


Data curation: A conceptual framework for the study of data quality

Working Paper

Author(s):

Leese, Matthias 

Publication date:

2023-02

Permanent link:

<https://doi.org/10.3929/ethz-b-000597540>

Rights / license:

In Copyright - Non-Commercial Use Permitted

Originally published in:

CURATE Working Paper 2

Data curation:

A conceptual framework for the study of data quality

CURATE Working Paper No. 2 – February 2023

Matthias Leese
mleese@ethz.ch

This Working Paper introduces the concept of data curation as a framework for the study of data quality from a social science perspective. Data curation can be understood as a specific set of data practices that is geared towards making and maintaining the trustworthiness of data and explicates how data require continuous care-taking. Analytically, it offers several advantages to the study of data quality. First of all, it draws attention to the fact that data quality is predicated on distributed forms of labour among various social actors and technical means. Secondly, it explicates the multiple interaction points between humans and data along the lives of data and their journeys through the world. And finally, it introduces a terminology that speaks to the needs of practitioners as well as to academic theory-building.

Databases have become key global infrastructures that provide the basis for decision-making and regulation in governance and public administration. Less attention has, however, so far been paid to the quality of the data that are stored in these databases (Leese, 2022). While some scholars have argued that data can indeed be “broken” (Pink et al., 2018), that digital archives need to be considered as epistemically “uncertain” (Thylstrup et al., 2021a), and that there can therefore be considerable trust issues when it comes to datasets (Kitchin, 2021: 9), there are so far no systematic studies that investigate how high data quality for governance and public administration is ensured in practice.

This is somewhat surprising, as unreliable data, even when processed by means of sophisticated algorithmic tools, can make for faulty decisions and interventions with unwanted outcomes. The aim of this Working Paper is develop the concept of data curation as a framework for the study of data quality from a social science perspective.

The idea of curation starts from the assumption that data must not be misunderstood as isolated from the social contexts within which they are created and used. As Thylstrup et al. (2021b: 1) have put forward, data “interact at every level with the human,” thus rendering such interactions key

sites for analyses of how data come to matter in the world. Curation can in this context be understood as a specific set of practices that is geared towards high data quality. It explicates how data require continuous care-taking in order to be made and to remain trustworthy and reliable.

As a conceptual framework, data curation provides several advantages. First of all, it draws attention to the fact that data quality is predicated on distributed forms of labour. Taking care of data inevitably involves various social actors and technical means that in turn come with coordination challenges and power relations. Secondly, data curation explicates the multiple interaction points between humans and data along the lives of data and their journeys through the world. As data are archived or enter different institutional and organizational contexts, they become subject of different use case specifications and corresponding care-taking regimes. And

Curation as care-taking

Per the Oxford Dictionary, the verb “to curate” has two closely related meanings. First, it means the collection, selection, and presentation of information or items for various purposes, including self-presentation, entertainment, learning, or professional knowledge. Second, it means the selection, organization, and presentation of artworks in museums and galleries. Originating from the Latin “curare” (“care”), curation thus primarily pertains to activities that structure the external perception of a person, material objects, or an abstract theme. A curator, in this sense, is someone who actively thinks about external perception of a phenomenon and devises strategies to realize a particular desired result. As a care giver, the curator is in a constant state of concern about the information or objects that they are entrusted with (Gampert, 2020). In art contexts, for example, the tasks of the curator include

finally, data curation introduces a terminology that speaks to the needs of practitioners as well as to academic theory-building. In this sense, it provides a reference point for both understanding and improving, and resonates with a long-standing tradition of action research in the social sciences.

This Working Paper proceeds as follows. It first engages the term curation and discusses how it has been used in different literatures to capture and study matters of organizing and presenting. It then builds on literature from the Library Sciences, Critical Data Studies, and Science and Technology Studies to engage in more depth with the themes of distributed labour, multiple interaction points, and the potential to bridge theory and practice. It concludes by spelling out how these themes relate to broader contexts of governance and public administration, and how data curation can serve as a productive conceptual lens for the study of data quality.

indexing, maintenance, and editing such that the presented works can unfold the desired effects (Campbell and Baars, 2019; Oliveira Lopes, 2020). And for individuals in everyday contexts, curation pertains to the organization of their appearance to others in selective ways, for instance in the context of online identity management (Caldeira et al., 2021; Mahoney, 2022).

In the Library Sciences, “digital curation” has become an important concept in regard to knowledge infrastructures. Sparked by the wide-ranging digitization of research activities and the ensuing establishment and growing importance of large-scale research databases in the 2000s, scholars have mobilized the idea of curation to understand and improve the systematic organization, storage, management, preservation, and access to unprecedented amounts of data in the most practical and goal-oriented fashion.

As Baker and Yarmey (2009: 14) have argued, data curators are in knowledge organizations concerned with particular practices of rendering data “truthful” and “trustworthy.” Curation has thus become a major concern and strategic priority for knowledge organizations such as university libraries, research foundations, or state agencies concerned with research and innovation (American Council of Learned Societies, 2006; Association of Research Libraries, 2006; National Science Foundation, 2003, 2007).

There are notable parallels between this literature and current concerns about data quality in governance and public administration contexts, where data just as

The distributed labour behind data quality

The first aspect of data curation to be discussed here concerns the distributed forms of labour that are involved in the making and maintaining of data quality. This perspective corresponds with the fact that data are made and re-made in relation to their social and technical environments. They are, as Kitchin and Lauriault (2014: 16) assert, part of “data assemblages” that are composed of both infrastructures (databases, repositories, archives) and all other “technological, political, social and economic apparatuses that frames their nature, operation and work.” The important point that Kitchin and Lauriault (2014: 6) highlight is that data and their surrounding assemblages exist in a perpetual state of mutual co-constitution, meaning that they are “bound together in a set of contingent, relational and contextual discursive and material practices and relations.”

Building on such a relational understanding of data, scholars have highlighted how taking care of data consists of various different activities, including the likes of “selection, classification, transformation, validation, and preservation” (Freitas and Curry, 2016: 88).

well need to be taken care of in order to ensure accurate knowledge production and a functional interface between state institutions and citizens (Leese, 2022). Moreover, approaches from the Library Sciences resonate closely with more recent social scientific approaches to the study of data, especially with work from the fields of Critical Data Studies and Science and Technology Studies that has explored digital forms of knowledge production and management. Taken together, these literatures can help to substantiate data curation as a conceptual framework for the study of data quality and will serve as reference points throughout the remainder of this Working Paper.

Importantly, these activities are not carried out by a single actor or technical process. Rather, as Baker and Yarmey (2009: 20) put it, data curation in the context of modern database architectures is “necessarily distributed, as no single person, technology or organization can support all data objects for all purposes including their final long-term maintenance.”

To analyze who is involved in the care-taking of data, Lyon (2007) proposes to map curation activities amongst networks of involved actors, thereby identifying roles, rights, and responsibilities in regard to data – and notably the (power) relationships between these dimensions. With regard to the effective government of data curation, she argues that “there is a real need for tangible leadership and cross-domain strategic co-ordination, ostensibly at the highest levels of research funding organisations, to put in place the infrastructure and services to effectively manage the burgeoning data deluge” (Lyon, 2007: 59). Such leadership can, for instance, involve the setting of standards for data formats, guidelines for metadata production, or the definition of roles and access rights. As

a report by the Association of Research Libraries (2006: 19) points out, curation must accordingly be “understood not merely as preservation of bits and the ability to decode them but also as a system that requires both cooperation across a diversity of organizations, uses, and stakeholders and sustainable models of technical and economic support.”

One aspect of such cooperation relates to the inclusion of data producers in curation activities. Data, as Gitelman and Jackson (2013) point out, are shaped in form and content by decisions that precede their existence. As they argue, “data need to be imagined as data to exist and function as such, and the imagination of data entails an interpretive base” (Gitelman and Jackson, 2013: 3). What they mean by this is that every dataset is created with a particular purpose in mind – and this purpose in turn affects what a dataset eventually looks like: the variables it contains and their characteristics, the classification systems that it uses to categorize empirical phenomena, the format in which the data will be saved, and so on. Moreover, the imagination of data is itself informed by considerations of later use cases, the technical characteristics of databases, or software.

In the Library Sciences, this perspective is mirrored by the acknowledgement that not only institutions, data centers, and end users should be involved in data curation – but

Multiple interaction points

A second aspect of data curation relates to the fact that data do not remain within one place throughout their lives. The ability to travel is, in fact, fundamental to the power and practical appeal of data, as it renders them transferrable across different domain contexts and combinable with other data. Their digital nature, moreover, allows them to be in more than one place at the same time, often

importantly also the scientists and researchers that decide how to produce data in the first place (Lyon, 2007: 9). Raising awareness for formats, structures, and documentation (i.e. metadata) can, so the rationale here, help to prevent potential issues with data quality at later points in time. For this purpose, as Yakei (2007: 338) has put forward, curation should be understood as an “active and potentially interactive process with record creators”, making sure that data production is already set up in a way that facilitates future care-taking tasks. Similarly, Abbott (2008: 1) suggests to see curation as an integral part of any research activities themselves, notably in regard to documentation best practices.

In summary, an important aspect of data curation consists of the coordination of the distributed socio-technical labour of taking care of datasets. Such labour is likely to involve different social actors and infrastructures. Notably, curation sets in even before the actual production of data through the definition of classification systems and the setting of standards. Coordination activities thus must not be limited to data handlers, but need to extend to data producers as well. To study how data quality comes into being in practice, socio-technical networks around data can be mapped, allowing for the identification of processes, structures, and involved power relations (Glouftsiou and Leese, 2023; Pelizza, 2016).

branching them off into multiple simultaneously existing versions that can be put to use within different contexts. To grapple with the mobility and multiplicity of data, scholars have proposed to think through the metaphors of lifecycles and journeys.

The lifecycle approach emphasizes the chronological sequence of the socio-technical relations that data are subjected to from their

creation to their use, migration, and eventually long-term storage (Corti et al.: Ch. 2). Thinking about data through the idea of a lifecycle implies that they undergo certain transformations as they come of age, and that they will be exposed to different interactions and experiences. Kaufmann (2020; see also Kaufmann and Leese, 2021) has in this regard argued that the lifecycle of data should be understood as non-linear and somewhat unpredictable. Notions of retirement or death are, according to this perspective, in fact reversible, as data can always be retrieved from archives and/or re-combined with data from other sources to put them into new use-case contexts. The idea of a data lifecycle thus always already includes the possibility for recycling, or what Bellanova and González Fuster (2019) have called “composting”, i.e. “dead” data turning into something new that comes with renewed analytical value.

The lifecycle model has two implications for data curation. First of all, scholars have highlighted that to ensure effective curation, it is important to cover the entire lifecycle of data, i.e. to start curation activities at the stage of data production and to extend care-taking throughout archives (Lyon, 2007: 48). Long-term storage and the potential for repurposing thereby benefit, as discussed earlier, not only from durable and accessible infrastructures, but also from standards and documentation best practices that have been agreed on in collaboration with data producers. And secondly, it means that data, even after being archived (i.e. put into long-term storage), are in fact never “mothballed” but might still be referenced or re-used in future contexts (Higgins, 2011: 79).

A complementary way to think about the mobility of data is provided by the journey approach (Bates et al., 2016; Leonelli and Tempini, 2020). A data journey, as Leonelli (2020: 9) has argued, can in simple terms be defined as “the movement of data from their production site to many other sites in which

they are processed, mobilised and repurposed”. Importantly, data journeys are not limited to geographical terms, but can also involve mobility across time, different jurisdictions, organizations, and not least cultures. Thinking along the lines of data journeys, as Bates et al. (2016: 3) argue, analytically provides the opportunity to “[follow] data through multiple interconnected organisations and projects within and across knowledge infrastructures.”

Importantly, data journeys are not likely to be smooth and straightforward. Edwards (2010: 84) refers to the notion of friction to describe “the resistance that must always be overcome” when “information systems transform data [...] into information and knowledge.” More specifically, in regard to data, friction for him pertains to “the costs in time, energy, and attention required simply to collect, check, store, move, receive, and access” (Edwards, 2010: 84) data.

Notably, as Bates (2018) has pointed out, the occurrence of data friction in complex socio-technical environments should be considered natural – and yet the reduction of such friction is a major priority for knowledge organizations. This becomes evident for example in cases where data quality requirements differ across domains and organizational contexts (Leonelli, 2020: 6). As Glouftsiou and Leese (2023) have shown, data that are considered of sufficient quality in one use-case context might have to be subjected to additional data quality measures to make them suitable for another one. Curation can in this context be seen as geared towards the reduction of the frictions that will inevitably occur when data travel and enter into new environments that come with particular requirements as to their trustworthiness.

In summary, the notions of both lifecycles and journeys emphasize the fact that data come into contact with multiple and variegated socio-technical contexts. Within these contexts, different data quality requirements

and curation activities can be expected. Moreover, the number of interaction points between humans and data are in theory unlimited, as data never really die but might at any point be resurrected from long-term storage and be used for new tasks. Tracking and reconstructing data lifecycles/journeys provides an analytical incision point to

Bridging theory and practice

Finally, apart from explicating the entanglement of data and care-taking efforts across different contexts, the concept of data curation can also serve to bridge the gap between theory and practice. Social scientific research on data is primarily interested in exploring the knowledge-making functions of data and their role in larger contexts of social and political ordering. In this regard, the idea of curation can accommodate a reflexive angle that engages how the epistemic value of data (i.e. their practical “truthfulness” within knowledge organizations) is enacted through the interactions between humans and data. It thus contributes to research that has challenged empiricist accounts of data by highlighting their constructedness and performativity (Arbia, 2021; Rieder and Simon, 2017).

At the same time, it ties in with a practical stance on data quality as a process within which curation presents a concrete approach to make data more trustworthy and reliable. As Jasanoff (2017: 11) points out, practitioners tend to think about data in representational terms, i.e. to consider them “either unbiased and valid or biased and corrupt.” From such a perspective, thinking about data curation can be understood as “the methodological and technological data management support to address data quality issues maximizing the usability of the data” (Freitas and Curry, 2016: 87). By appealing to such a point of view, data curation resonates closely with the professional lifeworlds and rationales of those who are involved in data management.

identify and explicate curation practices that are necessary for the alignment of data with use case specifications. The study of varying curation regimes across space, time, and culture can thereby help to understand the value given to data quality and the practical ways to achieve and maintain it.

As a conceptual framework, it can thus also help to find a common language for social scientists and practitioners to talk about the care-taking of data. Maybe even more importantly, once such common ground has been established, the framework of data curation put social scientists in a position where they can help address data quality challenges and their repercussions.

The use of theoretical and reflexive insights into complex social settings to help remedy functional and societal shortcomings in fact has a long tradition in the social sciences. Originally motivated by social injustices experienced by minority communities, researchers have under the label of action research (Lewin, 1948) been seeking to inspire social and political change. Born out of the desire to address societal issues, action research generally aims to “[investigate] reality in order to transform it” (Fals Borda, 1979), thereby building on an understanding of reflective thinking as a state of doubt (Dewey, 1933) and aspiring to open communicative spaces to resolve that doubt in a collaborative fashion (Loughran, 2010). In this capacity, it has become a widely applied approach, for example in contexts of education (Stringer, 2014a), health care (Hughes, 2008), or organizational change (Urquhart and Wearing, 2017).

In regard to data curation, an action research perspective can be productive in several ways. Firstly, it builds on the mutual constitution between knowledge and practice. Rather than

considering research as a linear process of producing knowledge and then applying this knowledge to practical contexts later on, action research already integrates the development of practice into the research process in a cyclical fashion. In doing so, it aims to co-develop theory and practice from inside a social setting. Research activities are carried out in close collaboration with research participants and can therefore speak closely to the lifeworlds and rationales of the involved persons. It thus offers the possibility to translate insights into data quality issues into practical remedies in collaboration with practitioners. Generally speaking, action research can be considered a highly reflexive, experiential, and participatory mode of research in which both researchers and research participants deliberately contribute to the research goals (Fals Borda and Rahman, 1991; Stringer, 2014b).

Conclusions

In summary, there are several theoretical, normative, and practical considerations of data quality that the concept of data curation can help to address. As this Working Paper has outlined, the notion of taking care of data, first of all, builds on the socio-technical composition and embeddedness of any dataset, thus highlighting the distributed labour that must by definition go into the making and maintaining trustworthiness and reliability in data. Secondly, it has highlighted the multiple interaction points between data and humans that occur as data travel from one place to another. This includes their subjection to different use-cases and data quality specifications, resulting in different care-taking activities. And finally, the Working Paper has outlined that the conceptual framework of data curation can serve as a connector between theory and practice, speaking to both the requirements of reflexive theory-building and practical improvement. As such, it resonates closely

And secondly, thinking data quality through an action research perspective highlights the normative dimension of data curation. Data curation is in itself a normative practice that seeks to render data as trustworthy and reliable as possible. Moreover, curation is particularly pertinent in light of the potential negative effects of low quality data. The terminology of curation can in this sense provide a productive angle for a constructive problematization of data practices that takes into consideration practitioners' perspectives on knowledge production and modes of meaning making (McTaggart et al., 2017: 27). It can thus be understood as geared towards developing a constructive form of critique that can assist practitioners in explicating and conceptualizing their tacit knowledges and practices and help them to reflect on the consequences of their professional actions (Kemmis, 2008).

with the transformative ambition of action research.

Databases for public administration and governance, so the argument put forward at the outset, are currently facing challenges that are similar to what research and science institutions have faced during the wave of digitization of science activities in the 2000s. When geared towards the study of data quality in contexts of governance and public administration, there are several lessons that can be drawn from literature from the Library Sciences, Science and Technology Studies, as well as Critical Data Studies.

Data curation, in this sense, first of all directs attention to the challenges and potential best practices of managing data from heterogeneous sources within complex, multi-level database architectures. Moreover, it draws attention to the sensitivities that are required to tend to data and that often stand in contrast to automated forms of ordering

data and thus helps to differentiate narratives of Big Data that tend to totalize knowledge production as a technical question of data exploitation. Finally, and similar to data in research and science, data in governance and public administration contexts, after their production and processing, end up in archives or might at some point be deleted because they

lost their informational value for particular use cases. Data curation highlights the continuous role of human care-taking activities that include the choice of appropriate infrastructures to house data long-term and keep them accessible such that they can serve as a reliable basis for as many use cases as possible.

References

- Abbott D (2008) What is Digital Curation? DCC Briefing Papers: Introduction to Curation. Edinburgh: Digital Curation Centre.
- American Council of Learned Societies (2006) Our Cultural Commonwealth: The Report of the American Council of Learned Societies Commission on Cyberinfrastructure for the Humanities and Social Sciences. New York: American Council of Learned Societies.
- Arbia G (2021) *Statistics, New Empiricism and Society in the Era of Big Data*. Cham: Springer.
- Association of Research Libraries (2006) To Stand the Test of Time: Long-term Stewardship of Digital Data Sets in Science and Engineering. Washington, DC: Association of Research Libraries.
- Baker K S and Yarmey L (2009) Data Stewardship: Environmental Data Curation and a Web-of-Repositories. *The International Journal of Digital Curation* 4(2): 12-27.
- Bates J (2018) The Politics of Data Friction. *Journal of Documentation* 74(2): 412-429.
- Bates J, Lin Y-W and Goodale P (2016) Data Journeys: Capturing the Socio-Material Constitution of Data Objects and Flows. *Big Data & Society* 3(2): 1-12.
- Bellanova R and González Fuster G (2019) Composting and Computing: On Digital Security Compositions. *European Journal of International Security* 4(3): 345-365.
- Caldeira S P, Van Bauwel S and Ridder S D (2021) 'Everybody Needs to Post a Selfie Every Once in a While': Exploring the Politics of Instagram Curation in Young Women's Self-representational Practices. *Information, Communication & Society* 24(8): 1073-1090.
- Campbell B A and Baars C (eds.) 2019. *The Curation and Care of Museum Collections*, London / New York: Routledge.
- Corti L, Van der Eynden V, Bishop L and Wollard M *Managing and Sharing Research Data: A Guide to Good Practice*. Los Angeles / London / New Delhi / Singapore / Washington: Sage.
- Dewey J (1933) *How We Think: A Restatement of the Relation of Reflective Thinking to the Educative Process*. Chicago: Henry Regnery Company.
- Edwards P N (2010) *A Vast Machine: Computer Models, Climate Data, and the Politics of Global Warming*. Cambridge: MIT Press.
- Fals Borda O (1979) Investigating Reality in Order to Transform it: The Colombian Experience. *Dialectical Anthropology* 4(1): 33-55.
- Fals Borda O and Rahman M A (eds.) 1991. *Action and Knowledge: Breaking the Monopoly with Participatory Action-Research*, New York: Intermediate Technology/Apex.
- Freitas A and Curry E (2016) Big Data Curation. In Cavanillas J M, Curry E & Wahlster W (eds.) *New Horizons for a Data-Driven Economy: A Roadmap for*

- Usage and Exploitation of Big Data in Europe*. Cham: Springer, 87-118.
- Gampert C (2020) Was machen Kuratorinnen und Kuratoren? *Deutschlandfunk*, 19 May 2020.
- Gitelman L and Jackson V (2013) Introduction. In Gitelman L (ed.) *"Raw Data" is an Oxymoron*. Cambridge/London: MIT Press, 1-14.
- Glouftsiou G and Leese M (2023) Epistemic Fusion: Passenger Information Units and the Making of International Security. *Review of International Studies* 49(1): 125-142.
- Higgins S (2011) Digital Curation: The Emergence of a New Discipline. *The International Journal of Digital Curation* 6(2): 78-88.
- Hughes I (2008) Action Research in Healthcare. In Reason P & Bradbury H (eds.) *The SAGE Handbook of Action Research: Participative Inquiry and Practice. Second Edition*. Los Angeles / London / New Delhi / Singapore: Sage, 381-393.
- Jasanoff S (2017) Virtual, Visible, and Actionable: Data Assemblages and the Sightlines of Justice. *Big Data & Society* (July-December): 1-15.
- Kaufmann M (2020) Vocations, Visions and Vitalities of Data Analysis: An Introduction. *Information, Communication & Society* 23(14): 1981-1995.
- Kaufmann M and Leese M (2021) Information In-Formation: Algorithmic Policing and the Life of Data. In Badalič V & Završnik A (eds.) *Automating Crime Prevention, Surveillance, and Military Operations*. Cham: Springer, 69-83.
- Kemmis S (2008) Critical Theory and Participatory Action Research. In Reason P & Bradbury H (eds.) *The SAGE Handbook of Action Research: Participative Inquiry and Practice. 2nd Edition*. London/Thousand Oaks/New Delhi/Singapore: Sage, 121-138.
- Kitchin R (2021) *Data Lives: How Data are Made and Shape our World*. Bristol: Bristol University Press.
- Kitchin R and Lauriault T P (2014) Towards Critical Data Studies: Charting and Unpacking Data Assemblages and Their Work. The Programmable City Working Paper 2.
- Leese M (2022) Data Quality in Governance: A Definition and a Research Agenda. CURATE Working Paper No. 1, November 2022. doi:10.3929/ethz-b-000581434.
- Leonelli S (2020) Learning From Data Journeys. In Leonelli S & Tempini N (eds.) *Data Journeys in the Sciences*. Cham: Springer, 1-24.
- Leonelli S and Tempini N (eds.) 2020. *Data Journeys in the Sciences*, Cham: Springer.
- Lewin K (1948) Action Research and Minority Problems. In Lewin G W (ed.) *Resolving Social Conflicts: Selected Papers on Group Dynamics*. New York: Harper & Brothers Publishers, 201-216.
- Loughran J (2010) Reflection Through Collaborative Action Research and Inquiry. In Lyons N (ed.) *Handbook of Reflection and Reflective Inquiry: Mapping a Way of Knowing for Professional Reflective Inquiry*. New York / Dordrecht / Heidelberg / London: Springer, 399-413.
- Lyon L (2007) Dealing With Data: Roles, Rights, Responsibilities and Relationships. UKOLN Consultancy Report. Bath: University of Bath.
- Mahoney C (2022) Is This What a Feminist Looks Like? Curating the Feminist Self in the Neoliberal Visual Economy of Instagram. *Feminist Media Studies* 22(3): 519-535.
- McTaggart R, Nixon R and Kemmis S (2017) Critical Participatory Action Research. In Rowell L L, Bruce C D, Shosh J M & Riel M M (eds.) *The Palgrave International Handbook of Action Research*. Basingstoke: Palgrave Macmillan, 21-35.
- National Science Foundation (2003) Revolutionizing Science and Engineering Through Cyberinfrastructure: Report of the

- National Science Foundation Blue-Ribbon Advisory Panel on Cyberinfrastructure. Alexandria: National Science Foundation.
- National Science Foundation (2007) Cyberinfrastructure Vision for 21st Century Discovery. Alexandria: National Science Foundation.
- Oliveira Lopes R (2020) Museum Curation in the Digital Age. In Hearn G (ed.) *The Future of Creative Work: Creativity and Digital Disruption*. Cheltenham: Edward Elgar Publishing, 123-139.
- Pelizza A (2016) Disciplining Change, Displacing Frictions: Two Structural Dimensions of Digital Circulation Across Land Registry Database Integration. *Tecnoscienza: Italian Journal of Science and Technology Studies* 7(2): 35-60.
- Pink S, Ruckenstein M, Willim R and Duque M (2018) Broken Data: Conceptualising Data in an Emerging World. *Big Data & Society* 5(1): 1-13.
- Rieder G and Simon J (2017) Big Data: A New Empiricism and its Epistemic and Socio-Political Consequences. In Pietsch W, Wernecke J & Ott M (eds.) *Berechenbarkeit der Welt? Philosophie und Wissenschaft im Zeitalter von Big Data*. Wiesbaden: Springer, 85-105.
- Stringer E (2014a) *Action Research in Education*. Harlow: Pearson.
- Stringer E T (2014b) *Action Research. 4th Edition*. Los Angeles/London/New Delhi/Singapore/Washington D.C.: Sage.
- Thylstrup N B, Agostinho D, Ring A, D'Ignazio C and Veel K (eds.) 2021a. *Uncertain Archives: Critical Keywords for Big Data*, Cambridge: MIT Press.
- Thylstrup N B, Agostinho D, Ring A, D'Ignazio C and Veel K (2021b) Big Data as Uncertain Archives. In Thylstrup N B, Agostinho D, Ring A, D'Ignazio C & Veel K (eds.) *Uncertain Archives: Critical Keywords for Big Data*. Cambridge: MIT Press, 1-27.
- Urquhart R and Wearing M (2017) Organisational Change in Non-Profit Human Services: Reflections on a Collaborative Action Research Approach to Working with Child, Youth, and Family Organisations. In Rowell L L, Bruce C D, Shosh J M & Riel M M (eds.) *The Palgrave International Handbook of Action Research*. Basingstoke: Palgrave Macmillan, 545-562.
- Yakel E (2007) Digital Curation. *OCLC Systems & Services: International Digital Library Perspectives* 23(4): 335-340.

Suggested citation: Leese M (2023) Data curation: A conceptual framework for the study of data quality. CURATE Working Paper No. 2, February 2023. doi:10.3929/ethz-b-000597540.