

# Big data, big literacies?

Tibor Koltay  
Szent István University, Jászberény, Hungary  
Faculty of Applied Arts and Pedagogy  
Department of Information and Library Studies

## Abstract

Digital research data is a hot topic today, which raises the need for a relatively new type of literacy, i.e. data literacy. After setting the scene, the role of data literacy is discussed in the light of data sharing, data quality, data management, data curation, and data citation, the majority of which are related to the activities of libraries and information professionals. Potentially relevant literacies are discussed in the light of these issues. Besides information literacy, the characteristics of digital literacy, scientific literacy and academic literacy are reviewed. Their relationship to data literacy, a concept in forming, is outlined.

## Keywords:

Data literacy, information literacy, data management, digital data research, data sharing, data quality, data curation, data citation, digital literacy, scientific literacy, academic literacy

## Prelude

The somewhat nonsensical question asked in the title leads first to a negation: While there is big data, "big literacy" does not exist. Notwithstanding, literacies are "big" in the sense that there are many literacies.

## Introduction

Big data is only one, though important facet of the data-rich world that has been around for a time, while it is still emerging. It exceeds the processing capacity of conventional database systems in capturing, storing, managing, and analyzing<sup>1</sup>. Nonetheless, the decisive factor is not the quantity of data, but the capacity to search, aggregate, and cross-reference large data sets by virtue of the processing power of computers and networks. Big data is no longer an exclusive issue for the natural sciences, but it is present in the social sciences, the humanities, arts and culture, as well. There is an aura of truth, objectivity, and accuracy around it, coupled with the belief that the existence of big data offers

a higher form of intelligence and knowledge<sup>2</sup>. While hopes and fears, related to this phenomenon may be exaggerated, the capacity to store massive amounts of data and other components of a highly developed information and communication technology infrastructure are beginning to bring with them changes to the nature and practice of research and scholarship<sup>3</sup>. The importance of data goes beyond big data issues. Research data has to be managed in the broad sense of the word, independently of the fact that we call it big or not. To make data accessible, we need to develop appropriate technical and organizational infrastructures for storage and retrieval. Incentives and policies for researchers to share data are also indispensable<sup>4</sup> and data literacy, which carries the potential of motivation and allows findability of data, is one of the essential elements of this infrastructure.

<sup>1</sup> Laura Gordon-Murnane, "Big Data: A big opportunity for librarians", *Online* (September – October 2012): 31.

<sup>2</sup> Dana D. Boyd and Kate Crawford, "Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon", *Information, Communication & Society* 15 (2012): 663.

<sup>3</sup> Jacob Carlson et al., "Determining data information literacy needs: A study of students and research faculty", *Portal: Libraries and the Academy* 11 (2011): 629.

<sup>4</sup> Stacy Kowalczyk and Kalpana Shankar, "Data sharing in the sciences", *Annual Review of Information Science and Technology* 45 (2011): 248.

## Data literacy defined

We find somewhat differing terms in the literature. Besides data literacy there is *data information literacy*,<sup>5</sup> *science data literacy*<sup>6</sup> and *research data literacy*<sup>7</sup>. Taking this variety, it seems to be expedient to define data literacy by ourselves as a set of skills and abilities related to accessing research data, critically evaluating and using it. This definition reflects the opinion of Qin and D'Ignazio, according to whom science data literacy (SDL) emphasizes the ability to understand, use, and manage science data<sup>8</sup>. It also subsumes the reasoning of the Association of College and Research Libraries, which focuses on understanding how to find and evaluate data, giving emphasis to the version of the given dataset, the person responsible for it, and does not neglect the questions of citing and ethical use of data<sup>9</sup>.

Data literacy accommodates not only the data consumer's viewpoint, but the data producer's, as well<sup>10</sup>. This is in accordance with novel views on information literacy, represented for example by the concept of Information Literacy 2.0<sup>11</sup>.

From the fields, where data literacy competencies are used, following ones have to be mentioned:

- Discovery and acquisition of data;
- Data management;
- Data conversion and interoperability (dealing with the risks and potential loss or corruption of information caused by changing data formats);
- Quality assurance;
- Metadata;
- Data curation and re-use;
- Data preservation;
- Data analysis;

- Data visualization;
- Ethics, including citation of data<sup>12</sup>.

## Data literacy and the literacies

Data literacy, we agree, can be treated as a new sub-discipline of information literacy<sup>13</sup>. With this we acknowledge that the views on the relationship between data and information can be different, depending on the approaches used. The two concepts can be regarded as identical, as different or as not having direct relationship with each other.<sup>14</sup>

It fits well into a framework of future work skills of any worker, which provides a wide context to information literacy. One of these work skills is the ability to translate vast amounts of data into abstract concepts, as well as to understand data-based reasoning<sup>15</sup>.

Data literacy also shares some distinguishing features with media literacy, especially in regard to the use and reuse of content in ways not imagined by the content creator<sup>16</sup>. Mentioning media literacy has two reasons. First, we know that there is a convergence among literacies caused by the convergence between different forms of media and information and communication technologies<sup>17</sup>. From a slightly different angle we can state that information literacy is the best contender to draw together the other literacy movements into a single emphasis<sup>18</sup>. Second, we are aware of the fact that the capacity and interest in data-related issues is to a substantial extent result of the appearance of the web 2.0. In this environment, users and their interests are represented in mediated spaces, which also serve as an environment to activate engagement with others<sup>19</sup>.

<sup>5</sup> Carlson et al., 629.

<sup>6</sup> Jian Qin and John D'Ignazio, "Lessons learned from a two-year experience in science data literacy education", in *Proceedings of the 31st Annual IATUL Conference* (June, 20-24 2010), <http://docs.lib.purdue.edu/iatul2010/conf/day2/5> (Date of access 17. 1. 2014).

<sup>7</sup> René Schneider, "Research Data Literacy", in *Worldwide Commonalities and Challenges in Information Literacy Research and Practice*, S. Kurbanoglu et al., eds. (Springer International Publishing, 2013), 135.

<sup>8</sup> Qin and D'Ignazio.

<sup>9</sup> Association of College and Research Libraries, *Intersections of scholarly communication and information literacy: Creating strategic collaborations for a changing academic environment* (Chicago, IL: ACRL 2013), 10.

<sup>10</sup> Carlson et al., 633.

<sup>11</sup> Sonja Špiranec and Michaela Banek Zorica, "Information Literacy 2.0: hype or discourse refinement?", *Journal of Documentation* 66 (2010): 140–153.

<sup>12</sup> Carlson et al., 652–653.

<sup>13</sup> René Schneider, 135.

<sup>14</sup> Lai Ma, "Is information still relevant?", *Information Research* 18 (2013), <http://InformationR.net/ir/18-3/colis/paperC33.html> (Date of access 17. 1. 2014).

<sup>15</sup> Anna Davies et al., *Future work skills, 2020*. (Palo Alto, CA: Institute for the Future, 2011), 10, <http://www.iftf.org/our-work/global-landscape/work/future-work-skills-2020/> (Date of access 17. 1. 2014).

<sup>16</sup> Association of College and Research Libraries (2013), 11.

<sup>17</sup> Sonja Livingstone et al., "Converging traditions of research on media and information literacies: Disciplinary and methodological issues", in *Handbook of Research on New Literacies*, edited by D. J. Leu et al. (Hillsdale: Lawrence Erlbaum Associates, 2008), 106.

<sup>18</sup> William Badke, "Media, ICT, and Information Literacy", *Online* 33 (2009): 49.

<sup>19</sup> Kylie Jarrett, "Interactivity is Evil! A critical investigation of Web 2.0", *First Monday* 13 (2008), <http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/viewArticle/2140/1947> (Date of access 17. 1. 2014).

This mediated nature influences data indirectly and to a lesser extent. Nonetheless, it requires that we approach data literacy not only as a manifestation of information literacy, but look at it through the lens of media literacy.

Adapting the ACRL information literacy competency standards<sup>20</sup> to data information literacy, examined by Carlson et al. offers an interesting perspective<sup>21</sup>. For example, this comparison shows that the cultural practices and norms of the data producer's discipline influence not only potential needs and uses of the data being generated, but effective sharing, re-using, curating, and preserving of data.

It is worth to put a strong emphasis on *Standard Three* that directs attention to evaluating information critically. If we translate this universal principle to data, we see that data information literate students are required not only to determine when and how to share data, but also document their own sources of data. This acquires significance in the light of the difficulty of documenting data, a fact that appeared in our writing several times. The reason for this is fairly simple. Critical examination of some information or data is practically impossible if we cannot rely on its documented feature, e. g. we cannot verify its provenance.

Differences between different (other) types of literacy are outlined by Qin and D'Ignazio as follows. Information literacy mobilizes the abilities and skills, related to finding, retrieving, analyzing, and using information. The same object, i.e. information, is aimed by digital literacy, which also accentuates creating it, while emphasizing the use of digital technology<sup>22</sup>. Without wanting to go into a detailed discussion on the differences and similarities between information literacy and digital literacy (see e.g.<sup>23</sup>), we can stress that digital literacy shares a high degree of interest in production with data literacy.

The relationship between data literacy and academic literacy has many faces. The latter has a grammatical dimension that information literate students must be taught and researchers must master.

It involves the comprehension of the entire system of thinking, values, cultural identity and information flows of academia, which results in the ability to read, interpret, and produce texts valued in academia<sup>24</sup>. Data literacy, affects the undergraduate student population to a lesser degree than academic literacy. Instead, data literacy has a true lifelong-learning character. The reason for this is in the fact that researchers in the past did not have to possess many of the data-related skills, at least not the ones related to data sharing and especially to citing data. In addition to this, the majority of the researchers have to acquire data literacy in the workplace, up to the point, when it becomes a standard literacy in graduate and doctoral education.

Data literacy is also connected to scientific literacy, which comprises methods, approaches, attitudes and skills, related to thinking scientifically and doing scientific research<sup>25</sup>. This implies that everybody should be scientifically literate, even if only a small number of graduate students become scientists.

Let us add that – with the similarities and differences between them – these literacies complement each other, while academic literacy is more closely associated with formal learning, especially in higher education. It seems to be clear, as well, that they are not valid without information literacy, at least not in research environments.

All the above fit well into the framework of *metaliteracy*, which is overarching by providing the foundation for media literacy, digital literacy and other literacies, and emphasizes *content*<sup>26</sup>, so the separation between information in *information literacy* from data in *data literacy* is less conspicuous. In this context, the border between the two concepts is not rigid, as information literacy has always been interested in the proper understanding and use of data that is converted into information<sup>27</sup>.

<sup>20</sup> Association of College and Research Libraries, *Information Literacy Competency Standards for Higher Education* (Chicago, IL: ACRL, 2000).

<sup>21</sup> Carlson et al., 649–653.

<sup>22</sup> Qin and D'Ignazio.

<sup>23</sup> David Bawden, "Origins and concepts of digital literacy", in *Digital Literacies: concepts, policies and practices*, edited by C. Lankshear and M. Knobel (New York, NY: Peter Lang, 2008), 17–32.

<sup>24</sup> James Elmborg, "Critical information literacy: Implications for instructional practice", *Journal of Academic Librarianship* 32 (2006): 196.

<sup>25</sup> National Academy of Sciences, *National Science Education Standards* (Washington D.C.: National Academy Press, 1996), <http://www.nap.edu/readingroom/books/nse> (Date of access 17. 1. 2014).

<sup>26</sup> Thomas Mackey and Trudy Jacobson, "Reframing Information Literacy as a Metaliteracy", *College & Research Libraries* 72 (2011): 62–78.

<sup>27</sup> Schneider, 135.

## Data literacy skills

Data literacy education is not only a relatively new field for libraries, but it is one that cuts across disciplinary boundaries and across the traditional structures of academic library organizations. One sign of this is that humanities and social science scholars became a new constituency for data literacy education<sup>28</sup>.

Data literacy education can have a dual purpose. One is rather self-explanatory, i.e. to achieve that students and researchers become data literate science workers. The other one is to educate data management professionals<sup>29</sup>.

Data literate graduate students and researchers have to be familiar with some of the questions that data curation poses:

- Who owns the data?
- What requirements are imposed by others (e.g. funding agencies or publishers)?
- Which data should be retained?
- For how long should data be maintained?
- How should digital data be preserved?
- Are there ethical considerations?
- What sort of risk management is needed for research data?
- How are data accessed?
- How open should the data be?
- What alternatives to local data management exist?<sup>30</sup>

The need for reviewing data periodically is also worth of attention as it forms the basis of responsible decisions about disposal of "unnecessary" data, and preventing hardware or software obsolescence<sup>31</sup>.

A data literate researcher does not need to be a data curator. However, being familiar with the competencies of the curators and the fields, where the competencies, listed below, are of good use helps in developing data literacy:

- The ability to collaborate and work in teams;
- Familiarity with scientific data sources;
- Familiarity with quantitative research methods;

- Knowledge of general metadata standards<sup>32</sup>.
- From the fields, where these and other competencies can be used the following ones can be singled out:
  - The data structure of different digital objects;
  - The ways to assess the digital objects' authenticity, integrity and accuracy over time;
  - Storage and preservation policies, procedures and practices;
  - Relevant quality assurance standards;
  - The risks of information loss or corruption of digital entities<sup>33</sup>.

Data literacy education should not forget about data citation, which allows the identification, retrieval, replication, and verification of data underlying published studies. Standardized forms of data citation could provide a motivation for researchers to share and publish their data, thus it has the potential to become a tool of reward and acknowledgment for them<sup>34</sup>. Unfortunately, despite numerous initiatives, at the moment, there is no standardization and consistency in data citation.

The quality of data is a driving force of data literacy, as well. It is determined by multiple factors. The first one is trust that depends on subjective judgments on authenticity, acceptability or applicability of the data. Trust is also influenced by the given subject discipline, the reputation of those responsible for the creation of the data, and the biases of the persons who are evaluating the data. The next factor is authenticity, which measures the extent to which the data is judged to represent the proper ways of conducting scientific research, including the reliability of the instruments used to gather the data, the soundness of underlying theoretical frameworks, the completeness, accuracy, and validity of the data. In order to evaluate authenticity, the data must be understandable. The condition for the evaluation of understandability of data is the presence of sufficient context in the form of

<sup>28</sup> Association of College and Research Libraries (2013), 11.

<sup>29</sup> Qin and D'Ignazio.

<sup>30</sup> Rick Erway, "Starting the conversation: University-wide research data management policy", *Educause Review Online* (December 6, 2013), <http://www.educause.edu/ero/article/starting-conversation-university-wide-research-data-management-policy> (Date of access 17. 1. 2014).

<sup>31</sup> G. Pryor, "Why manage research data?", in *Managing Research Data*, G. Pryor, ed. (London: Facet, 2012), 8.

<sup>32</sup> Li Si, et al., "The cultivation of scientific data specialists: Development of LIS education oriented to e-science service requirements", *Library Hi Tech* 31 (4): 705.

<sup>33</sup> Melody M. Madrid, "A study of digital curator competences: A survey of experts", *International Information & Library Review* 45 (2013): 154–155.

<sup>34</sup> Hailey Mooney and Mark P. Newton, "The anatomy of a data citation: Discovery, reuse, and credit", *Journal of Librarianship and Scholarly Communication* 1 (2012): 1.

documentation and metadata, and it requires the data to be usable. To be aware of data provenance is also crucial. If we want data to be usable, it has to be discoverable and accessible; and be in a usable file format. The individuals judging data quality need to have at their disposal an appropriate tool to access the data; which has to show sufficient integrity to be rendered. Integrity of data assumes that the data can be proven to be identical, at the bit level, to some previously accepted or verified state. Data integrity is required for usability, understandability and authenticity, thus it influences overall quality<sup>35</sup>.

<sup>35</sup> Michael Giarlo, "Academic libraries as quality hubs", *Journal of Librarianship and Scholarly Communication* 1 (2013): 4–5.

## End note

Those, who will use data, will need education about how to find it, how to understand, interpret, and apply what they find. So, educating students alongside with future and actual scholars to data literacy is a new challenge to libraries, which cuts across disciplinary boundaries.

Information professionals thus need to be involved in developing ways to handle key issues of data literacy as they are precisely in the intersection of scholarly communication and information literacy<sup>36</sup>.

<sup>36</sup> Association of College and Research Libraries (2013), 11.

## Literature:

1. Association of College and Research Libraries. *Information Literacy Competency Standards for Higher Education*. Chicago, IL: ACRL, 2000.
2. Association of College and Research Libraries. *Intersections of scholarly communication and information literacy: Creating strategic collaborations for a changing academic environment*. Chicago, IL: ACRL, 2013.
3. Badke, William. "Media, ICT, and Information Literacy". *Online* 33 (2009): 47–49.
4. Bawden, David. "Origins and concepts of digital literacy". In *Digital Literacies: concepts, policies and practices*. Edited by C. Lankshear and M. Knobel, 17–32. New York, NY: Peter Lang, 2008.
5. Boyd, Dana D. and Kate Crawford. "Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon". *Information, Communication & Society* 15 (2012): 662–679.
6. Carlson, Jacob et al. "Determining data information literacy needs: A study of students and research faculty". *Portal: Libraries and the Academy* 11 (2011): 629–657.
7. Davies, Anna, Devin Fidler and Marina Gorbis. *Future work skills, 2020*. Palo Alto, CA: Institute for the Future, 2011. <http://www.iftf.org/our-work/global-landscape/work/future-work-skills-2020/> (Date of access 17. 1. 2014).
8. Elmborg, James. "Critical information literacy: Implications for instructional practice". *Journal of Academic Librarianship* 32 (2006): 192–199.
9. Erway, Rick. "Starting the conversation: University-wide research data management policy". *Educause Review Online* (December 6, 2013). <http://www.educause.edu/ero/article/starting-conversation-university-wide-research-data-management-policy> (Date of access 17. 1. 2014).
10. Giarlo, Michael. "Academic libraries as quality hubs". *Journal of Librarianship and Scholarly Communication* 1 (2013): 1–10.
11. Gordon-Murnane, Laura. "Big Data: A big opportunity for librarians". *Online* (September – October 2012): 30–34.
12. Jarrett, Kylie. "Interactivity is Evil! A critical investigation of Web 2.0". *First Monday*, 13 (2008). <http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/viewArticle/2140/1947> (Date of access 17. 1. 2014).
13. Kowalczyk, Stacy and Kalpana Shankar. "Data sharing in the sciences". *Annual Review of Information Science and Technology* 45 (2011): 247–294.

Koltay T. "Big data, big literacies", 3–8

14. Livingstone, Sonja, Elizabeth van Couvering and Nancy Thumin. "Converging traditions of research on media and information literacies: Disciplinary and methodological issues". In *Handbook of Research on New Literacies*. Edited by D. J. Leu et al., 103–132. Hillsdale: Lawrence Erlbaum Associates, 2008.
15. Ma, Lai. "Is information still relevant?". *Information Research* 18 (2013). <http://InformationR.net/ir/18-3/colis/paperC33.html> (Date of access 17. 1. 2014).
16. Mackey, Thomas and Trudy Jacobson. "Reframing Information Literacy as a Metaliteracy". *College & Research Libraries* 72 (2011): 62–78.
17. Madrid, Melody M. "A study of digital curator competences: A survey of experts". *International Information & Library Review* 45 (2013): 149–156.
18. Mooney, Hailey and Mark P. Newton. "The anatomy of a data citation: Discovery, reuse, and credit". *Journal of Librarianship and Scholarly Communication* 1 (2012): 1–14.
19. National Academy of Sciences. *National Science Education Standards*. Washington D.C.: National Academy Press, 1996. <http://www.nap.edu/readingroom/books/nses> (Date of access 17. 1. 2014).
20. Pryor, G. "Why manage research data?". In *Managing Research Data*. G. Pryor, ed., 1-16. London: Facet, 2012.
21. Qin, Jian and John D'Ignazio. "Lessons learned from a two-year experience in science data literacy education". In *Proceedings of the 31st Annual IATUL Conference* (June, 20-24 2010). <http://docs.lib.purdue.edu/iatul2010/conf/day2/5> (Date of access 17. 1. 2014).
22. Schneider, René. "Research Data Literacy". In *Worldwide Commonalities and Challenges in Information Literacy Research and Practice*. 134–140. Springer International Publishing, 2013.
23. Si, Li et al., "The cultivation of scientific data specialists: Development of LIS education oriented to e-science service requirements". *Library Hi Tech* 31 (4): 700–724.
24. Špiranec, Sonja and Mihaela Banek Zorica. "Information Literacy 2.0: hype or discourse refinement?". *Journal of Documentation* 66 (2010): 140–153.

## Велики подаци, велике писмености?

### Сажетак

Дигитални истраживачки подаци представљају горућу тему данашњице, која намеће потребу за релативно новим типом писмености, названим писменост података или проналажење и коришћење података (data literacy). Након уводних поставки, у раду се улога овог вида писмености разматра у светлу дељења (заједничког коришћења) података, њиховог квалитета, управљања, чувања и цитирања, који су већином повезани са активностима библиотека и информационих стручњака. Потенцијално релевантне писмености размотрене су у светлу ових питања. Поред информационе писмености, приказане су и карактеристике дигиталне, научне и академске писмености. Истакнут је њихов однос према проналажењу и коришћењу података, које, као вид писмености, представља концепт у фази формирања.

### Кључне речи:

проналажење и коришћење података (data literacy), информациона писменост, управљање подацима, дигитални истраживачки подаци, дељење (заједничко коришћење) података, квалитет података, складиштење података, цитирање података, дигитална писменост, научна писменост, академска писменост

Примљено: 4. марта 2014.  
Исправке рукописа: 21. марта 2014.  
Прихваћено за објављивање: 29. марта 2014.