific data. She proposes four basic rationales for sharing data: 1) to reproduce and verify research, 2) to make results of publically funded research available for the public, 3) to enable others to ask new questions of extant data, and 4) to advance the state of research and innovation. Some of these, perhaps in a most direct sense 1-3, may seem somewhat intuitive, but the technical and political challenges in fulfilling these rationales and making them work are clouded by several issues ranging from the various kinds of data that are produced in the sciences, the social sciences and in the humanities.

Not only do data in different disciplines and research fields lend themselves to sharing differently, but there are in many situations ethical considerations to be made which are complicated through the ability of massing huge amounts of open data in, for instance, the social sciences (see, for example, [Metcalf, Crawford 2016]). Another problem which is seldom taken into consideration is not a technical one, but the simple fact that researchers often have proven to be reluctant to share their data. This hesitation does not necessarily come from an overly possessive relation to gathered data, but from the fact that it is often hard to distinguish in what state data is "sharable" (see, e.g., [Lassi, Johnsson, Golub 2016]). This too seems to relate more to disciplines new to the amount of data in digital scholarship, as is often the case within the humanities and social sciences, rather than in the traditional Big Data disciplines, where numerous researchers have instant access to data as it is produced, for example within astronomy.

The fourth rationale defined by Borgman is interesting in itself, as it represents what is sometimes referred to as the fourth paradigm or the fourth (industrial) revolution ([Floridi 2014], [Frederick, 2016]). The fourth paradigm or revolution is characterised by the convergence of numerous technologies and various developing sectors such as universities and industry. It is indirectly promoted through global development of entrepreneurial ideology of universities, creating new meaning of the demands of usability of research results. Examples of technologies and technology-related phenomena which are promoted through the thought of a new paradigm, apart from Big Data include: 3D printing, cloud computing, biotechnology, bringing together the human body and 'intelligent' technology. From an LIS point of view, this creates challenges in data curation, as data in many cases can no longer can be distinguished from either application or publications.

3.3 Data Curation

Data in LIS refers to data of all sizes, including small and big [De Giammarco, 2013], as well as different types of data, for example research data, governmental data etc. Considering that Digital Humanities (DH) is closely related to LIS especially when it comes to increasing importance of data-intensive research in both LIS and DH [Koltay 2016, p. 781], an example is given of Big Data in DH. Big Data in DH refer to large or dense cultural datasets, meaning that they are not necessarily big in the number of terabytes but in the fact that manual study becomes cumbersome and new methods are needed [Kaplan, 2015, p. 1-2]. Another important characteristic is that they are interconnected. In [Teets, Goldner 2013] an example is given of how OCLC's WorldCat catalog of global library holdings has started work on the knowledge graph in order to make their data useful and used throughout the Web. Other examples include large corpora of digitised books, millions of photographs and micro-messages shared in Web 2.0 services, big geographic information systems like Google Earth, and linked clouds of academic papers citing each other (ibid., p. 2-3). All these data sets may need to be collected, organized, preserved, and made retrievable, for future use by others.

Over the recent years many academic libraries have taken up the role of providing various services to support research data management, which has been considerably affected by data sharing developments discussed above (e.g., Purdue University Libraries, University of Michigan Library [Carlsson, Brandt, 2014]). Examples of Big Data curation and storage in libraries include map and GIS (Geographic Information Systems) data libraries which are facing dramatic changes resulting from cloud computing and big data [Goldberg et al. 2014]. Curation of non-research data also generally encompasses similar aspects as the ones outlined for research data below and these are therefore more or less applicable in other contexts as well.

The management of research data involves addressing a range of highly challenging issues considering the diversity of data types and disciplinary cultures which entail different data collecting, processing, managing and storing practices in researchers' everyday work. The nature of different data sets in terms of form, format, size and embodied differences affect data sharing, archiving, discovery, organization and description, linking, interoperability, and impact in services provided by academic libraries or related infrastructures.

As an example, we take the knowledge organization aspects. Generally, knowledge organization in LIS addresses issues of organizing data, information, resources and knowledge, for the purpose of retrieval, whereby topics include, but are not limited to, Knowledge Organization Systems (KOS) like ontologies, classification systems, thesauri, subject headings; semi-automated indexing and classification; social tagging; Semantic Web and Linked Data. Subjects (keywords, topics) are important as subject searching (searching by topic or theme) is one of the most common and at the same time the most challenging type of searching in information systems, due to the ambiguities of the natural language. Subject index terms taken from standardized KOS (e.g., thesauri, classification schemes) provide numerous benefits compared to free-text indexing: consistency through uniformity in term format and the assignment of terms, provision of semantic relationships among terms, support of browsing by provision of consistent and clear hierarchies (for a detailed overview, see, for example, [Lancaster 2003]).

Standardized metadata schemes are crucial in naming and organizing data and relationships among them and have “profound effects on the ability to discover, exchange, and curate data” [Borgman 2015, p. 65]. Particularly challenging is bringing together different metadata schemes used to describe data sets coming from different communities. This is even harder when it comes to mapping subjects (or topics, keywords) between the different disciplines, as each discipline often uses a different knowledge organization system, while some may not be using any at all [Golub, Johansson, Lassi 2016]. Still, challenges like mapping across the KOS need to be addressed in order to meet the established objectives of quality controlled information retrieval systems like those provided by libraries.

Apart from librarian KOS-based indexing, social tagging, author tagging and semi-automated subject indexing approaches like text categorization, document clustering and string-to-string matching [Golub 2006] may be used to more comprehensively describe resources. Visualisation would best be applied for browsing large collections, as well as large collections data sets and other types of resources (see, for example, [Whitelaw, 2009]).

4 Concluding Remarks

In their overview of dilemmas and opportunities for Big Data research, [Ekbia et al. 2015] conclude that Big Data is of such concern for the general public in their daily lives that opportunities are promising. This is in spite of the fact that challenges can be found within 1) the technology itself, 2) methodology reaching beyond traditional boundaries (e.g., the qualitative/quantitative dichotomy), 3) aesthetics because visualizations and model making have been taken on a whole new level; all these leading to questions concerning epistemology yet unsolved.

It does however require joint efforts by political actors and academia to point to the relevance of Big Data and to make the processes visible to ordinary people. Ethical and legal issues are therefore also pointed out as areas necessary to address. These necessities are also emphasized in [Day 2014] in relation specifically to social Big Data which are of such significance for so many people. He particularly underlines the challenges facing research as social Big Data relates to a neo-liberal logic rather than a traditionally epistemological one. Frické [Frické 2015] builds a related argument in a more pessimistic direction claiming that Big Data presents major challenges to research in that it opposes traditional epistemology through its elusive illusory character – increased knowledge does not necessarily follow from increased amount of data. Instead he finishes his argument by stating that "if anything, science needs more theories and less data" (ibid., p. 660).

Big Data will, however, continue to grow even bigger. New technologies and networked research environments will increase both in numbers and size and, as proposed by the fourth revolution proponents, integration between physical worlds and computer worlds will become more complex and in many aspects the boundaries between academia and industry will become increasingly blurred. In this environment it will become necessary to find new ways of organising documentation, information and data and to develop instruments for research evaluation also in situations where the differences between publications and data will decrease in significance and clarity. Big Data research challenges our view of what research is and LIS is rapidly developing tools to meet the opportunities arising – through educational initiatives and the development of new research areas such as (big) data curation and altmetrics. As social and political demands for open data grow, these issues are pressing.

References

[Abbas 2010] Abbas, J.: "Structures for Organizing Knowledge: Exploring taxonomies, Ontologies, and Other Schemas"; Neal-Schuman / New York (2010).

[Borgman 2007] Borgman, C.: "Scholarship in the Digital Age: Information, Infrastructure and the Internet"; MIT Press / Cambridge, MT (2007).

[Borgman 2012] Borgman, C.: "The Conundrum of Sharing Research Data"; JASIST (Journal of the American Society for Information Science and Technology), 63, 6 (2012), 1059-1078.

[Borgman 2015] Borgman, C.: "Big Data, Little Data, No Data: Scholarship in the Networked World"; MIT Press / Cambridge, MT (2015).

[Buckland 1991] Buckland, M.: "Information and Information Systems"; Praeger / Westport, CT (1991).

[Buckland 2012] Buckland, M.: "What Kind of Science *Can* Information Science Be?"; JASIST (Journal of the American Society for Information Science and Technology), 63, 1 (2012), 1-7.

[Carlson, Brandt 2014] Carlson, J., Brandt, S.D.: "Data Curation Profiles" (2014); available at: http://datacurationprofiles.org/.

[Chowdhury, Chowdhury 2007] Chowdhury, G.G., Chowdhury, S.: "Organizing Information: from the Shelf to the Web"; Facet / London (2007).

[Day 2001] Day, R.: "The Modern Invention of Information: Discourse, History and Power"; Southern Illinois Univ. Press / Carbondale (2001).

[Day 2006] Day, R.: "'A Necessity of Our Time': Documentation as 'Cultural Technique'"; What Is Documentation?: English Translation of the Classic French Text; Scarecrow / Lanham, MD, 47-63 (2006).

[Day 2014] Day, R.: “Indexing It All: The Subject in the Age of Documentation, Information, and Data”; MIT Press / Cambridge, MT (2014).

[De Mauro, Greco, Grimaldi 2016] De Mauro, A., Greco, M., Grimaldi, M.: “A Formal Definition of Big Data Based on Its Essential Features”; Library Review, 65, 3 (2016), 122-135.

[De Solla Price 1963] De Solla Price, D.: ”Little Science, Big Science”; Columbia Univ. Press / New York (1963).

[Di Giammarco 2013] Di Giammarco, F.: “Libraries and Big Data”; Biblioteche Oggi, 31, 10 (2013), 7-8.

[Ekbia et al. 2015] Ekbia, H., Mattioli, M., Kouper, I., Arave, G., Ghazinejad, A., Bowman, T., Ratandeep Suri, V., Tsou, A., Weingart, S., Sugimoto, C. R.: “Big Data, Bigger Dilemmas: A Critical Review”; JASIST (Journal of the Association for Information Science and Technology), 66, 8 (2015), 1523-1545.

[Farmer, Safer, Alan 2016] Farmer, L., Safer, S. J., Alan, M. “Library Improvement through Data Analytics”; Facet / London (2016).

[Fienberg, Martin, Straf 1985] Fienberg, S. E., Martin, M. E., Straf, M. L.: ”Sharing Research Data”; National Academic Press / Washington, DC (1985).

[Floridi 2014] Floridi, L.: “The 4th Revolution: How the Infosphere is Reshaping Human Reality”; Oxford Univ. Press / Oxford (2014).

[Frederick 2016] Frederick, D.: “Libraries, Data and the Fourth Industrial Revolution”; Library High Tech News, 5 (2016), 9-12.

[Frické 2015] Frické, M.: “Big Data and Its Epistemology”; JASIST (Journal of the Association for Information Science and Technology”; 66, 4 (2015), 651-661.

[Frohmann 2004] Frohmann, B.: “Deflating Information: From Science Studies to Documentation”; Univ. Of Toronto Press / Toronto (2004).

[Furner 2016] Furner, J.: “‘Data’: The Data”; Information Cultures In The Digital Age: A Festschrift in Honor of Raphael Capurro; Springer / Wiesbaden (2016), 287-306.

[Goldberg et al. 2014] Goldberg, D. et al.: “Maps & GIS Data Libraries in the Era of Big Data and Cloud Computing”; Journal of Map & Geography Libraries: Advances In Geospatial Information, Collections & Archives, 10, 1 (2014), 100-122.

[Golub 2006] Golub, K.: “Automated Subject Classification of Textual Web Documents”; Journal of Documentation, 62, 3 (2006), 350-371.

[Golub 2014] Golub, K.: “Subject Access to Information: An Interdisciplinary Approach”; Libraries Unlimited / Santa Barbara, CA (2014).

[Hansson 2004] Hansson, J.: “The Social Legitimacy of Library and Information Studies”; Aware and Responsible: Papers of the Nordic-International Colloquium on Social and Cultural Awareness and Responsibility in Library, Information and Documentation Studies (SCARLID); Scarecrow Press / Lanham, MD, 2004, 49-70.

[Ingwersen, Järvelin 2006] Ingwersen, P., Järvelin, K.: ”The Turn: Integration of Information Seeking and Retrieval in Context”; Springer / Dordrecht (2006).

[Kuhn 1996] Kuhn, T. S.: “The Structure of Scientific Revolutions”; University Of Chicago Press / Chicago, Ill (1996).

[Lancaster 2003] Lancaster, F.W.: "Indexing and Abstracting in Theory and Practice"; Facet: London (2003).

[Lassi, Johnsson, Golub 2016] Lassi, M., Johnsson, M., Golub, K.: " Research Data Services: An Exploration of Requirements at Two Swedish Universities"; IFLA Journal, 42, 4 (2016), 266-277.

[Kaplan 2015] Kaplan, F.: "A Map for Big Data Research in Digital Humanities"; Frontiers in Digital Humanities, 2, 1 (2015); available at https://www.frontiersin.org/Articles/10.3389/Fdigh.2015.00001/Full

[Koltay 2016], Koltay, T.: "Library and Information Science and the Digital Humanities"; Journal of Documentation, 72, 4 (2016), 781-792.

[Markey 2015] Markey, K.: "Online Searching: A Guide to Finding Quality Information Efficiently and Effectively"; Rowman & Littlefield / Lanham (2015).

[Meadows, O'Connor 1971] Meadows, A. J., O'Connor, J. G.: "Bibliographical Statistics as a Guide to Growth Points in Science"; Social Studies of Science; 1, 1 (1971), 95-99.

[Metcalf, Crawford 2016] Metcalf, J., Crawford, K.: "Where Are Human Subjects in Big Data Research? The Emerging Ethics Divide"; Big Data & Society, January–June 2016, 1–14.

[Miksa 1992] Miksa, Francis L.: "Library and Information Science: Two Paradigms"; Conceptions of Library And Information Science: Historical, Empirical and Theoretical Perspectives, 229-252.

[Moed 2012] Moed, H.: "The Use of Big Datasets in Bibliometric Research"; Research Trends, 30 (2012), available at: Https://Www.Researchtrends.Com/Issue-30-September-2012/The-Use-Of-Big-Datasets-In-Bibliometric-Research/

[Nolin, Åström 2010] Nolin, J., Åström, F.: "Turning Weakness into Strength: Strategies for Future LIS"; Journal of Documentation, 66, 1(2010), 7-27.

[Oliver, Harvey 2016] Oliver, G., Harvey, R.: "Digital Curation"; Facet: London (2016).

[Pettigrew, Fidel, Bruce 2001] Pettigrew, K. E., Fidel, R., Bruce, H.: "Conceptual Frameworks in Information Behavior"; ARIST (Annual Review of Information Science and Technology), 35 (2001), 43-78.

[Priem 2014] Priem, J.: "Altmetrics"; "Beyond Bibliometrics: Harnessing Multidimensional Indicators of Scholarly Impact"; MIT Press / Cambridge, MT (2014), 263-287.

[Rowley 2007] Rowley, J.: "The Wisdom Hierarchy: Representations of the DIKW Hierarchy"; Journal of Information Science, 33, 2(2007), 163-180.

[Saracevic 2009] Saracevic, T.: "Information Science"; Encyclopedia of Library and Information Science; Taylor & Francis / New York (2009), 2570-2586.

[Svenonius 2000] Svenonius, E. "The Intellectual Foundation of Information Organization"; MIT Press / Cambridge, Mass. (2000).

[Teets, Goldner 2013] Teets, M., Goldner, M.: "Libraries' Role in Curating and Exposing Big Data"; Future Internet, 5 (2013), 429-438., available at: http://www.Mdpi.Com/1999-5903/5/3/429

[Thelwall 2008] Thelwall, M.: "Bibliometrics to Webometrics"; Journal of Information Science 34, 4 (2008), 605-21.

[West, Vilhena 2014] West, J. D., Vilhena, D. A. “A Network Approach to Scholarly Evaluation”; Beyond Bibliometrics: Harnessing Multidimensional Indicators of Scholarly Impact; MIT Press / Cambridge, MT, 151-165.

[Whitelaw 2009] Whitelaw, M. “The Visible Archive”, available at http://Mtchl.Net/The-Visible-Archive/

[Zeng, Qin 2016] Zeng, M. L., Qin, J. “Metadata”; ALA: Chicago (2016).