

## RESEARCH PAPER

# Alter-Value in Data Reuse: Non-Designated Communities and Creative Processes

Guillaume Boutard

Université de Montréal – École de bibliothéconomie et des sciences de l'information  
Centre for Interdisciplinary Research in Music Media and Technology (CIRMMT), CA  
[guillaume.boutard@umontreal.ca](mailto:guillaume.boutard@umontreal.ca)

---

This paper builds on the investigation of data reuse in creative processes to discuss 'epistemic pluralism' and data 'alter-value' in research data management. Focussing on a specific non-designated community, we conducted semi-structured interviews with five artists in relation to five works. Data reuse is a critical component of all these works. The qualitative content analysis brings to light agonistic-antagonistic practices in data reuse and shows multiple deconstructions of the notion of data value as it is portrayed in the data reuse literature. Finally, the paper brings to light the benefits of including such practices in the conceptualization of data curation.

---

**Keywords:** Data Reuse; Data Value; Non-Designated Communities; New Media Arts; Sound Art; Epistemic Pluralism; Alter-Value; Creative Process

---

## Introduction

Data reuse has been an important domain of investigation as well as a critical topic for research institutions; conjointly, human and technical capabilities for data sharing have greatly increased during the last twenty years (see Bishop & Kuula-Luumi, 2017 for a discussion in the context of the UK). As Johnston (2020) states, "data reuse is a major focus for institutional research groups and their funders and it's easy to see why".

While the definition of data reuse is a matter of discussion (see van de Sandt, Dallmeier-Tiessen, Lavasa, & Petras, 2019), broadly speaking, data reuse relates to the "[...] research in which some or all of the data analyzed were collected by others besides the reuser or members of his/her research team" (Curty, Crowston, Specht, Grant, & Dalton, 2017). Such definitions of data reuse are often based on the assumption that the reuser is part of a specific designated community (as defined by the Open Archival Information System, 2012), usually, a part of the scientific community. Designated communities are evolving together with the data lifecycle (Baker, Duerr, & Parsons, 2015) in relation to monitoring activities (Consultative Committee for Space Data Systems, 2012, p. 4–19). A significant goal of data curation is then to go "from data that are invisible to researchers other than its creators to collections easily findable by other researchers" (Kethers, Treloar, & Wu, 2017, p. 2).

The notion of designated community has been debated in the literature. While "[...] a key premise of the Open Archival Information System reference model for the curation and preservation of digital objects is identifying designated communities to determine whether the context information needed to support reuse is being provided (Consultative Committee for Space Data Systems, 2012)" (Faniel, Frank, & Yakel, 2019, p. 1275), Bettavia (2016) argued that "[...] this term is the source of considerable power for a digital preservation repository in a way that privileges the institutions more so than the publics and audiences they serve".

In the specific context of data reuse, profiles of reusers have been discussed from a broader perspective to include 'non-professional' reusers, especially in relation to Open Government Data (OGD), but also in relation to Natural Science Data (e.g. Baker et al., 2015). Abella, Ortiz-de-Urbina-Criado, and De-Pablos-Heredero (2019) identified multiple open data reusers categories, considering that these categories "[...] would help open data portal managers to define more efficient and accurate promotion policies" (p. 299).

They proposed to distinguish between four types of end users: a) social users; b) citizens; c) professional users; d) academic users. Baker, Duerr and Parsons (2015) used the term ‘audience’ to encompass “[...] all groups and communities having interest in data products” (p. 113), further stressing that, “some audiences are not familiar with the sciences; they have a cultural knowledge base that differs significantly from that of a scientific community” (p. 113). The term ‘audience’, while inclusive, promotes a passive, consumerist view of data interaction. As such, an audience is not meant to produce anything that could be part of the data lifecycle, worthy of further data curation activities. The notion of non-designated community, that is still to be defined in this paper, builds upon the idea that these ‘audiences’ may be active members of the data reuse landscape. In the context of this study, we are interested in a very specific kind of reusers, who fail to be represented in previously described categorization: the artistic community, and more specifically, new media artists, sound artists and composers. We will thus discuss the specific relevance of this particular non-designated community, and its impact on data ontologies and data reuse epistemologies.

According to Yoon (2017), “while different disciplines have various cultures and practices relating to data, existing research presents some common challenges in data reuse across disciplines. The major challenge is transferring information regarding the contexts of data”. For Barry, Born and Weszkalnys (2008), “a commitment to a discipline is a way of ensuring that certain disciplinary methods and concepts are used rigorously and that undisciplined and undisciplinary objects, methods and concepts are ruled out. By contrast, ideas of interdisciplinarity and transdisciplinarity imply a variety of boundary transgressions, in which the disciplinary and disciplining rules, trainings and subjectivities given by existing knowledge corpuses are put aside or superseded” (p. 20–21). They defined three modes of interdisciplinarity, namely, 1) integrative-synthesis, 2) subordination-service, and 3) agonistic-antagonistic. The third mode “[...] stems from a commitment or desire to contest or transcend the given epistemological and ontological foundations of historical disciplines [...]” (Born, 2010, p. 211). Ontologically speaking, Massumi (2018)’s post-capitalist questioning of (economic) value and advocacy for alternative conceptions of value, notably as *alter-value* in relation to creativity, provides us with a relevant framework for discussing data ontologies through the prism of artistic practice.

The goal of this study is thus to investigate how data reuse embedded within creative processes questions research data management epistemologies and ontologies and the benefits we may gain from their inclusion in the conceptualization of data curation.

## From signal to data

The relevance of such a non-designated community and its practices in relation to the question of research data management and data curation is not arbitrary. It historically builds upon the intimate relation between creative processes, on the one hand, and data and information, on the other hand. Gottschalk (2016), reviewing contemporary music practice since the 1970s (as such a continuation of the work of Nyman in his 1974 book), considers that “sound is a type of information and can be treated as such, either spontaneously or as a document” (p. 155). In this circular definition, sound is, at the same time, signal and information. The questioning of boundaries is indeed a recurrent topic of artistic inquiry, and “[...] for decades many artists have disputed the notion that data or signals are not *already* information, and have explored varied practices of abstracting, manipulating, exposing, visualizing, sonifying, relocating or embodying data, in order to make messages into information, to make them meaningful in alternative ways” (Cook, 2016, p. 15). The question of visualisation, sonification, audification, and musification of data has a large body of literature. From a didactic perspective, for example, Paté et al. (2017) comment: “auditory display can complement visual representations in order to better interpret scientific data” (p. 2143). While there are significant forms of visualizing, sonifying, and embodying data in the works created by our participants, exploring utility versus creativity is not the scope of our research. In other words, we are not so much interested in the reverse process of data to signal but rather in the specific processes of data reuse, that is to say the ‘alternative ways’ alluded to by Cook (2016).

Since the 1960s, ‘earth’ and bio-signals have been regularly incorporated in artistic practices (see Mumma, 1974). Composer Gordon Mumma used P-wave and S-wave patterns generated by earthquakes and underground nuclear explosions during the early 1960s, as a structural basis for his series of *MOGRAPHS* – compositions for piano solo or various combinations of pianos. Charles Dodge’s famous 1970 piece *Earth’s Magnetic Field* used the Kp indices for the year 1961 as a basis for composition (Mumma, 1974, p. 332). According to Dodge (see Thieberger & Dodge, 1995), “the geophysicists at Goddard [Institute for Space Studies] had a way of recording the effects of the radiation of the sun on the magnetic field that resembled, in its notation, music” (p. 18). Alvin Lucier’s 1965 piece *Music for Solo Performer* is usually considered the first piece using ‘brainwaves’, followed by people such as Richard Teitelbaum and David Rosenboom (Miranda &

Brouse, 2005). Straebel and Thoben (2014) remind us that “Alvin Lucier’s Music for Solo Performer for enormously amplified brain waves and percussion (1965) is often considered an early example of artistic sonification” (p. 17). Other examples of biosignal use in composition and performance include composers of various generations such as Pauline Oliveros (see Mumma, 1974, p. 333) or Atau Tanaka (see Ortiz, 2012), among others. In 1985, composer Gérard Grisey discovered the signal of pulsars, thanks to astrophysicist Joe Silk (Grisey, 2019), which he used for his piece *Le Noir de l’Étoile* (1989–1990) as a live feed during performances.

From this history, which we only touched upon, Kahn (2013) concludes, “by the 1960s, composers and artists began considering electromagnetism per se as raw material for their craft, with Alvin Lucier working with what he called *natural electromagnetic sounds*, stretching from brainwaves to outer space; and by 1975, the composer Gordon Mumma could identify what he called an *astro-bio-geo-physical* trend within the ranks of live electronic music, tied to a plenitude of signals culled from scientific investigations of natural phenomena. It is an important distinction: electronic and experimental music did not merely use scientific signals; such music was already conducive to them” (p. 8). If by the 1960s, electronic and experimental music was conducive to the scientific signal, contemporary artistic practice should, arguably, be conducive to (research) data in the information age.

In 1985, French philosopher Jean-François Lyotard and Thierry Chaput curated the seminal, much discussed, exhibition *Les Immatériaux*, which has “[...] often been seen as a celebration of information technology [...]” (Birnbau & Wallenstein, 2015, p. 245), and which was based on a previous report by Lyotard, “[...] considered to be a response to another report by Simon Nora and Alain Minc, in the 1970s, which proposed the ‘computerisation of society’” (Hui & Broeckmann, 2015, p. 9). New Media art in the exhibition included works such as Rolf Gehlhaar’s *Son=Espace* that “[...] comprised a space which viewers walked through, with sensors that picked up the movements of the audience and turned them into sounds by means of an elaborate computerized system devised by the artist” (Gallo, 2015, p. 129). Gallo (2015) further comments: “it is difficult to say whether the philosopher of postmodernism would have appreciated a new kind of artwork, made with data flows on the web – one of the newest forms of interaction” (p. 131), which she exemplifies with Carlo Zanni’s work, including his practice of *Data Cinema*.

Furthermore, for new media theorist Manovich (2000), “database becomes the centre of the creative process in the computer age” (p. 182). Vesna (2007) stresses that “artists have long recognized the conceptual and aesthetic power of databases [...]” (p. xi). She further argues that there is a need “[...] to show not only how practicing artists think in relation to databases but also to raise the awareness of a wider audience about the importance of considering how our social data are being organized, categorized, stored, and retrieved” (p. xiv).

Beyond social data and beneath the conflation of signal, data, and information, our focus is the investigation of potential agonistic-antagonistic constructs in creative processes in discussion with data reuse and data value literature.

## Methodology

### *Methodological framework*

This study is grounded in five semi-structured interviews with artists from varied domains of practice. Participants were selected according to purposive sampling to maximize the range of practices and types of data used: Research Data, Open Government Data (OGD) and Public Sector Information (PSI). A pilot interview was conducted with one potential participant to establish the credibility (internal validity) of the protocol. The interview protocol was divided in three parts: a) the description of the work and its relation to the domain of research that produced the data; b) the process of working with the data and its significance; c) the relation between the work and the data.

The data analysis is based on two methodological frameworks: 1) Qualitative Content Analysis (Drisko & Maschi, 2016); and 2) Interpretative Phenomenological Analysis (Smith, Flowers, & Larkin, 2009). This paper presents the outcome of the QCA.

Drisko and Maschi (2016) state that “while qualitative content analyses may involve interpretations of latent content and meaning, broad critical analyses are not commonly their main research purpose” (p. 82). Latent content is not targeted by the analysis in this study. Still, as Strauss and Corbin (1998) put it, “[...] whenever we conceptualize data or develop hypotheses, we are interpreting to some degree”. The analysis used both inductive and deductive approaches, using categories emerging from the data through open coding and putting them in relation to categories from the data reuse literature.

In this study, we are not developing theory in the sense of Grounded Theory (see 1998), and therefore not trying to reach theoretical saturation, hence the small sample. The goal of purposive sampling (rather

than theoretical sampling) is used to “raise awareness, provide new perspectives, or provide descriptions of events, beliefs, and actions” (Drisko & Maschi, 2016, p. 99), rather than portraying a group which would require the transferability of the analysis to the group.

In a last phase, the outcomes of the analysis are discussed in relation to notions of alter-value and epistemic pluralism.

### **Participants**

The five participants relate to five artworks. The reason for limiting the data collection to one work per artist (and one participant per work in the case of a collective or collaboration) was primarily informed by the Interpretative Phenomenological Analysis with the goal to investigate the subjective experience of participants in relation to datasets, which is out of the scope of this paper. Interviews lasted between 25 and 90 minutes and were conducted, face to face or online, between August 2018 and August 2019. In the context of artistic production and the investigation of specific creative processes, the discussion of the work process of participants cannot be anonymized. This was documented in the ethics certificate which was provided and signed by participants before the data collection. Participants are in alphabetic order: Alice Guerlot-Kourouklis (AGK) from IAKERI collective (with Jimena Royo-Letelier and Aneymone Wilhelm) for their work *Invisible Walls*; Jasmine Guffond (JG) for her work *Anywhere All The Time: A Permanent Soundtrack to Your Life*; Elizabeth Hoffman (EH) for her work *Retu(r)nings*; Herman Kolgen (HK) for his work *SEISMIK*; and Sissel Marie Tonn (SMT) for her work *The Intimate Earthquake Archive*. Interviews with AGK and HK were conducted in French and all citations from them in this paper are our translation. The interview with EH was complemented by a couple of emails, following the interview and at her initiative, which are included in the analysis.

The IAKERI collective primarily reuses open government data, focussing on gender inequality statistics. The data, in terms of sound production, provides a ratio of inequality used as a parameter for multiple elements of signal processing applied to an electroacoustic medium previously composed by AGK. Both HK and SMT use seismic data in very different contexts, further described in this paper. SMT uses specific datasets related to seismic activity generated by human activity (gas extraction) in the Netherlands. After several stages of processing, the data is then used to produce an embodied experience for participants wearing ‘haptic cloaks’. EH uses the physical access logs of New York University’s Elmer Holmes Bobst library as material for a spatialized electroacoustic composition (based on an idea from J. Martin Daughtry and Kent Underwood) that is diffused within Bobst Library’s atrium. JG is using “the SSID names, the MAC addresses and then the frequency of the Wifi” (JG) as data related to surrounding networks for real-time sonification during listening walks.

## **Analysis**

### **Data interactions**

#### **Availability**

In the case of SMT, the data, which is publicly accessible, came from one institution, The Royal Netherlands Meteorological Institute (KNMI). SMT was in direct contact with data scientists at KNMI for some form of collaboration: “[...] with the help of the data scientist at the KNMI, we got a python-based program where you can extract those lines of data and then translate it into a sound file, which we worked with” (SMT). The selection process was based on the intensity of the man-made earthquakes (maximum of 3.6 on the Richter scale according to SMT), while SMT also mentioned being interested in working with the smaller ones in relation to “[...] the threshold of what we can experience when it comes to these kinds of climate-related or environmentally-related changes [...]” (SMT). Signal produced by the python program was then processed with a piece of software designed by collaborator Jonathan Reus in SuperCollider (a platform for audio synthesis, etc.). The creative process encompasses thus a multitude of levels of manipulation from the original data. For SMT, “it wasn’t so much about simulating the earthquake, but rather creating a compositional medium which could be thought of as an archive”. SMT also relates the experience of data in this context to a political act of deep listening (a reference to a concept by composer Pauline Oliveros, see Oliveros, 2005) “that manifests by a gesture of paying attention to this particular subject matter”.

For HK, the relation to data went, in the first place, through attempts at reaching out to scientists around the world. A process which was proven unsuccessful at the beginning: “I had my list of questions [...] and I had no feedback from any scientist” (HK, trans.). The questions were related to availability and accessibility of data. Later, HK developed a collaboration with two researchers at Grand Accélérateur National d’Ions Lourds (GANIL) in France for another project: ISOTOPP. According to his statements during the interview, HK used, in *SEISMIK*, live data feeds selected before and during performance through an interface developed

for him by an assistant collaborator. The interface of HK, that is to say the instrument, allows him to select a feed during performance according to his criteria (intensity was mentioned several times). The data is used as a control process for multiple media elements of the work through a matrix system. Grisey (2019), discussing his work *Le Noir de l'Étoile* recalls: "I was seduced by the sounds of the Vela Pulsar and immediately, I wondered, like Picasso picking up an old bicycle: 'what could I do with that?'" (p. 166, our translation). From a data reuse perspective, the notion of real-time in HK's work seems to parallel Grisey's use of the pulsar's signal in the sense that while there is no specific ontological alter-value attached to the data/signal (as per our analysis), the real-time aspect affects the nature of the performance, notably in relation to data availability (as highlighted by his set of questions to scientists). In that respect, the work of HK converges with the conceptualization of quality in data reuse literature, where availability is defined as a "[...] degree of convenience for users to obtain data and related information [...]" (Cai & Zhu, 2015).

In the work of JG, which also builds on real-time processing of data and where the element of sonification is direct, the notion of real-time is discussed in terms of both the technological domain of production, that is to say 'surveillance technologies', and the artistic medium itself: "[...] contemporary digital surveillance, [...] is decentralized and in a constant state of flux, and sound is similarly in a constant state of flux, so I feel like it is an appropriate medium in which to tangibly experience these networks that are for the most part intangible and yet constitute our everyday experience" (JG). JG applied similar principles in a few other works, for example her sonification of browser cookies, implemented as a browser plugin named *Listening back*, or her 2019 installation *Sonic Profiles* using social media' metadata (see Guffond & González Fuster, 2020). In comparison to research data management, the availability of data – emphasized by the real-time aspect of the performance (see Guffond, 2015) – is thus, in JG's work, the focus of a process for building awareness and knowledge about a sociopolitical environment rather than a dimension of quality.

### Curation

For the IAKERI Collective, as explained by AGK, the process of data collection was longitudinal (and still ongoing at the time of the interview, which happened during the exhibition of the piece at Eastern Bloc in Montreal, with the possibility to include new datasets), multi-level and multi-sources, purposive with elements of convenience sampling. With regards to this last point, they started at the French Institut national de la statistique et des études économiques (Insee) but integrated reports which emerged at that time on the topic of gender inequalities in the cultural field, identifying institutions who commissioned these surveys. They further included data from outside of France, for example data from Québec, as well as data provided by international organizations such as UN/UNESCO, and OECD. While no appraisal was mentioned in the interview, selection was made according to the diversity of domains represented in the statistics. Beside the modalities of 're-sampling' two other relevant points emerge from this analysis: 1) Data integration and the coherence of the juxtaposition of heterogeneous data, "we are still, at his moment, studying the relevance of these coexistences" (AGK, trans.); 2) Data harmonization (and re-purposing, somehow similarly to a secondary data analysis) as discrepancies (between gender), when it was not directly provided, and in terms of range (between 0 and 1) so that it could feed into the audio signal processing, "[...] with some code provided by Jimena [Royo-Letelier] [...]" (AGK, trans.), who is also a mathematician. Both of these elements have data-related ontological relevance. The first pertains to the relational value of data sets, potentially beyond their initial domain of relevance, and, as discussed later, beyond their initial epistemic framework. The second point relates to the question of data processing, which has gained research attention in term of contextualization of data with the recognized importance of research software, as Ramapriyan, Moses and Duerr (2012) state, in the context of Earth System Science, "instrument calibration data are sometimes acquired and applied separately from science data production, and it is appropriate to archive the data and source code that has led to the calibrated science data products". While we will discuss further the relation between data and process in relation to the work of EH, the most important outcome of the analysis of this work lies at the junction of these two elements, that is to say, the active production of a new space of signification by the combination of heterogeneous datasets and their normalization to fit this purpose. From a process point of view it seems to mirror parts of Brunner (2016)'s anarchival methodology proposal in relation to data and creative processes: "take all the data you might consider relevant. Add more to it by subtracting. Make sure that subtraction is not a diminishing or reductive procedure but an amplification of intensity. Co-inhabit a space and time with the data [...]" (p. 77). While we will not discuss the notion of anarchival (see Duguet, 2019) in the scope of this paper, we may relate IAKERI's process to our domain of interest.

The Consortia Advancing Standards in Research Administration Information's (CASRAI, 2020) defines data curation as "a managed process, throughout the data lifecycle, by which data & data collections are cleansed,

documented, standardized, formatted and inter-related". The word curation in the cultural domain brings another perspective, relevant to our discussion, especially in relation to new media art curation. Paul (2008) argues that "the role of a new media curator is increasingly less that of 'caretaker' of objects (as the original meaning of the word 'curator' suggests) and more that of a mediator and interpreter or even producer" (p. 65). The IAKERI collective produces an act of data curation which seems to join the notion of curation as expressed in data science as well as in cultural studies.

### Process

The question of software preservation in relation to research data management reuse has been discussed recently in the literature. Davenport, Grant and Jones (2020) argue that, "increasingly, research across disciplines depends upon software, for experimental control or instrumentation, simulating models or analysis and turning numbers into figures. It is vital that bespoke software is published alongside the journal article and the data it supports. While it doesn't ensure that code is correct, it does enable the reproducibility of analysis and experimental workflows to be checked, and validated against correct or 'expected' behaviour" (p. 2). The creative process of EH deviates from this specific causal dependency between data and process which mirrors the ontological status of data in the data-information-knowledge (DIK) hierarchy, adequately portrayed by the notion of representation information in the OAI (2012): "Data interpreted using its Representation Information yields Information" (p. 2–4). The work of EH critically interferes with this view, firstly by discussing productivity in terms of potentialities (her emphasis in the following quotes): "it is neither the algorithmic structure that I've created *NOR* the content (the data) that is particularly compelling experientially. In other words, there is no guarantee that something put through the structure will elucidate the structure of the algorithm in an elegant or expressive or informative way, nor that the algorithm will necessarily make every dataset artistically spectacular or intriguing. But, when combined, the algorithm and the data \*have the potential\* to produce some very compelling artistic results that would almost certainly not otherwise have been conjectured" (EH). And secondly, it reverses the causal dependency between data and process, and thus the ontological status of data in research data management: "I cannot predict *WHICH* dataset will yield an amazing artistic result and which may be simply a good result. *so/but* – the more datasets that I put through the algorithm the more I learn *ABOUT* the structure that I created" (EH).

### **Epistemic Pluralism and Non-Designated Communities**

The modalities of collaboration are a first axis of discussion which has epistemological relevance. When Charles Dodge released *Earth's Magnetic Field* on Nonesuch records, Bruce R. Boller, Carl Frederick and Stephen G. Ungar were acknowledged as scientific associates on the front cover. This integrative-synthesis and/or subordination-service (Barry et al., 2008) paradigm of disciplinary collaboration is also the one we can trace in our data analysis of HK's work, whether it is implemented or not during the artistic project (see above). The framework appears to be different, in particular, in the work of SMT.

Edwards (2009), retracing an encounter between artist Ryoji Ikeda (famous for his immersive datascares, see Weil, 2012, his work series *Datamatics*, see Rhee, 2017, etc.) and mathematician Benedict Gross, states that: "Benedict Gross is ultimately a mathematician and Ryoji Ikeda an artist and their truths sometimes find incompatible expression. But it is this incompatibility that outlines the experiment" (p. 13). This very generic statement may be put into perspective with the work of SMT.

*The Intimate Earthquake Archive* was presented, notably, in 2018 in Marfa, TX during an exhibition curated by Timothy Morton and Laura Copelin, called *Hyperobjects*, in reference to a concept by Morton (see Morton, 2013), which relates to ecological awareness. During the interview, SMT made explicit the specific interactions she had with scientists: "[...] I mentioned [to the scientists] that some of these people were waking up before the earthquakes, which I thought was a fascinating story, but the scientists got kind of angry, [...]. Even one person that I emailed said that he did not want to talk to me because this was anecdotal and he was doing real science and what I was doing was not helping the cause". From this perspective, SMT's interest in the perception of environmental change, expressed through a theoretically-rich body of works (see also an interview with SMT and collaborator Reus in Morris, 2020), can be related to the recontextualisation of the development of a general public interest in Sea Ice Data at the National Snow and Ice Data Center (NSIDC) by Baker, Duerr and Parsons (2015) in the mid-2000s, and the issues it raised: "[...] upon investigating a number of reports that incorrectly cited the data or analysis, it became apparent that with the great public interest in sea ice there was some misinformation proliferating on the web through social media" (p. 120). While understanding the broader sociopolitical context of data reuse, such as the one portrayed by Baker, Duerr and Parsons (2015) from a critical perspective, SMT is also interested in "what we consider legitimate knowledge" (SMT).

This position also reflects debates about dominant epistemologies in research data management as well as information science. Mauthner and Parry (2009), discussing the rise of qualitative data sharing debates in the UK in the mid-1990s, argued that “[...] contemporary data preservation and sharing debates, discourses, policies and practices are embedded within these foundational understandings of knowledge and its production, [...]” (p. 294). Similarly, Feldman and Shaw (2019) relate current practice to neopositivism: “these demands [to archive and share qualitative data] have gained currency in the era of big data in which large amounts of information, coupled with new technologies for analyses, make these data available for sharing and reuse. Advocates, who generally undertake research in the neopositivist tradition, contend that data sharing and reuse can improve transparency, avoid unnecessary replication, and contribute to theory building (Arzberger et al., 2004)” (p. 700). This question of qualitative research data sharing and reuse has received research attention (e.g. Mason, 2007; Slavnic, 2013; Kirilova & Karcher, 2017; McAllister, 2018). However, linking this question to agonistic-antagonistic data reuse artistic practice, such as the one from SMT, requires to build on another argument by Mauthner and Parry (2009): “[foundationalism] claims epistemic supremacy for itself while denying the epistemic status and legitimacy of other perspectives on knowledge” (p. 294). Their advocacy for epistemic pluralism parallels the questions raised by SMT in her creative practice.

As Cragin et al. (2010) put it, “in the discourse on data sharing, risks of data misuse (and other barriers to sharing) have been prominent themes [...]” (p. 4033). They further comment on the results of their study: “Misuse incidents experienced by scientists in this study influenced their views on the appeal of data sharing, decreasing their willingness to share and increasing their cynicism in data-sharing initiatives, but they also had a real impact on their behaviour” (p. 4033). Beyond the question of intellectual property – as Borgman, Scharnhorst and Golshan (2019) argue, “researchers often maintain a sense of ownership over their data, regardless of legal status” (p. 890) – the case of agonistic-antagonistic creative practice poses the question of which epistemic framework validates the idea of misuse. Using Baker, Duerr and Parsons’ (2015) characterization of data products, we may then propose a definition of non-designated communities, not just as a potential audience for data, but as those active communities whose epistemic perspective, manifested in the data products they create, is not yet being supported by data curation frameworks. We argue that this should lead to question the role assigned to non-designated communities and the place of epistemic pluralism in data curation.

### ***Alter-Value and Creative Process***

Harvey (2007), discussing selection and appraisal, proposes a few typical questions that pertain to that stage: “why are the materials worth keeping? What gives them the value that warrants the trouble of preserving them? Is that value associated with evidence, information, artistic or aesthetic factors, significant innovation, historic or cultural association, what a user can make the material do or do with the material, culturally significant characteristics?” We would argue that defining data value is not limited to a discussion on context.

Pasquetto, Randles and Borgman (2017) consider that “few data policy documents define ‘data’ explicitly” (p. 3). Johnston et al. (2018) provide us with a basic and pragmatic definition: “facts, measurements, recordings, records, or observations about the world collected by scientists and others, with a minimum of contextual interpretation” (p. 5). While the relation between data and fact is a matter of discussion (e.g. Chow-White & Sandy Green, 2013), this definition is consistent with the data-information-knowledge (DIK) hierarchy, which has often been commented, discussed and critiqued in information science (see Frické, 2008; Ma, 2012). Nielsen and Hjørland (2014), specifically, discussed it in the context of data curation. Building on Kaase (2001), they state that “data are always recorded on the basis of some interests, perspectives, technologies and situated practices that determine their meaning and usefulness in different contexts” (Nielsen & Hjørland, 2014, p. 225). The question of data recording parallels the question of reuse. Creative processes are situated practices which have demonstrated in our analysis their potential for producing alternative data ontologies. These new ontologies question the frameworks of definition for data value. Data value has been discussed along various dimensions relating to the research lifecycle. According to Johnston (2017), “[...] we are experiencing a dramatic shift in how data are reused, not only to ‘do new science,’ but also because data reuse may increase a paper’s potential research impact, provide greater transparency to the results, and in some cases, can even make or break an individual’s career” (p. 14). In terms of research cost, “archivists (and scientists) also frequently made a distinction between observational and experimental data, pointing out that observational data might have long-term value if they were irreplaceable or very costly to collect [...]” (Akmon, Zimmerman, Daniels, & Hedstrom, 2011, p. 332). Curty et al. (2017) note that “secondary data analysis has been suggested as a mechanism to lessen expenses related to data collection [...]” (p. 3), it shortens the research project and has potential for generating new results (e.g. Cragin et al., 2010, p. 4029).

These elements speak directly to the critique of value elaborated by Massumi (2018) from an economy standpoint, where “the first task of revaluation of value is to *uncouple value from quantification*” (p. 4).

From a broader perspective, data quality has been discussed beyond the characteristics of data *per se*. Cai and Zhu (2015) proposed five general categories of data quality: availability; usability; reliability; relevance; and presentation quality. In terms of trust factors, Yoon and Lee (2019) identified four categories, including data quality, namely: data producers, scholarly community, data quality, and intermediary (i.e. data repository). Maybe unsurprisingly, if we consider previous discussions about our participant’s works, when relevant, no reference to data quality *per se* nor to trust emerged from the interviews, nor to trust. On the other hand, in our analysis, many elements from these frameworks appear to be challenged by these practices from which emerge new data ontologies.

JG’s emphasis on availability, as argued previously, may be seen as a polarity (or valence, to use core affect terminology) reversal as compared to standard data ontologies. On a similar note, the same artist brings an important twist to the notion of accuracy (one of Cai and Zhu, 2015, defining elements of reliability), turning the imperfections of GPS tracking into a presence for the participants: “because this was all about how trackable we are via our everyday devices, so I had this idea that if someone was following you maybe you would hear their footsteps, or feel their breath on the back of your neck” (JG). EH’s critical conceptualization of data versus process is also disrupting a few elements of data quality frameworks, such as Yoon and Lee’s (2019) notion of comprehensiveness – requiring “[...] all aspects of the data to be understandable” – and Cai and Zhu’s (2015) *completeness* and *readability*.

Certain elements also bring questions beyond established categories. In the domain of archival science, Lemay and Klein (2012), have discussed the increasing interest of archivists for the emotional value of archives beyond their primary use as evidence. Klein (2019), who focuses on archives’ reuse in artistic practice, states: “the installations of [Bertrand] Carrière and [Christian] Boltanski highlight another aspect of the archives which is likely to generate an emotional response, namely, accumulation. The mass of documents contributes to the materiality of archives since, as conveyed by their definition, they are ‘a set of documents’. [...] However, what Carrière and Boltanski reveal is that behind this mass, [...] men and women lived” (p. 193, our translation). Away from Johnston et al. (2018)’s definition of data, IAKERI collective’s willingness to convey the idea of the ‘brutality of data’ (as they describe it, see Guerlot-Kourouklis, Royo-Letelier, & Wilhelm, 2019) seems to parallel this idea of accumulation and emotion as constitutive parts of the object itself (strictly from a curation perspective rather than an interpretation of the creative process or an aesthetical comparison to the works discussed by Klein), in the context of big data: “[...] when we take a look at all these data on the question of inequalities, we realize that there is something systemic about it, something that is immense, and it is something that we would like people to experience with sound, and that is why we chose the question of sound distortion” (AGK, trans.). To a certain extent, we find similar elements in SMT’s emphasis on environmental awareness and ecological grief.

These perspectives combine to the aforementioned dimensions of value to propose us with leads into the definition of data alter-value(s).

## Conclusion

Massumi (2018) argued in his manifest on economic value that, “to take back value is to revalue value, beyond normativity and standard judgment” (p. 19). Normativity in research data management has been discussed in the literature in relation to, notably, foundationalism and neopositivism. In the context of data reuse, we have discussed how non-designated communities in the artistic domain, through agonistic-antagonistic practices, may propose paradigmatic shifts – epistemologically and ontologically speaking – from research data management’s ‘normativity’. The variety of ontological deconstructions emerging from a limited set of case studies, hints at the benefits the data management and curation community may gain from discussing data reuse and data value in relation to creative processes. Specifically, we argue that including these practices in data curation – building on the polysemy of the word curation – may be a relevant way to provide a significant form of acknowledgment of epistemic pluralism in research data management.

## Competing Interests

The author has no competing interests to declare.

## References

**Abella, A, Ortiz-de-Urbina-Criado, M and De-Pablos-Heredero, C.** 2019. The process of open data publication and reuse. *Journal of the Association for Information Science and Technology*, 70(3): 296–300. DOI: <https://doi.org/10.1002/asi.24116>



- Akmon, D, Zimmerman, A, Daniels, M and Hedstrom, M.** 2011. The application of archival concepts to a data-intensive environment: working with scientists to understand data management and preservation needs. *Archival Science*, 11(3–4): 329–348. DOI: <https://doi.org/10.1007/s10502-011-9151-4>
- Arzberger, P, Schroeder, P, Beaulieu, A, Bowker, G, Casey, K, Laaksonen, L, Wouters, P, et al.** 2004. Promoting Access to Public Research Data for Scientific, Economic, and Social Development. *Data Science Journal*, 3: 135–152. DOI: <https://doi.org/10.2481/dsj.3.135>
- Baker, KS, Duerr, RE and Parsons, MA.** 2015. Scientific Knowledge Mobilization: Coevolution of Data Products and Designated Communities. *International Journal of Digital Curation*, 10(2): 110–135. DOI: <https://doi.org/10.2218/ijdc.v10i2.346>
- Barry, A, Born, G and Weszkalnys, G.** 2008. Logics of interdisciplinarity. *Economy and Society*, 37(1): 20–49. DOI: <https://doi.org/10.1080/03085140701760841>
- Bettivia, RS.** 2016. The Power of Imaginary Users: Designated Communities in the OAI Reference Model. In *Proceedings of the 79th ASIS&T Annual Meeting*, 38:1–38:9. Silver Springs, MD, USA: American Society for Information Science. DOI: <https://doi.org/10.1002/pr2.2016.14505301038>
- Birnbaum, D and Wallenstein, S-O.** 2015. From Immaterials to Resistance: The Other Side of *Les Immatériaux*. In Hui, Y and Broeckmann, A (eds.), *30 Years After Les Immatériaux: Art, Science and Theory*, 245–267. Meson Press.
- Bishop, L and Kuula-Luumi, A.** 2017. Revisiting Qualitative Data Reuse: A Decade On. *SAGE Open*, 7(1). DOI: <https://doi.org/10.1177/2158244016685136>
- Borgman, CL, Scharnhorst, A and Golshan, MS.** 2019. Digital data archives as knowledge infrastructures: Mediating data sharing and reuse. *Journal of the Association for Information Science and Technology*, 70(8): 888–904. DOI: <https://doi.org/10.1002/asi.24172>
- Born, G.** 2010. For a Relational Musicology: Music and Interdisciplinarity, Beyond the Practice Turn. *Journal of the Royal Musical Association*, 135(2): 205–243. DOI: <https://doi.org/10.1080/02690403.2010.506265>
- Brunner, C.** 2016. (Digital) anarchival platforming. In *The Go-To How-To Book of Anarchiving*, 75–79. Montreal, QC: The Senselab.
- Cai, L and Zhu, Y.** 2015. The Challenges of Data Quality and Data Quality Assessment in the Big Data Era. *Data Science Journal*, 14(0): 2. DOI: <https://doi.org/10.5334/dsj-2015-002>
- Chow-White, PA and Sandy Green, J.** 2013. Data Mining Difference in the Age of Big Data: Communication and the Social Shaping of Genome Technologies from 1998 to 2007. *International Journal of Communication*, 7(0): 28.
- Consortia Advancing Standards in Research Administration Information.** 2020. *Research Data Management Glossary*. Retrieved 2020-04-24, from <https://casrai.org/rdm-glossary/> (Library Catalog: casrai.org).
- Consultative Committee for Space Data Systems.** 2012. *Reference Model OAI (Tech. Rep. No. Magenta Book)*. Washington, DC: Consultative Committee for Space Data Systems.
- Cook, S.** 2016. *Information*. Cambridge, MA: MIT Press. (OCLC: 945730189).
- Cragin, MH, Palmer, CL, Carlson, JR and Witt, M.** 2010. Data sharing, small science and institutional repositories. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 368(1926): 4023–4038. DOI: <https://doi.org/10.1098/rsta.2010.0165>
- Curry, RG, Crowston, K, Specht, A, Grant, BW and Dalton, ED.** 2017. Attitudes and norms affecting scientists' data reuse. *PLOS ONE*, 12(12): e0189288. DOI: <https://doi.org/10.1371/journal.pone.0189288>
- Davenport, JH, Grant, J and Jones, CM.** 2020. Data Without Software Are Just Numbers. *Data Science Journal*, 19(1): 3. DOI: <https://doi.org/10.5334/dsj-2020-003>
- Drisko, JW and Maschi, T.** 2016. *Content analysis*. New York, NY: Oxford University Press. DOI: <https://doi.org/10.1093/acprof:oso/9780190215491.001.0001>
- Duguet, A-M.** 2019. The “anarchive” Series as a Challenge between Art and Information. A singular approach of media art history. In Grau, O, Hoth, J and Wandl-Vogt, E (eds.), *Digital art through the looking glass: new strategies for archiving, collecting and preserving in digital humanities*, 73–77. Donau: Edition Donau-Universität.
- Edwards, D.** 2009. DYAD. In Barral, X, Ikeda, R and Naphegyi, C (eds.), *V(\neq) L*. Paris: Éditions Xavier Barral.
- Faniel, IM, Frank, RD and Yakel, E.** 2019. Context from the data reuser's point of view. *Journal of Documentation*, 75(6): 1274–1297. DOI: <https://doi.org/10.1108/JD-08-2018-0133>
- Feldman, S and Shaw, L.** 2019. The Epistemological and Ethical Challenges of Archiving and Sharing Qualitative Data. *American Behavioral Scientist*, 63(6): 699–721. DOI: <https://doi.org/10.1177/0002764218796084>

- Frické, M.** 2008. The knowledge pyramid: a critique of the DIKW hierarchy. *Journal of Information Science*, 35(2): 131–142. DOI: <https://doi.org/10.1177/0165551508094050>
- Gallo, F.** 2015. Contemporary Art as “Immatériaux”: Yesterday and Today (M. Desmond & P. Polidori, Trans.). In Hui, Y and Broeckmann, A (eds.), *30 Years After Les Immatériaux: Art, Science and Theory*, 119–135. Meson Press.
- Gottschalk, J.** 2016. *Experimental Music Since 1970*. New York; London: Bloomsbury Academic.
- Grisey, G.** 2019. *Écrits ou l'invention de la musique spectrale* (2nd ed.; Lelong, G and Reby, A-M (eds.)). Paris: éditions MF.
- Guerlot-Kourouklis, A, Royo-Letelier, J and Wilhelm, A.** 2019. *Murs Invisibles*. Retrieved from [http://iakeri.fr/content/Murs\\_Invisibles\\_Collectif\\_Iakeri\\_2019.pdf](http://iakeri.fr/content/Murs_Invisibles_Collectif_Iakeri_2019.pdf).
- Guffond, J.** 2015. *Anywhere, All the Time, A Permanent Soundtrack to Your Life*. Retrieved 2020-02-27, from <http://www.jasmineguffond.com/?path=art/Anywhere+All+The+Time>.
- Guffond, J and González Fuster, G.** 2020. *Sounds Of Surveillance – The Wire*. Retrieved 2020-04-24, from <https://www.thewire.co.uk/audio/tracks/sounds-of-surveillance>.
- Harvey, R.** 2007. Appraisal and Selection. In Ross, S and Day, M (eds.), *DCC Digital Curation Manual* (1st ed.). Glasgow: HATII, University of Glasgow. Retrieved from <http://www.dcc.ac.uk/resource/curation-manual/chapters/appraisal-and-selection>.
- Hui, Y and Broeckmann, A.** 2015. Introduction. In Hui, Y and Broeckmann, A (eds.), *30 Years After Les Immatériaux: Art, Science and Theory*, 9–24. Meson Press.
- Johnston, LR.** 2017. Introduction to data curation. In Johnston, LR (Ed.), *Curating Research Data Volume One: Practical Strategies for Your Digital Repository*, 1: 1–30. Chicago, IL: Association of College and Research Libraries.
- Johnston, LR.** 2020, February. *How a network of data curators can unlock the tremendous reuse value of research data*. Retrieved 2020-02-21, from <http://www.oclc.org/blog/main/data-curators-network/>.
- Johnston, LR, Carlson, J, Hudson-Vitale, C, Imker, H, Kozlowski, W, Olendorf, R and Stewart, C.** 2018. How Important is Data Curation? Gaps and Opportunities for Academic Libraries. *Journal of Librarianship and Scholarly Communication*, 6(1): eP2198. DOI: <https://doi.org/10.7710/2162-3309.2198>
- Kaase, M.** 2001. Databases, core: political science and political behavior. In Smelser, NJ and Baltes, PB (eds.), *International encyclopedia of the social & behavioral sciences* (Vol. 5). Amsterdam: Elsevier. DOI: <https://doi.org/10.1016/B0-08-043076-7/00034-6>
- Kahn, D.** 2013. *Earth Sound Earth Signal: Energies and Earth Magnitude in the Arts*. Berkeley: University of California Press. DOI: <https://doi.org/10.1525/california/9780520257801.001.0001>
- Kethers, S, Treloar, A and Wu, M.** 2017. Building Tools to Facilitate Data Reuse. *International Journal of Digital Curation*, 11(2): 1–12. DOI: <https://doi.org/10.2218/ijdc.v11i2.409>
- Kirilova, D and Karcher, S.** 2017. Rethinking Data Sharing and Human Participant Protection in Social Science Research: Applications from the Qualitative Realm. *Data Science Journal*, 16(0): 43. DOI: <https://doi.org/10.5334/dsj-2017-043>
- Klein, A.** 2019. *Archive(s), mémoire, art: éléments pour une archivistique critique*. Québec, QC: Presses de l'Université Laval.
- Lemay, Y and Klein, A.** 2012. Les archives et l'émotion: un atelier d'exploration et d'échanges. *Archives*, 44(2): 91–109.
- Ma, L.** 2012, April. Meanings of information: The assumptions and research consequences of three foundational LIS theories. *Journal of the American Society for Information Science and Technology*, 63(4): 716–723. DOI: <https://doi.org/10.1002/asi.21711>
- Manovich, L.** 2000. Database as a genre of new media. *AI & SOCIETY*, 14(2): 176–183. DOI: <https://doi.org/10.1007/BF01205448>
- Mason, J.** 2007. 'Re-Using' Qualitative Data: On the Merits of an Investigative Epistemology. *Sociological Research Online*, 12(3): 39–42. DOI: <https://doi.org/10.5153/sro.1507>
- Massumi, B.** 2018. *99 theses on the revaluation of value: a postcapitalist manifesto*. Minneapolis, MN: University of Minnesota Press. DOI: <https://doi.org/10.5749/9781452958484>
- Mauthner, NS and Parry, O.** 2009. Qualitative data preservation and sharing in the social sciences: On whose philosophical terms? *Australian Journal of Social Issues*, 44(3): 291–307. DOI: <https://doi.org/10.1002/j.1839-4655.2009.tb00147.x>
- McAllister, JW.** 2018. Scientists' Reuse of Old Empirical Data: Epistemological Aspects. *Philosophy of Science*, 85(5): 755–766. DOI: <https://doi.org/10.1086/699695>

- Miranda, E** and **Brouse, A.** 2005. Toward direct brain-computer musical interfaces. In *Proceedings of the 2005 conference on New interfaces for musical expression*, 216–219. Vancouver, Canada.
- Morris, R.** 2020. *Intimate mapping – In conversation with Jonathon Reus & Sissel Marie Tonn*. Retrieved 2020-04-25, from <https://www.fiber-space.nl/news/intimate-mapping-in-conversation-with-jonathon-reus-sissel-marie-tonn/>.
- Morton, T.** 2013. *Hyperobjects: philosophy and ecology after the end of the world*. Minneapolis, MN: University of Minnesota Press.
- Mumma, G.** 1974. Live-Electronic Music. In Appleton, JH and Perera, RC (eds.), *The development and practice of electronic music*, 286–335. Englewood Cliffs, NJ: Prentice Hall.
- Nielsen, HJ** and **Hjørland, B.** 2014, March. Curating research data: the potential roles of libraries and information professionals. *Journal of Documentation*, 70(2): 221–240. DOI: <https://doi.org/10.1108/JD-03-2013-0034>
- Nyman, M.** 1974. *Experimental music: Cage and beyond*. New York: Schirmer Books.
- Oliveros, P.** 2005. *Deep listening: a composers's sound practice*. New York; Shanghai: iUniverse. (OCLC: 470106497).
- Ortiz, M.** 2012. A Brief History of Biosignal-Driven Art: From biofeedback to biophysical performance. *eContact!*, 14(2). Retrieved 2020-02-26, from [https://econtact.ca/14\\_2/ortiz\\_biofeedback.html](https://econtact.ca/14_2/ortiz_biofeedback.html).
- Pasquetto, IV, Randles, BM** and **Borgman, CL.** 2017. On the Reuse of Scientific Data. *Data Science Journal*, 16: 1–9. Retrieved from <http://datascience.codata.org/articles/10.5334/dsj-2017-008/>. DOI: <https://doi.org/10.5334/dsj-2017-008>
- Paté, A, Boschi, L, Dubois, D, Le Carrou, J-L** and **Holtzman, B.** 2017. Auditory display of seismic data: On the use of experts' categorizations and verbal descriptions as heuristics for geoscience. *The Journal of the Acoustical Society of America*, 141(3): 2143–2162. DOI: <https://doi.org/10.1121/1.4978441>
- Paul, C.** 2008. Challenges for a ubiquitous museum: From the white cube to the black box and beyond. In Paul, C (ed.), *New Media in the White Cube and Beyond: Curatorial Models for Digital Art*, 53–75. Berkeley, CA: University of California Press.
- Ramapriyan, H, Moses, J** and **Duerr, R.** 2012. Preservation of data for Earth system science – Towards a content standard. In *Proceedings of 2012 IEEE International Geoscience and Remote Sensing Symposium*, 5304–5307. DOI: <https://doi.org/10.1109/IGARSS.2012.6352411>
- Rhee, J.** 2017. The Materialities of Big Data. In Willis, D, Braham, WH, Muramoto, K and Barber, DA (eds.), *Energy Accounts: Architectural Representations of Energy, Climate, and the Future*, 68–77. New York, NY: Routledge.
- Slavnic, Z.** 2013. Towards Qualitative Data Preservation and Re-Use—Policy Trends and Academic Controversies in UK and Sweden. *Forum: Qualitative Social Research*, 14(2).
- Smith, JA, Flowers, P** and **Larkin, M.** 2009. *Interpretative phenomenological analysis: theory, method and research*. Los Angeles, CA: SAGE.
- Straebel, V** and **Thoben, W.** 2014. Alvin Lucier's Music for Solo Performer: Experimental music beyond sonification. *Organised Sound*, 19(1): 17–29. DOI: <https://doi.org/10.1017/S135577181300037X>
- Strauss, AL** and **Corbin, JM.** 1998. *Basics of qualitative research: techniques and procedures for developing grounded theory* (2nd ed.). Thousand Oaks, CA: Sage Publications.
- Thieberger, EM** and **Dodge, C.** 1995. An Interview with Charles Dodge. *Computer Music Journal*, 19(1): 11–24. DOI: <https://doi.org/10.2307/3681298>
- van de Sandt, S, Dallmeier-Tiessen, S, Lavasa, A** and **Petras, V.** 2019. The Definition of Reuse. *Data Science Journal*, 18(1): 1–19. Retrieved 2020-02-06, from <https://datascience.codata.org/articles/10.5334/dsj-2019-022/>. DOI: <https://doi.org/10.5334/dsj-2019-022>
- Vesna, V.** 2007. *Database Aesthetics: Art in the Age of Information Overflow*. Minneapolis, MN: University of Minnesota Press. DOI: <https://doi.org/10.5749/j.ctts7q7>
- Weil, B.** 2012. Notes on the immersive datascares of Ryoji Ikeda. In *Ryoji Ikeda: datamatics*, 121–129. Milano: Charta. (OCLC: 769430108).
- Yoon, A.** 2017. Role of Communication in Data Reuse. In *Proceedings of the 80th Annual Meeting of the Association for Information Science and Technology*. Crystal City, VA.
- Yoon, A** and **Lee, YY.** 2019. Factors of trust in data reuse. *Online Information Review*, 43(7): 1245–1262. DOI: <https://doi.org/10.1108/OIR-01-2019-0014>

**How to cite this article:** Boutard, G. 2020. Alter-Value in Data Reuse: Non-Designated Communities and Creative Processes. *Data Science Journal*, 19: 23, pp.1-12. DOI: <https://doi.org/10.5334/dsj-2020-023>

**Submitted:** 28 February 2020      **Accepted:** 01 May 2020      **Published:** 22 May 2020

**Copyright:** © 2020 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <http://creativecommons.org/licenses/by/4.0/>.

**]u[** *Data Science Journal* is a peer-reviewed open access journal published by Ubiquity Press.

**OPEN ACCESS** 