# Curating/Fermenting Data: data workflows for semantic web applications

Curating/Fermenting Data

DR MAGDALENA TYŻLIK-CARVER

Associate Professor, Dept. Digital Design and Information Studies, Aarhus University, magdatc@cc.au.dk

DR LOZANA ROSSENOVA

Associate Researcher, Open Science Lab, TIB – Leibniz Information Centre for Science and Technology, lozana.rossenova@tib.eu

DR LUKAS FUCHSGRUBER

Post-doctoral Researcher, TU Berlin, fuchsgruber@tu-berlin.de

This one-day workshop will bring together HCI researchers, Linked Open Data (LOD) experts, curators, designers and cultural practitioners to explore applications of LOD for the creation of participatory datasets. The aim of the workshop is to work with a collection of small unstructured data while engaging with questions of ethical data use in HCI design and working with data against a "colonial impulse" [22]. Starting from the Curating Data diagram (Appendix A) and working with data gathered during a series of workshops in the project Fermenting Data [48], we will create speculative ontologies for the Fermenting Data dataset. The practice component of the workshop will interrogate the design of data curation workflows for semantic web applications, while working with actual ferments in order to physically engage with the process at the source of this dataset.

**CCS CONCEPTS** • HCI • Interaction paradigms • Data models • Data provenance • Data management systems • Data structures;

**Additional Keywords and Phrases:** Curating data • Linked Open Data • Decolonial Computing • Wikidata • Data lifecycle

**ACM Reference Format:**
Magdalena Tyżlik-Carver, Lozana Rossenova, and Lukas Fuchsgruber. 2018. Curating/Fermenting Data: data workflows for semantic web applications. Workshop Proposal Submission for NordiCHI2022 Participative Computing for Sustainable Future Conference, October 08-12, 2022, Aarhus, Denmark.

## 1 WORKSHOP BACKGROUND AND MOTIVATION

In both machine learning and semantic web applications, a primary assumption about data – namely that data are abstract and as such universal – continues to prevail, while data are also considered a natural resource like oil, air or water. However, as a number of information studies, digital humanities and critical data studies scholars have argued, data are not raw [7, 29] because in order to exist and function as data they have to be imagined as such [29]. Against the etymological roots of the word data in the Latin language, some have argued that data should be understood as 'capta' [38, 23], not given but taken, to account for processes that facilitate data capture. Such processes do not mirror the social [39] but rather depend on interpretations and abstractions of objects, people and relations into data. In effect, data models which are designed to process myriad data points and network relations obscure the fact that these data come from somewhere and that they had been collected under specific conditions that are social, historical, cultural and technological. We believe that cross-disciplinary collaborations are necessary to practically and theoretically interrogate systems of data processing. The methodology that we propose to adopt in this workshop, which takes the form of curating data [50], can facilitate this because it combines 1) curatorial care for data as digital objects that are part of specific systems of knowledge, with 2) practical work on data and the associated systems for storage, structuring and retrieval.

### 1.1 Background

Having worked in the curatorial and digital art and design contexts, as well as in the field of information and data management (also in museums), we observe that calls to situate knowledges [31] and recognition of data as local [40] identify and question totalising tendencies of presumably universal data. In fact, an increasing number of critical data studies and computer science projects show how data are biased [10, 43] and how they discriminate [21, 24]. Studies of labeling of datasets in machine learning (ML) [18], facial recognition models [10] and biometrics [41, 27] establish how these technologies are linked to racist and gendered assumptions that undergird the data categories used in building datasets. The situation in the field of semantic web applications and linked open data is not much different. For example, the largest open and publicly curated data platform – Wikidata – is routinely selected as a source to study pervasive biases and universalist tendencies among the communities that decide on the ontological order of the data and debate the nuance of how unstructured information can be turned into structured data, and ultimately a knowledge base [37, 45]. At the same time, Wikidata offers a uniquely diverse, networked and social system for data processing - wherein open data from many different knowledge domains gets enmeshed - and possibilities for intersection and symbiosis exist even against the grain of the totalising impulse [45]. We find this tension productive and we will utilise the relation between the universalising systems and their data sources in the workshop.

### 1.2 Troubling data as part of colonial epistemologies

Not only is this data age a moment of profound data extraction and surveillance [17, 53] but these exploitative and universalising tendencies in data practices are historically rooted in colonial and racist laws [9, 6, 15]. Data colonising logic [15, 4] is present in data processing as a form of knowledge-making and in "the conception of colonialism as a knowledge enterprise" [22]. In order to tackle the problems of data in the algorithmic age there is a growing need to investigate very closely how technology and culture "intersect and collide" because, as Wendy Hui Kyong Chun states, "drilling down into any technology reveals profound social and cultural assumptions, and vice versa" [12] which bleed into how we live, what we know, and how we know it. We are mindful that interventions into data technologies and related practices require more than just critique [33]. What might be the practice of "drilling down" into digital data

systems so that the critique of data as extractive is accompanied by interventions that recognise data as representation of "colonial ways of knowing and being" [4] and propose ways of doing data-based practices otherwise? The question of which practices get to produce knowledge [8] demands that we consider the political dimensions of datasets [46] while also learning from archives [36] and decolonial activism in art, museums and beyond [5, 20, 47, 19]. Drawing on the critical scholarship referenced so far, this workshop engages interdisciplinary collaborations with the intention to practically interrogate systems of data processing, such as categorisation and ontology modeling, as part of the Curating Data methodology.

### 1.3 Troubling linked open data while learning form archives

Archival science theory of the past several decades has reoriented itself towards the digital paradigm with a particular focus on acknowledging the subjective position of the archivist and the colonial roots of this position [14, 11, 3], while aiming to engage stakeholder communities to take up an active role in processes of categorisation, preservation and access. Within this *post-custodial* approach the emphasis is on the importance of "the context, purpose, intent, interrelationships, functionality, and accountability of the record and especially its creator and its creation processes" [14]. However, while there has been a shift away from static archival cataloguing towards "mapping dynamic relationships" in theory, for the most part, digital archive systems remain unable to account for this in practice [34, 51]. At the same time, linked open data and the semantic web have been widely discussed in this context – as technical paradigms that can answer the needs of cultural heritage data and digital preservation [35, 25].

Data traceability and provenance play important roles in the conceptualization and specification of the semantic web protocols [30]. In broad terms, the semantic web is the proposition to move towards a web of data, which is machine-readable, so that computers can handle web data in more useful ways than was previously possible through the capabilities of HTML documents alone [32]. The related concept of Linked Open Data (LOD) proposes concrete protocols that specify how machine-readable data should be made openly available and interoperable on the web. In this web of connected data, the ability to read the source of the data and the agent who creates data is crucial to semantic interpretation of data, i.e. to making meaning of data (either as a human user, or as a software application). However, software applications that are designed to interface with linked open data remain largely inaccessible to more general audiences [42; 26; 44]. The requisite high degree of data and computer science literacy perpetuate an exclusive environment, often ignorant of its own inherent biases [37, 45]. This situation, nonetheless leaves plenty of space for HCI designers and other concerned stakeholders to intervene and develop alternative paradigms for user participation amidst the increasing datafication of the open web.

### 1.4 From data curation to Curating Data

Data provenance and data quality are some of the cornerstones of data curation. The processes of data curation follow data lifecycles so that the value and quality of data is sustained. However, in the wider ecologies of curating data within increasingly complex networks across machine-readable databases, algorithmic timelines, Internet of Things, and webs of meanings and sites, people connect with other people, while also connecting with things that are linked in a wide web of relations. Within this wide range of data practices, the question of data curation needs to move beyond the assurances of data quality to include development of ethical frameworks, and corresponding HCI design patterns, in which these data function. Such frameworks need to incorporate questioning of the physical origin and circumstance of dataset creation, as well as the methods applied in the modelling and categorisation of data. Making visible the conditions of datasets creation and their material origins contributes to efforts to increase the quality of data, while developing

ways to make the models and categorisations applied to the data more open to scrutiny, supports data literacies beyond the computer and data sciences professionals.

As Gebru, et al. [28] have asserted, questions of data provenance are discussed extensively in the databases communities, but less so in the machine learning communities. Equally, such questions are rarely discussed in interface design and HCI contexts [44]. Although increasingly HCI is moving away from the hyper focus of studying and designing for users as discrete individuals – as in the case of more-than-human-design (MTHD) approaches [16] or approaches concerned with a holistic view of the environment [21] – still the design of products that deal with large quantities of curated data rarely engages with, or questions, the process of curation. This is particularly well exhibited in the context of semantic web applications and linked open data wherein interaction requires a change in the user's mental model away from traditional understanding of datasets as tables or hierarchical trees towards the model of a graph – a network of nodes (i.e. things in the world) and edges (the relations between them). This mental model mismatch, paired with scarcity of adequate interface design patterns to represent graph data models [42], results in high learning curves. The prerequisite for data science expertise results in the low adoption of semantic web applications in the workflows of data collectors, curators, and users, and consequently also in the low capacity to critically engage with the technical capabilities and limitations of these applications [44].

Whether part of a database or a set, data start with bodies, human or not, and they index relations in the most abstract way. Hence, "the task in curating data is to reclaim their traceability, and to account for their lineage" [49]. The attention to this places the value of data in strategies and actions that are about sustaining links to data sources in the form of annotations and data sheets [28] that explain the conditions of making datasets. Curating Data methods open the possibility to keep data 'lively' through curation workflows while actively engaging with questions about the design and implementation of technical systems that can support varying levels of interaction with data and their traces.

## 2 CALL FOR WORKSHOP PARTICIPATION

We invite designers, data analysts, curators, computer scientists, archivists and others to collaboratively curate data. This workshop is organised as part of NordICHI22 conference in Aarhus and aims to experiment with speculative methods for curating data, while working practically with Linked Open Data tools and protocols for data modelling and annotation. During the workshop we will collaboratively create a structured dataset from 'raw' and unstructured data and we will document this process in order to critically reflect on decisions made during our workflows, also speculating on its possible future use and the way such dataset might be (mis)managed.

### Workshop methods:

- Fermenting data: Work with raw (physical) materials alongside their digital data proxies to create a case study dataset following the structuring principles of the semantic web.
- Prototyping: Collaboratively sketch out strategies towards making the underlying data model more visible and transparent to outside interpretation.
- Documentation: Discuss together decolonial frameworks around data publishing protocols through documenting the process of curation and annotation of the case study dataset.

### Workshop benefits:

- Learn about the processes involved in curating and fermenting data .

- Learn about the organising principles of the semantic web and what roles they could play in improving data provenance and quality.
- Explore the linked open data environment of Wikidata.
- Contribute towards making data workflows more intelligible to fellow interdisciplinary researchers and end users alike.

### Two main questions frame the workshop:

*How to curate data so that the creation of a dataset delivers information about how it was created, by whom, for what reason and under what circumstances:* We believe such information is important as it has the potential to "increase transparency and accountability."[28]. Unpacking this problem requires attending to workflows more closely connected to the social processes surrounding data collection and the abstraction of data from real-world items and events into digital proxy concepts and processes. This is where the Fermenting Data project [48] will provide a rich context for workflow experimentation and reflection.

*How to create and publish datasets that can be interpreted semantically, i.e. meaningfully, by both humans and machines:* This is an HCI and more-than-human-design problem concerning the digital infrastructure that facilitates data curation workflows. We take the concrete environment of semantic web applications, more specifically – the public LOD platform Wikidata – as spaces that require design intervention and reconceptualising design patterns that befit mental models for interaction with non-hierarchical, complexly networked, yet traceable datasets.

## 3 WORKSHOP STRUCTURE

### 3.1 Planned activities

We propose a full day, practice-based workshop (9:30am-5:15pm) which will start with introductions of the participants and framing statements from the workshop facilitators. During this warming up period participants will have the opportunity to smell and taste the ferments which will serve the "raw" basis of the dataset work. The workshop will consist of morning and afternoon sessions which will focus on the creation of a structured dataset, and its integration into a linked open data environment, respectively.

The morning session is titled '**Getting to know your data'** and will include physical activities with our data subjects (selecting and chopping vegetables, reviewing recipes), as well as time to familiarise oneself with the available data from the Fermenting data project, which all participants will have access to and be able to use as building blocks for the structured dataset. During this session participants will engage in a speculative data modeling exercise beginning to categorise the unstructured data within the framework of semantic triples, i.e. data statements consisting of entities (e.g. a particular fermentation process, a recipe, or an individual vegetable) and the relations between them.

The afternoon session – '**Visualising Data Ontologies**' – will be dedicated to working within Linked Open Data environments, thinking through the problems of data ontologies and the ethical questions around data publication and dataset annotation. We will continue hands-on activities with our ferments: salting, crunching and massaging, placing into jars. The temporal breaks required by the fermenting process will leave us time to explore the open, collaborative environment of Wikidata. We will analyse how the speculative categorisations and data relations developed in the morning fit (or not) within the ontology of Wikidata. Participants will be encouraged to investigate different ways to visualise data entities and the connections between them – either using the existing capabilities of the Wikidata interface, or prototyping new interaction design proposals.

Discussions and reflections will be encouraged throughout the session, and we will share findings and prototypes at the end of the workshop, and participants will be able to take their jars of ferments with them so they can be consumed once the fermentation sets in.

### 3.2 Workshop outcomes

The workshop will generate curating data workflows with accompanying documentation and it will also help to clarify how the Curating Data methodology might be helpful in addressing questions of data quality in interdisciplinary collaborations. These results will be the basis for an article submitted to a relevant journal. If agreed with participants, we will submit the visualisations of data models generated during the workshop, to Exhibit-X section of the ACM interactions magazine. As all workshop activities will accrue a layer of documentation – be it data spreadsheets, data model sketches, or jarred ferments – this documentation will be used as a prompt to address the 'afterlife' of dataset creation: licensing concerns, data source annotation, potential future (mis)use.

### 3.3 Promotional strategy

We plan to set up a dedicated web page for the workshop and an online form for prospective participants to express interest in joining the workshop. This promotional page will then be widely distributed on relevant academic and professional mailing lists. We may also reach out directly – via email or social media channels – to practitioners and researchers whose work is relevant to the topics of this workshop, inviting them to sign up. We will aim to recruit a diverse range of participants representing different disciplinary and cultural perspectives. We plan for a maximum number of 30 participants.

### 3.4 Organisers' background

Dr Magdalena Tyżlik-Carver is Associate Professor of Digital Communication and Culture at the Department of Digital Design and Information Studies at Aarhus University and Associate Researcher at Centre for the Study of the Networked Image at London South Bank University. She is curator of digital art and design and arts and humanities scholar of critical data studies. Her interdisciplinary research into computational cultures focuses on data as material practice where she explores how curating data can be a methodology for decolonial data practice.

Dr Lozana Rossenova is a digital humanities researcher and designer. Her doctoral research focused on questions related to presentation and performativity in the online archive of born-digital art examined through the lens of interface design (theory and practice). Rossenova is currently based at the Open Science Lab at TIB, Hanover, working on the NFDI4Culture project towards a national research infrastructure of cultural heritage data. She is an active member of the Wikidata and Wikibase open source development communities, and a co-founder of the Wikibase Stakeholder Group.

Dr Lukas Fuchsgruber is a post-doctoral researcher at TU Berlin in the collaborative project "Museums and Society - Mapping the Social" of TU Berlin, HU Berlin, the Institute for Museum Research of the State Museums in Berlin, and the Museum for Natural History Berlin. He studied art history and published his dissertation on the 19th century auction market in Paris (Das Spektakel der Auktion, 2021). Currently, he is investigating the digital image worlds created around museums, focusing on historical and contemporary social aspects of interfaces for digital commons.

### 3.5 Flexibility in terms of hybrid participation

Due to the hands-on nature of the activities proposed in this workshop, it is planned as an in-person event. However, in the case of a change of circumstances connected to the COVID-19 pandemic and increased restrictions on physical contacts, the workshop can be reorganised into an online event.

## REFERENCES


[1] Russell L. Ackoff. 1989. From Data to Wisdom. *Journal of Applied Systems Analysis* 16: 3–9.

[2] Mortimer J. Adler 1986. *A Guidebook to Learning: For a Lifelong Pursuit of Wisdom*. New York: Macmillan.

[3] Daniela Agostinho, Catherine D'Ignazio, Annie Ring, Nanna Bonde Thylstrup, and Kristin Veel. 2019. Uncertain Archives: Approaching the Unknowns, Errors, and Vulnerabilities of Big Data through Cultural Theories of the Archive. *Surveillance & Society* 17(3/4):422–41. doi: 10.24908/ss.v17i3/4.12330.

[4] Syed Mustafa Ali. 2016. A Brief Introduction to Decolonial Computing. *XRDS: Crossroads, The ACM Magazine for Students* 22(4):16–21. doi: 10.1145/2930886.

[5] Ariella Aïsha Azoulay. 2019. *Potential History: Unlearning Imperialism*. Verso.

[6] Ruha Benjamin. ed. 2019. *Captivating Technology: Race, Carceral Technoscience, and Liberatory Imagination in Everyday Life*. Durham: Duke University Press Books.

[7] Geoffrey C. Bowker. 2008. *Memory Practices in the Sciences*. 1. paperback ed. Cambridge, Mass.: MIT.

[8] Loren Britton and Helen Pritchard. 2020. For CS. *IX Interactions*. Retrieved November 2, 2020 (https://interactions.acm.org/blog/view/for-cs).

[9] Simone Browne. 2015. *Dark Matters: On the Surveillance of Blackness*. Durham: Duke University Press Books.

[10] Joy Buolamwini and Timnit Gebru. 2018. Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification. Pp. 1–15 in *Proceedings of Machine Learning Research*. Vol. 81.

[11] Michelle Caswell. 2019. Dusting for Fingerprints: Introducing Feminist Standpoint Appraisal. *Journal of Critical Library and Information Studies* 3, 1. Retrieved May 20, 2022 from https://journals.litwinbooks.com/index.php/jclis/article/view/113/67

[12] Wendy Hui Kyong Chun. 2021. *Discriminating Data: Correlation, Neighborhoods, and the New Politics of Recognition*. Cambridge, MA, USA: MIT Press.

[13] Terry Cook. 2007. Electronic records, paper minds: the revolution in Information Management and Archives in the post-custodial and post-modernist era. *Archives & Social Studies: A Journal of Interdisciplinary Research* 1, 0, 399–443

[14] Terry Cook. 2001. Archival science and postmodernism: New formulations for old concepts. *Archival Science* 1, 1, 3–24. DOI:10.1007/BF02435636.

[15] Nick Couldry and Ulises A. Mejias. 2019. Data Colonialism: Rethinking Big Data's Relation to the Contemporary Subject. *Television & New Media* 20(4):336–49. doi: 10.1177/1527476418796632.

[16] Paul Coulton and Joseph Galen Lindley. 2019. More-Than Human Centred Design: Considering Other Things. *The Design Journal* 22(4):463–81. doi: 10.1080/14606925.2019.1614320.

[17] Kate Crawford and Vladan Joler. 2018. Anatomy of an AI system. -. Retrieved May 19, 2022 (https://anatomyof.ai/).

[18] Kate Crawford and Trevor Paglen. 2019. Excavating AI. Retrieved May 19, 2022 (https://www.excavating.ai).

[19] Subhadra Das and Miranda Lowe. 2018. Nature Read in Black and White: Decolonial Approaches to Interpreting Natural History Collections. *Journal of Natural Science Collections* 4–14.

[20] Clémentine Deliss. 2020. *The Metabolic Museum*. Berlin: H. Cantz Verlag.

[21] Paul Dourish. 2010. HCI and Environmental Sustainability: The Politics of Design and the Design of Politics. In *Proceedings of the 8th ACM Conference on Designing (Aarhus, Denmark '10). Association for Computing Interactive Systems (DIS Machinery, New York)*, NY, USA, 1–10. doi.org: 10.1145/1858171.1858173

[22] Paul Dourish and Scott D. Mainwaring. 2012. Ubicomp's Colonial Impulse. P. 133 in *Proceedings of the 2012 ACM Conference on Ubiquitous Computing - UbiComp '12*. Pittsburgh, Pennsylvania: ACM Press.

[23] Johanna Drucker. 2012. Humanistic Theory and Digital Scholarship. Pp. 85–95 in *Debates in the Digital Humanities*, edited by M. K. Gold. University of Minnesota Press.

[24] Virginia Eubanks. 2018. *Automating Inequality: How High-Tech Tools Profile, Police and Punish the Poor*. First Edition. New York, NY: St. Martin's Press.

[25] Lukas Fuchsgruber. 2022. Image Archives in the Age of Surveillance Capitalism. In Moritz Neumüller (ed.): *The Routledge Companion to Photography, Representation and Social Justice*. Routledge (forthcoming).

[26] Roberto García, Rosa Gil, Juan Manuel Gimeno, Eirik Bakke & David R. Karger. 2016. BESDUI: A benchmark for end-user structured data user interfaces. In *The Semantic Web – ISWC 2016. ISWC 2016. Lecture Notes in Computer Science*, vol 9982. Springer, Cham. https://doi.org/10.1007/978-3-319-46547-0_8

[27] Kelly Gates. 2011. *Our Biometric Future: Facial Recognition Technology and the Culture of Surveillance*. New York: New York University Press.

[28] Timnit Gebru, Jamie Morgenstern, Briana Vecchione, Jennifer Wortman Vaughan, Hanna Wallach, Hal Daumé III, and Kate Crawford. 2021. Datasheets for datasets. *Commun. ACM* 64, 12 (December 2021), 86–92. doi.org: 10.1145/3458723

[29] Lisa Gitelman and Virginia Jackson. 2013. *"Raw Data" is an Oxymoron*. MIT Press.

[30] Paul Groth and Luc Moreau. 2013. PROV-Overview: An Overview of the PROV Family of Documents. *W3C Working Group Note 30 April 2013*. Retrieved May 20, 2022 https://www.w3.org/TR/2013/NOTE-prov-overview-20130430/

[31] Donna J. Haraway. 1988. Situated Knowledges: The Science Question in Feminism and the Privilege of Partial Perspective. *Feminist Studies* 14(3):575–99. doi: 10.2307/3178066.

[32] James Hendler, Ora Lassila, and Tim Berners-Lee. 2001. The semantic web. *Scientific American* 284, 5, 34-43. Retrieved May 20, 2022 from http://www.jstor.org/stable/26059207

[33] Anna Lauren Hoffmann. 2021. Even When You Are a Solution You Are a Problem: An Uncomfortable Reflection on Feminist Data Ethic. *Global Perspectives* 2(1). doi: 10.1525/gp.2021.21335.

[34] Chris Hurley. 2005. Parallel Provenance. Retrieved May 20, 2022 https://www.descriptionguy.com/images/WEBSITE/parallel-provenance.pdf

[35] Eero Hyvönen. 2012. *Publishing and using cultural heritage linked data on the semantic web. Synthesis Lectures on the Semantic Web: Theory and Technology* 2, 1, 1–159.

[36] Eun Seo Jo and Timnit Gebru. 2020. Lessons from Archives: Strategies for Collecting Sociocultural Data in Machine Learning. Pp. 306–16 in *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency, FAT* '20*. New York, NY, USA: Association for Computing Machinery.

[37] Os Keyes. 2019. Questioning Wikidata. In *WikidataCon 2019*, Berlin, October 24–26, 2019.Retrieved May 20, 2022 from https://www.wikidata.org/wiki/Wikidata:WikidataCon_2019/Program/Sessions/Keynote:_Questioning_Wikidata

[38] Rob Kitchin. 2014. *The Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequences*. SAGE Publications Ltd.

[39] Ganaele Langlois, Joanna Redden, and Greg Elmer, eds. 2015. *Compromised Data: From Social Media to Big Data*. New York, NY ; London: Bloomsbury Academic, an imprint of Bloomsbury Publishing, Inc.

[40] Yanni A. Loukissas. 2019. *All Data Are Local: Thinking Critically in a Data-Driven Society*. Cambridge, Massachusetts: The MIT Press.

[41] Mark Maguire. 2009. The Birth of Biometric Security. *Anthropology Today* 25:9–14. doi: 10.1111/j.1467-8322.2009.00654.x.

[42] M. C. Schraefel and David Karger. 2006. The pathetic fallacy of rdf. *International workshop on the semantic web and user interaction (SWUI)*. Vol. 2006. Retrieved May 20, 2022 from https://eprints.soton.ac.uk/262911/

[43] Sofiya Noble. 2018. *Algorithms of Oppression: How Search Engines Reinforce Racism*. 1 edition. New York: NYU Press.

[44] Lozana Rossenova. 2021. *Model–Database–Interface: A study of the redesign of the ArtBase, and the role of user agency in born-digital archives*. Ph.D. Dissertation. London South Bank University School of Art and Creative Industries. https://doi.org/10.18744/lsbu.8wz7x

[45] Anasuya Sengupta. 2021. Decolonising Wikidata: Why Does Knowledge Justice Matter for Structured Data?. In *WikidataCon 2021*, online, October 29–31, 2019. Retrieved May 20, 2022 from https://commons.wikimedia.org/wiki/File:Anasuya_Sengupta_Whose_Knowledge_WikidataCon_2021_Slides.pdf

[46] Nanna Bonde Thylstrup. 2022. The Ethics and Politics of Data Sets in the Age of Machine Learning: Deleting Traces and Encountering Remains. *Media, Culture & Society* 01634437211060226. doi: 10.1177/01634437211060226.

[47] Eve Tuck and K. Wayne Yang. 2012. Decolonization Is Not a Metaphor. *Decolonization: Indigeneity, Education & Society* 1(1).

[48] Magdalena Tyżlik-Carver. 2020. Fermenting Data. Retrieved May 19, 2022 (https://fermentingdata.org).

[49] Magdalena Tyżlik-Carver. 2021. Curating Data: Infrastructures of Control and Affect … and Possible Beyonds. *Stages* (9).

[50] Magdalena Tyżlik-Carver. 2022 (forthcoming). Curating/Fermenting Data: speculative practices for curatorial futures. *Companion to Curatorial Futures*. edited by B. el Baroni and B. Crone. Edinburgh University Press

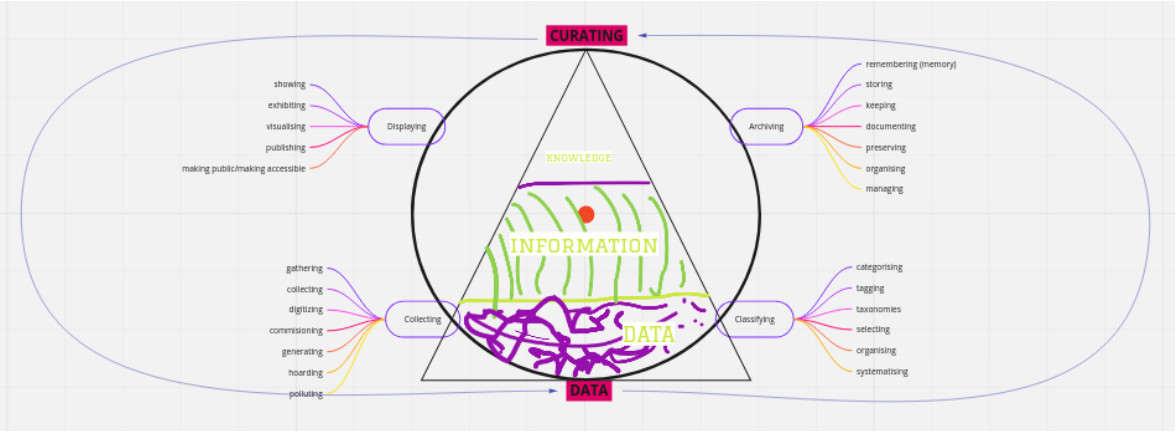
[51] Gregory Wiedeman. 2019. The Historical Hazards of Finding Aids. *University Libraries Faculty Scholarship* 124. Retrieved May 20, 2022 https://scholarsarchive.library.albany.edu/ulib_fac_scholar/124

[52] Milan Zeleny. 1987. Management Support Systems: Towards Integrated Knowledge Management. *Human Systems Management* 7, no. 1: 59–70.

[53] Shoshana Zuboff. 2015. Big Other: Surveillance Capitalism and the Prospects of an Information Civilization. *Journal of Information Technology; London* 30(1):75–89. doi: 10.1057/jit.2015.5.

# A APPENDICES

## A.1 Curating Data Diagram




Working Model for Curating Data prepared by M. Tyżlik-Carver for the course Curating Data in the Critical Data Studies program at Aarhus University. The diagram maps activities that are part of a typical data lifecycle into four main categories: collecting, classifying, archiving, displaying; which are then organised into subcategories. All categories and subcategories are not final and they are open to definitions, remodelling and deconstructions. This model is then projected onto the 'Data, Information, Knowledge' pyramid, a hierarchical organisation model proposed in the 1980s by information science scholars [1, 2, 52]. This is to further critically consider the organisation of knowledge models and standards and bring it into the context of curating data activities – to ask what assumptions are present at the very start of doing knowledge work.