

# Secondary Publication



Scheltjens, Werner

## Upcycling historical data collections : A paradigm for digital history?

Date of secondary publication: 18.04.2023

Accepted Manuscript (Postprint), Article

Persistent identifier: urn:nbn:de:bvb:473-irb-590876

### Primary publication

Scheltjens, Werner (2023): Upcycling historical data collections : A paradigm for digital history?, in: Journal of documentation, Bradford: Emerald, Vol. 79, Nr. 6, pp. 1325–1345, doi: 10.1108/JD-12-2022-0271.

### Legal Notice

This work is protected by copyright and/or the indication of a licence. You are free to use this work in any way permitted by the copyright and/or the licence that applies to your usage. For other uses, you must obtain permission from the rights-holders.

This document is made available under a Creative Commons license.



The license information is available online:

<https://creativecommons.org/licenses/by-nc-nd/4.0/legalcode>



**Upcycling historical data collections. A paradigm for digital history?**

|                  |  |
|------------------|--|
| Journal:         | <i>Journal of Documentation</i>  |
| Manuscript ID    | JD-12-2022-0271.R1   |
| Manuscript Type: | Article  |
| Keywords:        | Digital history, Reference works, Legacy collections, Retrodigitization, FAIR principles, Data Collection, Data rescue, Information creation, Paratext, Paradata |
|                  |  |

SCHOLARONE™  
Manuscripts

1  
2  
3 Upcycling historical data collections. A paradigm for digital history?  
4  
5

## 6 **Structured Abstract**

### 7 **Purpose**

8  
9 Upcycling is conceptualised as a digital historical research practice aimed at increasing the  
10 scientific value of ~~texts~~ historical data collections produced in print or in electronic form  
11 between the eighteenth and the late twentieth centuries. ~~Applied to historical data~~  
12 ~~collections,~~ ~~†~~The concept of upcycling facilitates data rescue and reuse as well as the study  
13 of information creation processes deployed by previous generations of researchers.  
14  
15  
16  
17  
18

### 19 **Design/methodology/approach**

20 Based on a selection of two historical reference works and two legacy collections, an  
21 upcycling workflow consisting of three parts (input, processing and documentation, and  
22 output) is developed. The workflow facilitates the study of historical information creation  
23 processes based on paradata analysis and targets the cognitive processes that precede and  
24 accompany the creation of historical data collections.  
25  
26  
27  
28

### 29 **Findings**

30  
31 The proposed upcycling workflow furthers our understanding of computational methods and  
32 their role in historical research. Through its focus on the information creation processes that  
33 precede and accompany historical research, the upcycling workflow contributes to historical  
34 data criticism and digital hermeneutics.  
35  
36  
37  
38

### 39 **Originality/value**

40 Many historical data collections produced between the eighteenth and the late twentieth  
41 century do not comply with the principles of FAIR data. The paper argues that ignoring the  
42 work of previous generations of researchers is not an option, because it would make current  
43 research practices more vulnerable, and would result in losing access to the experiences  
44 and knowledge accumulated by previous generations of scientists. The proposed upcycling  
45 workflow takes historical data collections seriously and makes them available for future  
46 generations of researchers.  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

## 1. Introduction

Since its inception in the early 2000s, upcycling has become a buzzword among entrepreneurs and environmental activists, eager to underline the forward-looking, sustainable nature of their product or initiative (McDonough & Braungart, 2013). Upcycling depicts the process of taking something old, outworn and seemingly worthless to make something qualitatively new, fashionable and of higher value (Zimring, 2016, p.45). In recent years, scientists have discovered the concept as a means to describe methods that move beyond data rescue and curation towards data reuse and integration (Vearncombe *et al.*, 2017; Avesani *et al.*, 2019). This paper argues that, due to its focus on scientific value gains, upcycling has a significant role to play in the historical sciences as a means to make historical data collections interoperable and reusable, on the one hand, and to examine the methods of previous generations of researchers, on the other hand. Dealing with historical data collections, upcycling also calls for a more general reflection on the achievements and remaining challenges of mass digitization.

The paper conceptualizes upcycling as a digital historical research practice aimed at increasing the scientific value of texts that have been produced in print or in electronic form between the eighteenth and the late twentieth centuries. Following Don McKenzie's seminal work, 'text' is understood as "(...) verbal, visual, oral, and numeric data, in the form of maps, prints, and music, of archives of recorded sound, of films, videos, and any computer-stored information, everything in fact from epigraphy to the latest forms of discography" (McKenzie 1999, p.13). McKenzie's broad definition of 'text' highlights the significance of non-book texts as one of many kinds of recorded forms (McKenzie 1999, p.12). It indicates that, beyond the examples of this paper, upcycling may be applied in various other research contexts as well, such as oral history (Thieberger 2018; McCartney 2018), film studies (Cherchi Usai 2010; Ricci 2008), or even 3D modelling (De Kramer 2022). Within this broad definition, This paper upcycling targets historical data collections (short: HDCs) as a particular kind of text. HDCs compile, structure and interpret data, typically (but not necessarily) from multiple sources. Many HDCs were produced before the invention of the computer; many more were produced before the era of laptops, digital cameras, stable internet connections and mass digitization. HDCs comprise *reference works*, such as dictionaries, lexica and encyclopaedia, and *legacy collections* that document the results of historical research projects focused on collecting data from primary sources.

## 2. Historical data collections

### 2.1. General characteristics

HDCs do not contain *raw data* like primary sources in archives, but *processed data* taken from these primary sources. A significant feature of both raw and processed data is that they are

1  
2  
3 meaningless when observed in isolation, i.e. as atomic, individual data items. If, however, data  
4 items are interpreted in relation to their surrounding data items, they become information. This  
5 requires structuring the data, ~~for example, horizontally and vertically~~. We call this process  
6 information creation. Research based on information, or information analysis, generates  
7 knowledge. In the context of HDCs, an inextricable link between *raw* and *processed* data  
8 remains, even if it is not explicitly documented and may have become invisible in the course  
9 of data collection. ~~HDCs were produced at a time when the mass digitization of archival~~  
10 ~~collections had yet to begin; in fact, many HDCs were produced even before the invention of~~  
11 ~~the computer~~. As a result, primary sources often exist independently from the HDCs in which  
12 they were used. Today, the situation is different. On the semantic web, Linked Historical Data  
13 express this inextricable link and may thus be understood as the union of primary and  
14 secondary sources (Meroño-Peñuela & Hoekstra, 2014, p.286).  
15  
16  
17  
18  
19  
20  
21  
22  
23

## 24 2.2. Reference works

25 Reference works that aimed at collecting and distributing (historical) data and knowledge  
26 about a variety of topics became increasingly popular since the eighteenth century. Their  
27 proliferation intertwined with advances in book production and trade (Baten & Van Zanden,  
28 2008; Buringh & Van Zanden, 2009; Bellingradt & Salman, 2017), a growing interest in the  
29 collection and diffusion of 'useful knowledge', and the emergence of the 'first knowledge  
30 economy' in Western Europe in the eighteenth century (Mokyr, 2002; Berg, 2007; Jacob,  
31 2014). Few of these reference works are available as digital editions, like Johann Georg  
32 Krünitz' *Ökonomische Enzyklopädie* (Krünitz, 1773-1858).<sup>1</sup> A much larger number, for  
33 example, reference works in the realm of commerce and trade are merely available in print or  
34 as scans in digital library collections. Digital editions of merchant manuals, handbooks on  
35 weights, measures and currencies, and commercial practices, or of encyclopaedia for the  
36 handicrafts, industry and commerce are not available.  
37  
38  
39  
40  
41  
42  
43

44 Sometimes, automatically generated OCR was produced during the scanning process. This  
45 offers some immediate new possibilities for working with historical reference works, but is  
46 insufficient for more encompassing analyses of their contents (\*\*). There are two important  
47 reasons. First, as a subgroup of digital library collections for the humanities, historical  
48 reference works often use different character sets and have a complex page structure  
49 including tables, in-text columns and significant typographical features, which the commonly  
50 used (commercial) OCR pipelines do not handle particularly well (Cordell, 2017; Chiron *et al.*,  
51 2017). With the broader category of humanities library collections, many ~~historical data~~  
52 ~~collections~~HDCs share a 'circuitous path to digitization', which includes sd the production of  
53  
54  
55  
56  
57  
58  
59

---

60 <sup>1</sup> <http://www.kruenitz1.uni-trier.de/>

scans from microfilms and has a negative impact on the OCR accuracy (Christy *et al.*, 2017). Second, a particular feature of historical reference works is that they hide a significant amount of implicit knowledge, which, due to organisation of the materials, dispersed across the entire work and is therefore almost impossible to retrieve using standard searches.

### 2.3. *Legacy collections*

The term legacy collections originates from archaeology, but is also used in other domains such as cultural heritage (Heitman *et al.*, 2017), library and archival collection management (Jaillant, 2022; Shiue *et al.*, 2021) and linguistics (Dobrin & Schwartz, 2021). Legacy, or 'orphaned', collections consist of "(...) older materials that do not meet modern 'best practice' curation standards and require considerable resources to be preserved for future research." (MacFarland & Vokes, 2016, p.161). As a rule, legacy collections were produced at a time when there was little concern for, or knowledge about, making data available to others (Vearncombe *et al.*, 2017; Clarke & Shiue, 2020). This is true for most legacy collections created before the year 2000 (cf. Kelly *et al.*, 2022)<sup>2</sup> and for much of the data and documentation collected by researchers in the course of their careers (Shiue *et al.*, 2021, p.308). In line with other disciplines dealing with legacy collections, many HDCs have "(...) lost curatorial support or w[ere] abandoned prior to transfer to a curatorial facility" after research results were published (MacFarland & Vokes, 2016, p.162), while the collections themselves remain "(...) as evidence of historical research and as potentially reusable data to support new scientific inquiries" (Shiue *et al.*, 2021, p.308). Insofar as HDCs are concerned, the term 'historical' has a double meaning, indicating both research that has been conducted in the past and is now historical and research about the past itself. HDCs are both. They contain evidence of the past, collected in the past, using past methodologies and techniques.

### 2.4. *Where HDCs are kept*

Many printed reference works and legacy collections are kept at memory institutions (galleries, libraries, archives and museums, short: GLAM), with many more hidden in personal libraries, index card systems, or computers. Thanks to large-scale efforts such as the digitization of sixteenth-, seventeenth- and eighteenth-century printed works in German-speaking countries (Sommer, 2010; Herr, 2012),<sup>3</sup> the Early English Books Online (EEBO) project,<sup>4</sup> or the Internet Archive,<sup>5</sup> a large and growing number of HDCs has resurfaced. They are available online as

---

<sup>2</sup> Today, most researchers collect their data digitally, which leads Kelly *et al.* (2022) to conclude that the issue with legacy collections is a finite problem. This, however, presupposes that researchers today apply accepted standards for data collection, which is not always the case.

<sup>3</sup> <http://www.vd16.de/>, <http://www.vd17.de/>, <http://www.vd18.de/>.

<sup>4</sup> <https://proquest.libguides.com/eebopqp>

<sup>5</sup> <https://archive.org/>

1  
2  
3 scans of book pages; sometimes complemented with automatically produced OCR. It is hard  
4 to estimate how many reference works would fall into this category, but even a quick look at  
5 some bibliographies and catalogues about a specialist topic, such as commercial dictionaries  
6 and merchant manuals printed before 1850, shows that their number easily runs in the  
7 thousands (see e.g. Besomi, 2012; Jeannin *et al.*, 1991-2001). Other HDCs have been  
8 deposited as datasets in online repositories, which exist in a variety of different forms at  
9 different institutional levels, ranging from university to discipline-oriented and national data  
10 infrastructures. Clearly, the number of HDCs that could be upcycled is potentially huge, but it  
11 is almost impossible to obtain an estimate of their number.  
12  
13  
14  
15  
16  
17  
18

### 19 2.5. Examples

20 In the remainder of the paper, I refer to a selection of four HDCs from the disciplines of  
21 maritime history, an interdisciplinary field of research that combines aspects of economic,  
22 social and global history (e.g. Harlaftis, 2020), and historical metrology, an important (but  
23 underrated) auxiliary science of history and object of study from a history of science  
24 perspective (e.g. Kramper, 2019) (see table 1). They will support for my argument. Two pre-  
25 1950 HDCs are included to underline that data collection and computation are by no means  
26 recent innovations related to the advent of the computer. Two more recent HDCs allow gaining  
27 insight into the development of computational methods during the early phase of using  
28 computers for the collection, analysis and dissemination of historical data. Two HDCs are  
29 reference works; two are legacy collections.  
30  
31  
32  
33  
34  
35  
36  
37

38 <<insert table 1 here>>  
39  
40

41 The oldest HDC is that of Christian and Friedrich Noback's "comprehensive pocketbook" of  
42 weights and measures (Noback & Noback, 1851). On just over 1,900 pages, this book  
43 provides detailed information about everything a merchant had to know to conduct his  
44 business (Denzel, 2002). Even today, researchers consult works like the "Noback" to look up  
45 the metric equivalents of pre-modern units of measure, but they tend to ignore that the book  
46 remains vague and often silent about the sources of this information. Ronald Zupko's  
47 *Dictionary of Weights and Measures for the British Isles* (Zupko, 1985) is the more recent  
48 counterpart of the "Noback". Zupko traces mentions of weights and measures in a large  
49 number of primary sources. Although created with the assistance of a computer, the dictionary  
50 was published as a typewritten manuscript, which is now available online in various digital  
51 formats.<sup>6</sup> The third HDC emerged roughly in the same period as Zupko's computer-assisted  
52  
53  
54  
55  
56  
57  
58  
59

60 <sup>6</sup> [https://archive.org/details/bub\\_gb\\_0l\\_k-XMliQIC/mode/2up](https://archive.org/details/bub_gb_0l_k-XMliQIC/mode/2up)

1  
2  
3 work on historical metrology. Between 1975 and 1990, a group of British historians created a  
4 *Computerized Edition of the Liverpool Plantation Registers, 1744-1786* (Richardson,  
5 Beedham & Schofield, 1992). In 2007, the UK data service (or its predecessors) converted  
6 the data files into a platform-independent format under the name *Liverpool Shipping and*  
7 *Trade, 1744-1786* (further: LST).<sup>7</sup> The creators of LST were inspired by the Danish historian  
8 Hans Christian Johansen. In the form of an electronic database of the Danish Sound toll  
9 registers (STR) for the late eighteenth and early nineteenth centuries, Johansen (1983) aimed  
10 to produce a follow-up of the so-called Sound Toll Tables (short: STT; original Danish: *Tabeller*  
11 *over skibsfart og varetransport gennem Øresund*) (Bang & Korst, 1906-1953). This  
12 monumental work in seven volumes was created in the first half of the twentieth century.  
13 Alongside the LST, the STT are our second legacy collection. The STT contain about 3,000  
14 pages of statistics about ship movements and commodity flows between the North Sea and  
15 the Baltic from 1497 to 1783. Despite substantial source criticism (e.g. Jeannin 1964; Gøbel  
16 2010; \*\*\*), the STT have had enormous significance for the study of early modern European  
17 transport and trade (Veluwenkamp *et al.*, 2021, pp.148-152). Between 2009 and 2020, an  
18 electronic database of ships' passages through the Danish Sound, including the cargoes  
19 carried and taxes paid, was created in the Sound Toll Registers Online project (short: STRO).<sup>8</sup>  
20 This electronic database links entries directly to the corresponding scans of the register pages  
21 where the entry was found.

22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33 STRO readily shows how the results of digitization projects carried out in recent years may be  
34 employed to further our understanding of previous historical research methods. Alongside the  
35 electronic database, the entire archival collections that had been processed manually to create  
36 the STT has been made available as well. This gives us two reasons for including the STT as  
37 an exemplar. Firstly, there is a claim, or rather a belief, that STRO replaces the STT and  
38 overcomes its limitations (see Veluwenkamp *et al.*, 2021). While this may be true from an  
39 historian's perspective, the claim is unwarranted from an epistemological perspective. Even if  
40 the STT are 'outdated' by modern standards, they still are an exceptional computational effort.  
41 Remarkably, little is known about the creation of the STT. As an exemplar of an HDC created  
42 in the time preceding the invention of computers, the STT are a highly relevant object of  
43 historical inquiry. Thanks to STRO, a massive amount of digital materials is now available to  
44 support research on the epistemology of the STT.

## 51 52 53 54 2.6. A dialogue with the first retrodigitization

55  
56  
57  
58  
59 <sup>7</sup> <https://doi.org/10.5255/UKDA-SN-2923-1>

60 <sup>8</sup> [www.soundtoll.nl](http://www.soundtoll.nl)

1  
2  
3 Each case in this paper should be understood as a dialogue with the accomplishments and  
4 limitations of the first mass retrodigitization of the last 20 years. All cases benefit from the  
5 availability of digital materials related to the HDC, such as scans of archival documents,  
6 historiography or OCR'ed texts. At the same time, the quality of the scans of the selected  
7 historical reference works is often insufficient, which results in 'noisy' OCR. For the available  
8 scans of the printed legacy collection STT there are some additional problems. The middle of  
9 the table was consistently lost during the scanning process. Moreover, the OCR that was  
10 automatically produced for the work cannot be used, because it does not account for the  
11 tabular structure of the source. In the case of the LST, the files could be opened and read after  
12 some trial and error, but they were almost incomprehensible due to deficiencies in their  
13 structure and the abundant use of abbreviations. In all cases, part of the work done in the first  
14 retrodigitization needs to be revised or at least requires significant post-processing when novel  
15 research questions are to be answered based on their contents.

16 For example, tests performed with the "Noback" have shown that the OCR quality is sufficient  
17 for answering only a few basic questions that are essentially limited to ad-hoc fact-checking  
18 (\*\*\*). Although the quality of the OCR is much less of a problem for Zupko's dictionary, it  
19 suffers from the same issues. Novel insights about the composition of past metrological  
20 systems can only be obtained through comparative analysis of a large number of articles in  
21 the book, for example, all articles about places belonging to the same economic region. But  
22 scrolling through the scans or searching in the OCR'ed text are inefficient, if not impossible,  
23 methods for conducting this kind of analysis. Even though historical reference works, such as  
24 the "Noback" or Zupko's dictionary, are available in digital form, it requires upcycling rather  
25 than a bunch of scans and automatically generated OCR to preserve the historical knowledge  
26 they contain in a findable, accessible, interoperable and reusable way.

### 3. The issue of un-FAIR-ness

#### 3.1. *The FAIR principles and historical data collections*

27 Across all sub-disciplines of the historical sciences and related areas of historical inquiry,  
28 many HDCs exist that could be upcycled. Yet often, it is not a sinecure, and sometimes even  
29 outright impossible to *Find* these HDCs. Focusing on research data that was produced at  
30 universities, the Swedish archivist Jakobsson (2021) laments that "[r]esearch data created  
31 within universities often ended up in a box on a bookshelf, despite university and researcher  
32 requirements to archive and share the data. (...) In addition, university researchers have  
33 typically considered such data to be their private property, to be used only for their own  
34 publications." Such HDCs, if they have survived, are found by chance, through personal  
35 contacts, or thanks to traces and hints left in historiography. Indeed, the search for HDCs often  
36 involves serendipity (Martin & Quan-Haase, 2017).

1  
2  
3 HDCs are hard, and sometimes impossible, to Access, most often because of the format in  
4 which the data are stored. Some HDCs are only available in print, mostly in thick volumes of  
5 tabular data, like the STT. Other HDCs are kept on index cards, punched cards, magnetic  
6 tapes, floppy disks, microfiches, or CD-ROMS. Each of these storage media is either outdated,  
7 or inefficient, or both, and hinders access to the data in its own specific ways.

8  
9  
10  
11 HDCs are hard to make *Interoperable*, because of design choices, incompatibility of  
12 (proprietary) data formats, and lack of documentation about the collection.

13  
14 These issues are widely known and discussed among librarians, archivists and historians  
15 (Rosenzweig, 2003). Initiatives promoting the production of [FAIR](#) historical data ~~in accordance~~  
16 ~~with the FAIR principles~~ have become widespread since 2016, when the FAIR principles were  
17 first published (Wilkinson *et al.*, 2016). Recent work on making historical data 'FAIR' has  
18 focused on the issue of interoperability, which is much harder to achieve and implement than  
19 the more 'technical' recommendations about making data findable, accessible and reusable  
20 with the help of persistent identifiers, metadata, data access rules and user licenses (Beretta,  
21 2021).

22  
23 Finally, HDCs were often conceived without anticipating an interest from future generations of  
24 researchers (Barats *et al.*, 2020; Kelly *et al.*, 2022), and were distributed in print or – since the  
25 1960s – in electronic formats that more or less severely limit their *Reuse*. Often they are poorly  
26 documented and hard to understand for outsiders. As a result, there is a risk that access to  
27 the HDCs is eventually lost.

### 28 29 30 31 32 33 34 35 36 3.2. *Why bother?*

37  
38 So why bother? Clearly, it is hard, and sometimes even impossible, to *Reuse* HDCs let alone  
39 make them *Interoperable*. They do not comply with the principles of FAIR data (cf. Wilkinson  
40 *et al.*, 2016), often suffer from methodological deficiencies and technical hindrances (cf.  
41 Huvila, 2022, p.30), and are "(...) embedded in a local context, which makes it difficult for  
42 reusers to understand and trust the data" (Faniel & Zimmerman, 2011, p.60). Reciprocally,  
43 "[t]he same factors that make it hard for scientists to reuse data collected by those from a  
44 different community also make it difficult for data producers to share data" (Faniel &  
45 Zimmerman, 2011, p.61). No wonder that "data reuse is not (...) as widespread as it might be  
46 imagined (...)" (Huvila, 2022, p.30). Yet at the same time, "[t]here is almost universal  
47 agreement that scientific data should be shared for use beyond the purposes for which they  
48 were initially collected" (Faniel & Zimmerman, 2011, p.59) and within the digital humanities  
49 there have been repeated calls to embrace the 'data deluge' (e.g. Guldi & Armitage, 2014).  
50 Moreover, in recent years, researchers have raised "(...) legitimate concerns of the present,  
51 often limited consideration of the usefulness, limits and contextuality of the currently available  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 and (re)used research data” (Huvila, 2022, p.30; see also: Faniel & Zimmerman, 2011; Bevan,  
4 2015; De Freitas, 2017; Smith, 2022).

5  
6 We have to bother. Deliberately throwing unFAIR HDCs out, or ignoring them, because they  
7 do not comply with current standards disregards the time and effort that had already been  
8 invested in their production. Doing the same work all over again with novel tools and methods  
9 is likely to be more expensive than reappraising what had already been achieved. More  
10 importantly, perhaps, choosing not to use existing HDCs because they are unFAIR would  
11 result in losing access to the experiences and knowledge accumulated by previous  
12 generations of scientists. This, in turn, would stand in the way of "capitalisation on existing  
13 structures and scientific legacy to avoid the duplication of means, supports and tools” (Barats  
14 *et al.*, 2020). Dismissing unfair HDCs would also make current research practices more  
15 vulnerable as it increases the risk of making the same mistakes again (Russell & Vinsel, 2018).  
16 Most importantly, however, unFAIR HDCs are essential for the systematic preservation of past  
17 methodologies for data collection and information creation.

18  
19 For a very long time, it seems that the acts of selecting, collecting, and grouping data, and of  
20 making calculations, in one word: computing, were seen as something straightforward, simple,  
21 theory-less and purely pragmatic that had its place in introductions to quantitative methods for  
22 historians, or manuals for constructing historical databases (e.g. Harvey & Press, 1996;  
23 Cameron & Richardson, 2005). Fierce debates between proponents and sceptics of  
24 computational methods accompanied the early adoption of computers in historical research,  
25 but the discussion evolved around their utility *per se* rather than around the methodological  
26 implications of their use (Thomas, 2004; Greenstein, 1989). Even today, general textbooks on  
27 methods for historical research pay limited attention to computation, and if they do, they stick  
28 to the pragmatics of using computers for historical research and refrain from exploiting its  
29 methodological implications (e.g. Gunn & Faire, 2016).

#### 44 **4. The pursuit of scientific value gains**

45  
46 By targeting unFAIR HDCs, upcycling pursues two kinds of scientific value gains that directly  
47 address this lacuna in historical methodology. The first way to achieve scientific value gains  
48 is through *data rescue*, or the process of making HDCs FAIR and thus available to all. The  
49 second way is through *documentation and analysis* of the *processing steps* necessary to make  
50 HDCs FAIR. As will be discussed later, in the pursuit of these scientific value gains, the  
51 discipline of digital history unveils its paradigmatic character.

52  
53  
54 <<insert figure 1 here. Caption: Value gains from data rescue>>

1  
2  
3 Insofar as HDCs are only available in print or as data files stored on magnetic tapes, diskettes,  
4 or someone's computer, *data rescue* is primarily a technical process coordinated by memory  
5 institutions around the world. For the purpose of upcycling, however, generating or restoring  
6 access to HDCs is barely enough. Indeed, "(...) a loss of intelligibility – that is an understanding  
7 of the context in which data are generated and interpreted – is as critical as the technical or  
8 material loss of data, if not more" (Barats *et al.*, 2020). The goal of data rescue is to improve  
9 access through assessment of the design, structure and contents of the HDC and facilitate  
10 the HDC's subsequent reconfiguration in accordance with current data management  
11 standards (see figure 1). In that sense, the availability of images of books or data files in online  
12 repositories is the starting point rather than the outcome of data rescue.

13  
14 From the point of view of historical methodology, *documentation and analysis* of the  
15 *processing steps* in upcycling workflows are where the most significant scientific value gains  
16 can be achieved. Meticulous documentation of upcycling workflows is deemed to reveal the  
17 technical, methodological, interpretative and creative processes, including their flaws and  
18 deficiencies, that led to the creation of an HDC. Documentation is bound to bring historical  
19 information creation processes (further: ICPs) to the surface. It puts the interactions of  
20 previous generations of researchers with primary sources in the spotlight, asking where, why  
21 and how they collected historical data; how available technical means might have affected  
22 their decisions; how their worldviews might have influenced data selection; and - last but not  
23 least - how they decided upon the format for their HDC. Upcycling thus entails a shift of focus  
24 from the results of ICPs to the actions and cognitive processes that precede and accompany  
25 them.

26  
27 This shift of focus aligns well with recent developments in the study of book history. Previously,  
28 much effort has been devoted to understanding the *results* of book production and knowledge  
29 distribution in the context of grand narratives about modernization, the knowledge economy,  
30 and the rise of the West. More recently, however, researchers have started to focus on the  
31 actions and decisions that *precede, support and impact* the emergence of texts, data and  
32 information. For example, a recent discussion about the sustainability of digital data urged to  
33 "[make] the researcher's 'toolbox' readable and [explain] her/his 'ways of doing things' (...)",  
34 arguing that "(...) aspects that can be understated, obscured or considered of little value in the  
35 production of knowledge (...) could (...) play a role in reinforcing the reflective approach that is  
36 central to research" (Barats *et al.*, 2020). A comparable interest in ICPs can also be found in  
37 a variety of other disciplines, most importantly literary studies with the concept of *paratext*  
38 (Genette & MacLean, 1991; Genette, 1997; Collins & Skover, 1992, 2010; Ciotti & Lin, 2016;  
39 Skare, 2021) and information science with the concept of *paradata* (Börjesson, Skold & Huvila,  
40 2020; Huvila, 2022). Both concepts have been adopted in the upcycling workflow that will be  
41 described will be explained in more detail below. Although their similarity may sound confusing  
42

1  
2  
3 at first, both concepts play distinct roles within the upcycling workflow. They address and  
4 document different information creation processes (see section 6 for more detail). The paratext  
5 focuses on the documentation surrounding an unFAIR HDC and uses this documentation to  
6 gain insight into the 'initial' HDC. Dealing with the paratext is a preparatory step in the  
7 upcycling workflow. Contrastingly, paradata is produced during the upcycling process and  
8 documents changes to the HDC to make it FAIR.- In applying these concepts, uUpcycling  
9 applies these concepts to HDCs and thus aims to contribute to historical information science  
10 (Boonstra *et al.*, 2006; Thaller, 2017), and in particular to the history of computational methods.  
11  
12  
13  
14  
15  
16

## 17 **5. The upcycling workflow**

### 18 **5.1. Buildings blocks**

19 In the previous sections, I have made a case for upcycling HDCs. I have argued that the HDCs  
20 themselves, the community of historians, and the scientific community in general might benefit  
21 from it. In the next sections, I introduce the building blocks of an upcycling workflow, and  
22 explain how they contribute to achieving scientific value gains. My proposed upcycling  
23 workflow consists of three parts: input, processing and documentation, and output (see figure  
24 2).  
25  
26  
27  
28  
29

30 Input consists of the unFAIR HDC and its accompanying paratext. The paratext comprises  
31 textual, iconic, material and factual elements that surround and give meaning to the HDC- and  
32 aim at making it comprehensible, like which may be understood broadly as surrounding  
33 information, like the table of contents, indices, user guides, etc. (see section 5.2. for more  
34 detail) that aim at making the HDC comprehensible. The paratext supports the main processes  
35 of the upcycling workflow: data rescue, assessment and reconfiguration. These processes go  
36 hand in hand with two tasks: (1) formulating research questions that could only be answered  
37 with better access to the data in the HDC and (2) conceptualising an 'ideal' data model for the  
38 upcycled HDC. Reconnecting the unFAIR HDC with its original 'raw' data might enhance the  
39 main upcycling processes by providing a means to verify, trace and assess the actions and  
40 decisions that were made by the creators of the HDC.  
41  
42  
43  
44  
45  
46

47 Data rescue might have to start with technical procedures to generate or restore access to  
48 HDCs, but they are not our main concern. Data rescue is understood here as part of an  
49 iterative process to improve access to an unFAIR HDC and make it interoperable and  
50 reusable. Starting point of data rescue are the scans or data files of the unFAIR HDCs as well  
51 as the paratext that documents their structure and contents. Insofar as the paratext is  
52 concerned, data rescue secures the availability in digital format of parts of the paratext that  
53 can significantly improve access to and use of the HDC. Assessment examines what needs  
54 to be done to reconfigure the unFAIR HDC and make it FAIR. Answers to this question are  
55 sought through manual and digital exploration of the HDC, which is supported by comparison  
56  
57  
58  
59  
60

1  
2  
3 of the HDC with the paratext and reconnected primary source documents. The goal of  
4 assessment is to identify and remove errors and omissions in the data model of the HDC. In  
5 doing so, it explicates how the data model can be reconfigured and why this should be done.  
6 Removing data modelling errors effectively means changing the data model and thus  
7 reconfiguring the HDC.  
8

9  
10  
11 Paradata documents the implementation of the iterative processes of data rescue,  
12 assessment and reconfiguration, and captures the essential steps of data rescue, assessment  
13 and reconfiguration in a systematic way. The goal of paradata capturing is to facilitate the  
14 subsequent analysis of the ICPs related to the HDC [\(see section 6 for more detail\)](#). The output  
15 of upcycling is not a single product, like a book, or an article, or an electronic dataset, but an  
16 upcycled HDC, which comprises the 'original' HDC, the paradata produced during upcycling,  
17 the 'new' HDC and its reconnection to the original 'raw' data. Reconfigured in that way, the  
18 upcycled HDC serves as a starting point for the study of historical ICPs using paradata  
19 analysis.  
20  
21  
22  
23  
24

25  
26  
27 <<insert figure 2 here. Caption: The upcycling workflow>>  
28  
29

### 30 5.2. Using the paratext

31 Upcycling uses the paratext found in documentation, guides, descriptions, prefaces, indexes  
32 and the HDC itself. The concept of paratext as introduced by the French literary scholar Gérard  
33 Genette (1930-2018) targets the information that surrounds the actual text of a book, starting  
34 with advertisements about forthcoming publication, title page, front and back cover, and  
35 ending with prefaces, indexes and reviews (Genette, 1997). According to Genette, each  
36 element of the paratext of a book provides insight into the creation of the book and contributes  
37 to understanding its contents. Thus, the paratext is "(...) the means by which a text makes a  
38 book of itself and proposes itself as such to its readers, and more generally to the public"  
39 (Genette & MacLean, 1991, p.261). Developed in the realm of literary studies, the concept of  
40 paratext focuses heavily on books (Ciotti & Lin, 2016, p.VI-VII), although its applicability is not  
41 limited to literary works (e.g. Birke & Christ, 2013; Skare, 2020, 2021). Therefore, Ciotti and  
42 Lin suggested defining paratext more broadly as a reflection of "the activities of everyone  
43 involved in the production, transmission, dissemination and reception of the manuscript and  
44 its content: authors, editors, scribes, artisans, commentators, readers, sellers, owners and so  
45 on..." (Ciotti & Lin, 2016, p.VIII). [Significantly, the paratext also comprises iconic, material and  
46 factual elements, such as illustrations, the typographical features of a work, or the author's  
47 gender, age and reputation \(Skare 2020, p.512\), which target, among other things, the HDC's  
48 physical form. These elements contribute to overcoming the limitations of a one-sided  
49 hermeneutic approach that focuses on the immaterial and contribute to the renewed interest](#)  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 [in materiality in the digital age \(for more information, see Skare, 2021\)](#). A comprehensive  
4 analysis of the paratext of an HDC sheds light on the intentions, methods, creative decisions  
5 and implementation strategies of the creators of the collection and is therefore a prerequisite  
6 for upcycling. Analysis of the paratext is essentially an inductive **hermeneutic** process (Genette  
7 & MacLean, 1991, pp.270-271) that is similar to the methods of historical praxeology, the study  
8 of practices in the past (Haasis & Rieske, 2015; Füssel, 2022).

9  
10  
11 In the “Noback”, Zupko’s dictionary, and the STT, elements of the paratext are found in the  
12 works themselves, before and after their ‘core’. In the ‘Vorwort’ (*Eng.* preface) and ‘Einleitung’  
13 (*Eng.* introduction) (Noback & Noback, 1851, pp.V-LII), the ‘Introduction’ (Zupko, 1985, pp.x-  
14 xiii) or the ‘Vorord / Avant-Propos’ (Bang & Korst 1906, pp.III-X; 1922, pp.III-X; 1930, pp.VII-  
15 XX), the authors explain their motives and describe, in more or less detail, how they collected  
16 and systematized their data. Additionally, depending on the example, one or more indexes are  
17 included for geographical locations, commodities and abbreviations. Alongside the  
18 introduction and acknowledgements, the paratext of Zupko’s dictionary also consists of  
19 several auxiliary tables, e.g. for abbreviations (Zupko, 1985, pp.xxii-xxiv), which precede the  
20 dictionary, and an extensive bibliography, which is added at the end (Zupko, 1985, pp.447-  
21 520). In the LST, a separate ‘guide’ accompanies the data files uploaded to the UK data  
22 archive. This ‘guide’ describes a relational database constructed in *dBase IV*, explains the  
23 elements of each table, and provides ‘codenames for occupations of owners of vessels’,  
24 ‘codenames for places included in database’ and abbreviations of the ‘sources used in the  
25 database’ as appendices to the guide (Richardson *et al.*, 1992). By processing parts of the  
26 paratext, access to the HDC is easily improved. For example, the appendices in the ‘guide’ to  
27 the database of the Liverpool Plantation Registers can be processed with an OCR tool to make  
28 them machine-readable and linkable to the data files in the collection.

### 5.3. *Adopting an emic perspective*

29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44 The elements of the paratext that can be found in the immediate neighbourhood of the ‘core’  
45 of the HDC provide valuable support for understanding the intricacies of their composition, but  
46 based on these elements alone, a comprehensive understanding of the HDC is hard to attain.  
47 Where the documentation is insufficient, exploratory analysis of the HDC itself should be  
48 conducted to fill remaining gaps in the documentation (cf. Ciotti & Lin, 2016, p.VIII). The extent  
49 of such posthumous paratext creation varies. In some cases, very little documentation is  
50 available; in other cases, the documentation is abundant and raises only minor questions.  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65  
66  
67  
68  
69  
70  
71  
72  
73  
74  
75  
76  
77  
78  
79  
80  
81  
82  
83  
84  
85  
86  
87  
88  
89  
90  
91  
92  
93  
94  
95  
96  
97  
98  
99  
100  
101  
102  
103  
104  
105  
106  
107  
108  
109  
110  
111  
112  
113  
114  
115  
116  
117  
118  
119  
120  
121  
122  
123  
124  
125  
126  
127  
128  
129  
130  
131  
132  
133  
134  
135  
136  
137  
138  
139  
140  
141  
142  
143  
144  
145  
146  
147  
148  
149  
150  
151  
152  
153  
154  
155  
156  
157  
158  
159  
160  
161  
162  
163  
164  
165  
166  
167  
168  
169  
170  
171  
172  
173  
174  
175  
176  
177  
178  
179  
180  
181  
182  
183  
184  
185  
186  
187  
188  
189  
190  
191  
192  
193  
194  
195  
196  
197  
198  
199  
200  
201  
202  
203  
204  
205  
206  
207  
208  
209  
210  
211  
212  
213  
214  
215  
216  
217  
218  
219  
220  
221  
222  
223  
224  
225  
226  
227  
228  
229  
230  
231  
232  
233  
234  
235  
236  
237  
238  
239  
240  
241  
242  
243  
244  
245  
246  
247  
248  
249  
250  
251  
252  
253  
254  
255  
256  
257  
258  
259  
260  
261  
262  
263  
264  
265  
266  
267  
268  
269  
270  
271  
272  
273  
274  
275  
276  
277  
278  
279  
280  
281  
282  
283  
284  
285  
286  
287  
288  
289  
290  
291  
292  
293  
294  
295  
296  
297  
298  
299  
300  
301  
302  
303  
304  
305  
306  
307  
308  
309  
310  
311  
312  
313  
314  
315  
316  
317  
318  
319  
320  
321  
322  
323  
324  
325  
326  
327  
328  
329  
330  
331  
332  
333  
334  
335  
336  
337  
338  
339  
340  
341  
342  
343  
344  
345  
346  
347  
348  
349  
350  
351  
352  
353  
354  
355  
356  
357  
358  
359  
360  
361  
362  
363  
364  
365  
366  
367  
368  
369  
370  
371  
372  
373  
374  
375  
376  
377  
378  
379  
380  
381  
382  
383  
384  
385  
386  
387  
388  
389  
390  
391  
392  
393  
394  
395  
396  
397  
398  
399  
400  
401  
402  
403  
404  
405  
406  
407  
408  
409  
410  
411  
412  
413  
414  
415  
416  
417  
418  
419  
420  
421  
422  
423  
424  
425  
426  
427  
428  
429  
430  
431  
432  
433  
434  
435  
436  
437  
438  
439  
440  
441  
442  
443  
444  
445  
446  
447  
448  
449  
450  
451  
452  
453  
454  
455  
456  
457  
458  
459  
460  
461  
462  
463  
464  
465  
466  
467  
468  
469  
470  
471  
472  
473  
474  
475  
476  
477  
478  
479  
480  
481  
482  
483  
484  
485  
486  
487  
488  
489  
490  
491  
492  
493  
494  
495  
496  
497  
498  
499  
500  
501  
502  
503  
504  
505  
506  
507  
508  
509  
510  
511  
512  
513  
514  
515  
516  
517  
518  
519  
520  
521  
522  
523  
524  
525  
526  
527  
528  
529  
530  
531  
532  
533  
534  
535  
536  
537  
538  
539  
540  
541  
542  
543  
544  
545  
546  
547  
548  
549  
550  
551  
552  
553  
554  
555  
556  
557  
558  
559  
560  
561  
562  
563  
564  
565  
566  
567  
568  
569  
570  
571  
572  
573  
574  
575  
576  
577  
578  
579  
580  
581  
582  
583  
584  
585  
586  
587  
588  
589  
590  
591  
592  
593  
594  
595  
596  
597  
598  
599  
600  
601  
602  
603  
604  
605  
606  
607  
608  
609  
610  
611  
612  
613  
614  
615  
616  
617  
618  
619  
620  
621  
622  
623  
624  
625  
626  
627  
628  
629  
630  
631  
632  
633  
634  
635  
636  
637  
638  
639  
640  
641  
642  
643  
644  
645  
646  
647  
648  
649  
650  
651  
652  
653  
654  
655  
656  
657  
658  
659  
660  
661  
662  
663  
664  
665  
666  
667  
668  
669  
670  
671  
672  
673  
674  
675  
676  
677  
678  
679  
680  
681  
682  
683  
684  
685  
686  
687  
688  
689  
690  
691  
692  
693  
694  
695  
696  
697  
698  
699  
700  
701  
702  
703  
704  
705  
706  
707  
708  
709  
710  
711  
712  
713  
714  
715  
716  
717  
718  
719  
720  
721  
722  
723  
724  
725  
726  
727  
728  
729  
730  
731  
732  
733  
734  
735  
736  
737  
738  
739  
740  
741  
742  
743  
744  
745  
746  
747  
748  
749  
750  
751  
752  
753  
754  
755  
756  
757  
758  
759  
760  
761  
762  
763  
764  
765  
766  
767  
768  
769  
770  
771  
772  
773  
774  
775  
776  
777  
778  
779  
780  
781  
782  
783  
784  
785  
786  
787  
788  
789  
790  
791  
792  
793  
794  
795  
796  
797  
798  
799  
800  
801  
802  
803  
804  
805  
806  
807  
808  
809  
810  
811  
812  
813  
814  
815  
816  
817  
818  
819  
820  
821  
822  
823  
824  
825  
826  
827  
828  
829  
830  
831  
832  
833  
834  
835  
836  
837  
838  
839  
840  
841  
842  
843  
844  
845  
846  
847  
848  
849  
850  
851  
852  
853  
854  
855  
856  
857  
858  
859  
860  
861  
862  
863  
864  
865  
866  
867  
868  
869  
870  
871  
872  
873  
874  
875  
876  
877  
878  
879  
880  
881  
882  
883  
884  
885  
886  
887  
888  
889  
890  
891  
892  
893  
894  
895  
896  
897  
898  
899  
900  
901  
902  
903  
904  
905  
906  
907  
908  
909  
910  
911  
912  
913  
914  
915  
916  
917  
918  
919  
920  
921  
922  
923  
924  
925  
926  
927  
928  
929  
930  
931  
932  
933  
934  
935  
936  
937  
938  
939  
940  
941  
942  
943  
944  
945  
946  
947  
948  
949  
950  
951  
952  
953  
954  
955  
956  
957  
958  
959  
960  
961  
962  
963  
964  
965  
966  
967  
968  
969  
970  
971  
972  
973  
974  
975  
976  
977  
978  
979  
980  
981  
982  
983  
984  
985  
986  
987  
988  
989  
990  
991  
992  
993  
994  
995  
996  
997  
998  
999  
1000

1  
2  
3 documentation requires looking at unFAIR HDCs from the perspective of the human agents  
4 that were involved in its creation. In other words, it requires an emic perspective towards  
5 historical information creation processes.  
6

7  
8 The mass digitization pursued at memory institutions around the world offers profoundly new  
9 possibilities for adopting an emic perspective. In many cases, the HDC can be reconnected to  
10 scans of archival sources or library collections. For example, the extensive bibliography at the  
11 end of Zupko's dictionary could be used as a starting point for tracing online versions of these  
12 works, which could then be connected to the dictionary itself. In the case of the STT and LST,  
13 scans of the original archival sources are readily available on dedicated websites.<sup>9</sup> These  
14 scans allow reiterating the data creation processes of previous generations of researchers,  
15 tracing the missing elements in the paratext, and improving our understanding of the  
16 processes of seeking, collecting, selecting and systematising historical data. The example of  
17 the "Noback", however, shows that reconnection cannot be guaranteed.  
18

19  
20 The potential benefits of reconnecting the unFAIR HDC to scans of archival documents and  
21 books for the purpose of an emic perspective towards historical information creation processes  
22 are high, but the complexity of the process should not be underestimated. Reconnection  
23 requires careful preparation and should only be carried out if significant value gains can be  
24 expected. It adds an extra dimension to the upcycling workflow, but it also comes with  
25 additional technical requirements.  
26  
27

#### 28 29 30 31 32 33 34 35 *5.4. The scenario-based method*

36 The processes of assessment and reconfiguration employ techniques from the realms of  
37 automated text recognition, natural language processing and data modelling (see figure 1).  
38 Before embarking upon these processes, it is necessary to formulate possible research  
39 questions and conceptualise an 'ideal' data model for the upcycled HDCs (see figure 3). The  
40 process of anticipating research questions for the benefit of streamlining data modelling efforts  
41 is known in knowledge engineering as the scenario-based method. It relates the planning  
42 process to so-called competency questions, which are questions that specialist users of the  
43 source would like to examine and answer with the help of a data model (Kendall &  
44 McGuinness, 2019). Lodi *et al.* (2017) and Carriero *et al.* (2021) have successfully adapted  
45 this renowned scenario-based method to questions of digital cultural heritage. However,  
46 addressing the processed data in HDCs requires modelling approaches that are distinct from  
47 those that target the 'data about data', or metadata, in digital heritage collections (\*\*\*)  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58

59  
60 <sup>9</sup> [www.soundtoll.nl](http://www.soundtoll.nl); <https://microform.digital/boa/series/20/the-transatlantic-slave-trade-1675-1907>;  
[www.slavevoyages.org](http://www.slavevoyages.org)

<<insert figure 3 here. Caption: Applications of the scenario-based method

For the “Noback”, we assume an upcycled HDC that could be described as a ‘Web of Science’, in which the contents of the “Noback” are made fit for text reuse detection in earlier (pre-1850) metrological reference works. This will make it possible to tackle research questions about the origins of the data compiled by Christian and Friedrich Noback. For Zupko’s dictionary, we assume an upcycled HDC that could be described as hierarchical ‘Tree of Knowledge’. This will allow answering questions about long-term processes of accumulation and dissemination of metrological knowledge as well as about the long-term development of metrological systems on the British Isles. For the LST, we anticipate a normalized relational data model with links to original archival documents (where available) as the result of upcycling. This would provide a novel basis for answering enduring questions about ownership relations in the shipping industry and the role of Liverpool in the trans-Atlantic slave trade in the eighteenth century. Finally, for the STT, we assume an upcycled HDC that links the semantic data in the STT to the electronic database of the Sound toll registers online (STRO), and thus also to the scans of the registers. This combination of sources makes it possible to develop a kind of ‘genetic criticism’ for HDCs (cf. Van Hulle, 2022). Rather than focusing on the published statistics *per se* to reach a verdict about their reliability, the upcycling workflow allows tracing the cognitive processes and pragmatic decisions made by the compilers of the STT in an unprecedented way.

## 6. Paradata production and analysis

The previous section did not provide lengthy descriptions of all the different steps that were (or could be) taken to improve access to and use of unFAIR HDCs, but instead limited itself to a few examples showing what data rescue, assessment and reconfiguration entail in practice. These few examples are deemed sufficient for a discussion of the production of paradata as a basic requirement for gaining insight into the historical ICPs that are ‘hidden’ in each HDC.

### 6.1. The concept of paradata

Paradata is a well-known concept in survey research and archaeology (e.g. Bentkowska-Kafel & Denard, 2012; Huvila, 2022). It recently started to attract the attention of digital humanists as well, in particular in relation to the issue of reproducibility of research results (Schubert, 2021; Schöch *et al.*, 2020). Isto Huvila summarises the importance of paradata (and more generally of the study of information creation processes) as follows:

“Without documentation of how data was captured or created, in what types of conditions, what tools were used, who did what and when, what qualifications and

1  
2  
3 experience the individuals had, and for instance, who made decisions and on what  
4 premises it is impossible to know the extents of the data, what was left out and why,  
5 and in general, how well the data suits for addressing other types of research questions  
6 than the original ones, and what needs to be taken into account when using it.” (Huvila,  
7 2022, p.31).  
8  
9

10 Whereas the need for and usefulness of paradata is undisputed and various technical  
11 solutions for capturing paradata are available (for an overview, see Huvila *et al.*, 2021;  
12 Börjesson *et al.*, 2020), a methodology for producing paradata is still in its infancy (Huvila  
13 2022). Recurring questions related to the use of paradata include: What information can  
14 function as paradata? Where can it be found? And how much capturing is enough? (Huvila,  
15 2022).  
16  
17  
18  
19

## 20 21 22 6.2. Capturing paradata

23 Paradata is captured in the form of narrative statements describing each action in a machine-  
24 readable and exchangeable format. However diverse and individual these narrative  
25 statements may seem, they fall within a certain range of viable options and share certain  
26 characteristics. In many cases, capturing is informed by the paratext as ‘contextual information  
27 provider’. Capturing paradata relies on a vocabulary and a set of rules that need to be  
28 elaborated (Huvila, 2022). In the upcycling workflow, paradata has a double scope.  
29  
30  
31  
32

33 On the one hand, paradata captures information about the methods employed by researchers  
34 in the past, i.e. information about the behaviour of previous generations of historians towards  
35 the raw data that they worked with. The paratext delivered with, or produced posthumously  
36 about the unFAIR HDC, informs the rescue, assessment and reconfiguration stages of  
37 upcycling and thus becomes part of the paradata.  
38  
39  
40

41 On the other hand, paradata also captures information about the methods employed in the  
42 present, i.e. information about the steps taken towards data rescue, assessment and  
43 reconfiguration. The latter kind of paradata is of essential importance for the reproducibility of  
44 the upcycling workflow. At least in part, it can be captured automatically, whereas the former  
45 kind requires a vocabulary, or a description language, and description rules to facilitate  
46 paradata analysis.  
47  
48  
49  
50

## 51 52 6.3. Analysing paradata

53 Upcycled HDCs comprise the ‘original’ HDC, the paradata produced during the iterative  
54 process of upcycling, the ‘new’ HDC and - if available - links to the original ‘raw’ data (see  
55 [figure 2](#)). Having been reconfigured in that way, upcycled HDCs may now serve as starting  
56 points for the study of historical ICPs using paradata analysis. Inspired by Gardin's logicist  
57 programme, a long-term project developed in archaeology since the 1970s (Gardin, 1999;  
58  
59  
60

1  
2  
3 Marlet *et al.*, 2019; Buard *et.al.*, 2020), paradata analysis aims at identifying the cognitive  
4 structure of HDCs.  
5

6 Paradata analysis studies the actions undertaken to rescue, assess and reconfigure unFAIR  
7 HDCs insofar as they had been captured during upcycling. The goal of paradata analysis is to  
8 further our understanding of historical information creation processes. To achieve this goal,  
9 paradata analysis relies on upcycled HDCs, which comprise both the 'original' [unFAIR](#) HDC,  
10 the documented upcycling process and the 'new' HDC. Therefore, upcycled HDCs can provide  
11 novel insights into two different ICPs (see figure 4).  
12  
13  
14  
15

16  
17 <<insert figure 4 here. Caption: The study of information creation processes using paradata  
18 analysis>>  
19  
20  
21

22 Starting from the [unFAIR](#) HDC, the first ICP documents how data were rescued and made fit  
23 for reuse. The first ICP comprises the workflows that digital historians and data scientists  
24 develop and implement to achieve an upcycled HDC. The second ICP documents how 'raw  
25 data' from primary sources was put into an HDC by previous generations of scientists. The  
26 second ICP targets their workflows. In principle, at least part of the information for the second  
27 ICP is found in the paratext of HDCs. However, even today, the practice of documenting  
28 decisions made during the process of collecting data is not a priority in the historical sciences,  
29 and the further we go back, the poorer the available documentation becomes. Therefore, the  
30 second ICP also entails producing a posthumous paratext during upcycling (see above). Thus,  
31 while rescuing an HDC and preparing it for reuse, valuable insights can be obtained about the  
32 methods applied by previous generations of scientists.  
33  
34  
35  
36  
37  
38

39 Paradata analysis can be understood as a condensation process that aims at unveiling the  
40 logico-semantic structure of data rescue, assessment and reconfiguration procedures  
41 implemented by 'upcyclers', on the one hand, and of data collection and systematisation  
42 procedures of previous generations, on the other hand. The result is a tree structure in which  
43 propositions (narrative statements) about the actions undertaken with the data are condensed  
44 into more general procedures, practices, methods and – eventually – cognitive processes. In  
45 doing so, paradata analysis sheds light on reasoning processes in the field of historical  
46 information science and contributes to the study of the historical methods that were - and still  
47 are – applied to 'raw' historical data sources and 'processed' historical data collections.  
48 Ultimately, the interpretation of these reasoning processes is the task of the digital historian,  
49 who - through his analysis of ICPs – contributes to the history of computational methods.  
50  
51  
52  
53  
54  
55  
56  
57

## 58 **7. Discussion**

59 My proposed upcycling workflow consists of three parts: input, processing and documentation,  
60

1  
2  
3 and output. Input consists of the unFAIR historical data collection and its accompanying  
4 paratext, which informs and supports the main data rescue processes: exploration,  
5 assessment and reconfiguration. Reconnection of the unFAIR historical data collection with  
6 its original 'raw' data supports these iterative processes. The implementation of these  
7 processes is captured in paradata. The output of upcycling is not a single product, like a book,  
8 or an article, or an electronic dataset, but an upcycled historical data collection, which  
9 comprises the 'original' collection, the paradata produced during upcycling, the 'new' collection  
10 and its reconnection to the original 'raw' data. Reconfigured in that way, the upcycled historical  
11 data collection serves as the starting point for the study of historical information creation  
12 processes based on paradata analysis. The latter entails a shift of focus from the results of  
13 ICPs, which we can find in historical data collections, to the cognitive processes that precede  
14 and accompany their creation. In the pursuit of this shift of focus, upcycling unveils its  
15 paradigmatic character.

16  
17 Upcycling's first contribution is to digital history, in general, and to the history of computational  
18 methods, in particular. Adopting a long-term view that also includes HDCs that precede the  
19 advent of computers, the study of historical information creation processes furthers our  
20 understanding of computational methods and their role in historical research. Upcycling in  
21 general, and paradata analysis in particular, address the intimate relationship between the  
22 history of computing and developments in historical methods, and thus contribute to a better  
23 understanding of the 'early history', or should we say the 'pre-history', of digital history. Indeed,  
24 through analysis of ICPs in HDCs, I hope to draw attention to two elements in the relationship  
25 between historians and computers that seem to have fallen off the radar in recent years. The  
26 first element is that of the continuities between 'analog' and 'digital' methods of historical  
27 research. The second element is that of the parallel development and application of  
28 computational methods in the humanities, on one hand, and the historical social sciences, on  
29 the other. Indeed, as early as 2011, the late Peter Haber already pointed out that the parallel  
30 development of computer-assisted research using historical data collections seems to have  
31 been lost in recent accounts of the history of 'digital humanities' and its predecessors (Haber,  
32 2011).

33  
34 Upcycling's second contribution is to the emerging field of digital hermeneutics. Upcycling  
35 responds to Andreas Fickers' call for an 'Update for Hermeneutics' in the form of a  
36 praxeological reflection on the impact of digital infrastructure, tools and sources on the practice  
37 of historical research (Fickers, 2020, p.2; Fickers *et al.*, 2022). Through its focus on the ICPs  
38 that precede and accompany historical research, upcycling contributes in a specific and  
39 hitherto under-exploited way to historical data criticism as an essential part of digital  
40 hermeneutics (Fickers, 2020, pp.6-7). In particular, upcycling suggests that ICPs are an  
41 important category in the practice of digital historians. Often, however, ICPs remain implicit or  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

are subsumed under the more general heading of 'data collection' in current data management plans (see e.g. Barats *et al.*, 2020). This, in turn, has a negative impact on scientific transparency, interoperability and, ultimately, sustainability of HDCs (Börjesson *et al.*, 2020).

## 8. Conclusions

Upcycling is an agenda for research in the field of digital history. In the preceding sections, the constituent parts of this agenda have been outlined. The next step should be to start implementing upcycling workflows in order to test and improve its conceptualization. To do so, it is of vital importance that the selected use cases are elaborated further. Moreover, it is essential to start working on a vocabulary and description rules for paradata capturing and analysis. And last but not least, following the example set by the National Agricultural Library in the US (Clarke & Shiue, 2020; Shiue *et al.*, 2021; see also Faniel & Zimmerman, 2011, pp.66-67), it is essential to start working on an assessment framework for selecting or not-selecting historical data collections for upcycling based on clearly defined feasibility requirements and expected value gains.

## References

- Avesani, P., *et al.* (2019), "The open diffusion data derivatives, brain data upcycling via integrated publishing of derivatives and reproducible open cloud services", *Scientific Data*, Vol. 69 No. 6, pp.1-13. <https://doi.org/10.1038/s41597-019-0073-y>
- Bang, N.E. & Korst, K. (Ed.s), *Tabeller over skibsfart og Varetransport gennem Øresund / Tables de la navigation et du transport des marchandises passant par le Sund 1497–1783*. Gyldendalske Boghandel – Nordisk Verlag, København / Otto Harrassowitz, Leipzig, 1906-1953.
- Barats, C., Schafer, V., & Fickers, A. (2020), "Fading Away... The challenge of sustainability in digital studies", *Digital Humanities Quarterly*, Vol. 14 No. 3. <https://www.proquest.com/scholarly-journals/fading-away-challenge-sustainability-digital/docview/2553557106/se-2>
- Baten, J., & Van Zanden, J.L. (2008), "Book production and the onset of modern economic growth", *Journal of Economic Growth*, Vol. 13 No. 3, pp.217-235.
- Bellingrad, D., & Salman, J. (2017), "Books and book history in motion: Materiality, sociality and spatiality", Bellingrad, D., Nelles, P., & Salman, J. (Ed.s), *Books in Motion in Early Modern Europe: Beyond production, circulation and consumption*. Palgrave Macmillan, Cham, pp.1-11.
- Bentkowska-Kafel, A., Denard, H., & Baker, D. (Ed.s) (2012), *Paradata and transparency in virtual heritage*, Ashgate, Farnham.

- 1  
2  
3 Beretta, F. (2021), "A challenge for historical research: making data FAIR using a  
4 collaborative ontology management environment (OntoME)", *Semantic Web*, Vol. 12 No. 2,  
5 pp.279-294. DOI: 10.3233/SW-200416  
6  
7  
8 Berg, M. (2007), "The Genesis of 'Useful Knowledge'", *History of Science*, Vol. 45 No. 2, pp.  
9 123–133. <https://doi.org/10.1177/007327530704500201>  
10  
11 Besomi, D. (Ed.) (2012), *Crises and Cycles in Economics Dictionaries and Encyclopedias*,  
12 Routledge, Abingdon / New York.  
13  
14 Bevan, A. (2015), "The data deluge", *Antiquity*, Vol. 89 No. 348, pp.1473-1484.  
15 doi:10.15184/aqy.2015.102  
16  
17 Birke, D., & Christ, B. (2013), "Paratext and Digitized Narrative: Mapping the Field",  
18 *Narrative*, Vol. 21 No. 1, pp.65-87. [doi:10.1353/nar.2013.0003](https://doi.org/10.1353/nar.2013.0003)  
19  
20 Boonstra, O., Breure, L., & Doorn, P. (Ed.s) (2006), *Past, present and future of historical*  
21 *information science*, DANS. 10.26530/OAPEN\_353255  
22  
23 Börjesson, L., Sköld, O., & Huvila, I. (2020), "Paradata in Documentation Standards and  
24 Recommendations for Digital Archaeological Visualisations", *Digital Culture & Society*, Vol. 6  
25 No. 2, pp.191-220. <https://doi.org/10.14361/dcs-2020-0210>  
26  
27 Buard, P.Y., Zadora-Rio, E., Chauveau, J., Roger, J., & Marlet, O. (2020), *Publishing an*  
28 *Archeological Excavation Report in a Logicist Workflow*.  
29 <http://dx.doi.org/10.15496/publikation-43222>  
30  
31 Buringh, E., & Van Zanden, J.L. (2009), "Charting the "Rise of the West": Manuscripts and  
32 Printed Books in Europe, A Long-Term Perspective from the Sixth through Eighteenth  
33 Centuries", *The Journal of Economic History*, Vol. 69 No. 2, pp.409-445.  
34 doi:10.1017/S0022050709000837  
35  
36 Cameron, S. & Richardson, S. (2005), *Using computers in history*, Palgrave Macmillan,  
37 Basingstoke / New York, 2005.  
38  
39 Carriero, V.A., Gangemi, A., Mancinelli, M.L., Nuzzolese, A.G., Presutti, V., & Veninata, C.  
40 (2021), "Pattern-based Design Applied to Cultural Heritage Knowledge Graphs", *Semantic*  
41 *Web*, Vol. 12, pp.313 – 357. 10.3233/SW-200422  
42  
43 Cherchi Usai, P. (2010), "The Conservation of Moving Images", *Studies in Conservation*,  
44 Vol. 55, pp. 250-257, DOI: 10.1179/sic.2010.55.4.250  
45  
46 Chiron, G., Doucet, A., Coustaty, M., Visani, M. & Moreux, J.-P. (2017), "Impact of OCR  
47 Errors on the Use of Digital Libraries: Towards a Better Access to Information", *2017*  
48 *ACM/IEEE Joint Conference on Digital Libraries (JCDL)*, pp.1-4, doi:  
49 10.1109/JCDL.2017.7991582  
50  
51 Christy, M., Gupta, A., Grumbach, E., Mandell, L., Furuta, R., & Gutierrez-Osuna, R. (2017),  
52 "Mass Digitization of Early Modern Texts With Optical Character Recognition", *Journal on*  
53 *Computing and Cultural Heritage*, Vol. 11 No. 1, pp.1-25. <https://doi.org/10.1145/3075645>  
54  
55  
56  
57  
58  
59  
60

- 1  
2  
3 Ciotti, G. & Lin, H. (Ed.s) (2016), *Tracing Manuscripts in Time and Space through Paratexts*,  
4 De Gruyter, Berlin / Boston. <https://doi.org/10.1515/9783110479010>  
5  
6 Clarke, C.T. & Shiue, H.S.Y. (2020), *Final Report and Recommendations of the Data*  
7 *Rescue Project at the National Agricultural Library*. <https://doi.org/10.13016/kpt7-cqgr>  
8  
9 Collins, R., & Skover, D. (2010), "Paratexts as Praxis", *Neohelicon*, Vol. 37 No. 1, pp.33-51.  
10 <https://doi.org/10.1007/s11059-010-0050-0>  
11  
12 Cordell, R. (2017), "'Q i-jtb the Raven": Taking Dirty OCR Seriously", *Book History*, Vol. 20,  
13 pp.188-225. [10.1353/bh.2017.0006](https://doi.org/10.1353/bh.2017.0006)  
14  
15 de Freitas, E. (2017), "The temporal fabric of research methods: Posthuman social science  
16 and the digital data deluge", *Research in Education*, Vol. 98 No. 1, pp.27–43.  
17 <https://doi.org/10.1177/0034523717723386>  
18  
19 [De Kramer, M. \(2022\), "3D models are easy. Good 3D models are not", Fickers, A.,](https://doi.org/10.1177/0034523717723386)  
20 [Tatarinov, J. \(Ed.s\), \*Digital History and Hermeneutics: Between Theory and Practice\*, De](https://doi.org/10.1177/0034523717723386)  
21 [Gruyter Oldenbourg, Berlin / Boston, pp. 239-258. https://doi.org/10.1515/9783110723991-](https://doi.org/10.1177/0034523717723386)  
22 [012](https://doi.org/10.1177/0034523717723386)  
23  
24 Denzel, M.A. (2002), "Handelspraktiken als wirtschaftshistorische Quellengattung vom  
25 Mittelalter bis in das frühe 20. Jahrhundert. Eine Einführung", Denzel, M.A., Hocquet, J.-C.,  
26 & Witthöft, H. (Ed.s), *Kaufmannsbücher und Handelspraktiken vom Spätmittelalter bis zum*  
27 *beginnenden 20. Jahrhundert — Merchant's Books and Mercantile Pratiche from the Late*  
28 *Middle Ages to the Beginning of the 20th Century*, Steiner Verlag, Stuttgart, pp.11-45.  
29  
30 Dobrin, L.M. & Schwartz, S. (2021), "The social lives of linguistic legacy materials",  
31 *Language Documentation and Description*, Vol. 21, pp.1-36.  
32 <http://www.e-publishing.org/PID/244>  
33  
34 Faniel, I.M., & Zimmerman, A. (2011), "Beyond the Data Deluge. A research agenda for  
35 large-scale data sharing and reuse", *International Journal of Digital Curation*, Vol. 6 No. 1,  
36 pp.58-69. <https://doi.org/10.2218/ijdc.v6i1.172>  
37  
38 Fickers, A. (2020), "Update für die Hermeneutik. Geschichtswissenschaft auf dem Weg zur  
39 digitalen Forensik?", *Zeithistorische Forschungen—Studies in Contemporary History*, Vol. 17  
40 No. 1, pp.157-168. <https://doi.org/10.14765/zzf.dok-1765>  
41  
42 Fickers, A., Tatarinov, J., & Heijden, T. van der (2022), "Digital history and hermeneutics –  
43 between theory and practice: An introduction", in: Fickers, A. & Tatarinov, J. (Ed.s), *Digital*  
44 *History and Hermeneutics. Between Theory and Practice*, De Gruyter Oldenbourg, Berlin /  
45 Boston, pp.1-19. <https://doi.org/10.1515/9783110723991>  
46  
47 Füssel, M. (2022), "Praxeologie als Methode", Haas, S. (Ed), *Handbuch Methoden der*  
48 *Geschichtswissenschaft*, Springer, Wiesbaden, pp.1-19. [https://doi.org/10.1007/978-3-658-](https://doi.org/10.1007/978-3-658-27798-7_18-1)  
49 [27798-7\\_18-1](https://doi.org/10.1007/978-3-658-27798-7_18-1)  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

- 1  
2  
3 Gardin, J.-C. (1999), "Archéologie, formalisation et sciences sociales", *Sociologie et*  
4 *sociétés*, Vol. 31 No. 1, pp.119–127. <https://doi.org/10.7202/001282ar>  
5  
6 Genette, G. (1997). *Paratexts. Thresholds of Interpretation*. Cambridge University Press,  
7 Cambridge.  
8  
9 Genette, G., & Maclean, M. (1991), "Introduction to the Paratext", *New Literary History*, Vol.  
10 22 No. 2, pp.261–72. <https://doi.org/10.2307/469037>  
11  
12 Gøbel, E. (2010), "The Sound Toll Registers Online project, 1497–1857", *International*  
13 *Journal of Maritime History*, Vol. 22 No. 2, pp.305–324.  
14  
15 Greenstein, D.I. (1989), "A Source-Oriented Approach to History and Computing: The  
16 Relational Database" *Historical Social Research / Historische Sozialforschung*, Vol. 14 No.  
17 3, pp. 9–16. <http://www.jstor.org/stable/20754411>  
18  
19 Guldi, J. & Armitage, D. (2014), *The History Manifesto*. Cambridge University Press,  
20 Cambridge. <https://doi.org/10.1017/9781139923880>  
21  
22 Gunn, S. & Faire, L. (Ed.s) (2016), *Research methods for history*. Edinburgh University  
23 Press, Edinburgh.  
24  
25 Haasis, L. & Rieske, C. (2015), *Historische Praxeologie : Dimensionen vergangenem*  
26 *Handelns*, Verlag Ferdinand Schöningh, Paderborn.  
27  
28 Haber, P. (2011), *Digital Past: Geschichtswissenschaft im digitalen Zeitalter*, Oldenbourg  
29 Wissenschaftsverlag, Berlin / Boston. <https://doi.org/10.1515/9783486712339>  
30  
31 Harlaftis, G. (2020), "Maritime history: A new version of the old version and the true history  
32 of the sea", *International Journal of Maritime History*, Vol. 32 No. 2, pp.383–402.  
33  
34 <https://doi.org/10.1177/0843871420924243>  
35  
36 Harvey, C. & Press, J. (Ed.s) (1996), *Databases in Historical Research*. Palgrave, London.  
37  
38 [https://doi.org/10.1007/978-1-349-24392-1\\_1](https://doi.org/10.1007/978-1-349-24392-1_1)  
39  
40 Heitman, C., Worthy, M., & Plog, S. (2017), "Innovation through Large-Scale Integration of  
41 Legacy Records: Assessing the "Value Added" in Cultural Heritage Resources", *Journal on*  
42 *Computing and Cultural Heritage* Vol. 10 No. 3, Article 17, 10 pages.  
43  
44 <https://doi.org/10.1145/3012288>  
45  
46 Herr, W. (2012), "VD 16,17 und 18: Retrospektive nationalbibliographische Verzeichnung vor  
47 dem Hintergrund sich verändernder technischer Rahmenbedingungen", *Perspektive*  
48 *Bibliothek* Vol. 1 No. 2, pp.106-136.  
49  
50 Huvila, I. (2022), "Improving the usefulness of research data with better paradata", *Open*  
51 *Information Science*, Vol. 6 No. 1, pp.28-48. <https://doi.org/10.1515/opis-2022-0129>  
52  
53 Huvila, I., Sköld, O. and Börjesson, L. (2021), "Documenting information making in  
54 archaeological field reports", *Journal of Documentation*, Vol. 77 No. 5, pp. 1107-1127.  
55  
56 <https://doi.org/10.1108/JD-11-2020-0188>  
57  
58  
59  
60

- 1  
2  
3 Jacob, M.C. (2014), *The first knowledge economy: Human capital and the European*  
4 *economy, 1750–1850*, Cambridge University Press, Cambridge.
- 5  
6 Jaillant, L. (2022), "Design Thinking, UX and Born-digital Archives: Solving the Problem of  
7 Dark Archives Closed to Users", Jaillant, L. (Ed.), *Archives, Access and Artificial Intelligence.*  
8 *Working with Born-Digital and Digitized Archival Collections*, Transcript, Bielefeld, pp. 83-  
9 107.
- 10  
11  
12 Jakobsson, U. (2021), "Digital Archaeological Archiving in Sweden: the Swedish National  
13 Data Service perspective", *Internet Archaeology*, Vol. 58. <https://doi.org/10.11141/ia.58.18>
- 14  
15 Jeannin, P. (1964), "Les comptes du Sund comme source pour la construction de l'activité  
16 économique en Europe (XVIe–XVIIIe siècle)", *Revue historique*, Vol. 121, pp.55–102 and  
17 307–340.
- 18  
19  
20 Jeannin, P., Kaiser, W. & Hoock, J. (Ed.s) (1991-2001), *Ars mercatoria. Handbücher und*  
21 *Traktate für den Gebrauch des Kaufmanns / Manuels et traités à l'usage des marchands,*  
22 *1470-1820. Eine analytische Bibliographie*, Ferdinand Schöningh, Paderborn.
- 23  
24  
25 Johansen, H.C. (1983), *Shipping and Trade between the Baltic and Western Europe 1784–*  
26 *95*. Odense: Odense University Press.
- 27  
28 Kelly, J.A., Farrell, S.L., Hendrickson, L.G., Luby, J. & Mastel, K.L. (2022), "A Critical  
29 Literature Review of Historic Scientific Analog Data: Uses, Successes, and Challenges",  
30 *Data Science Journal*, Vol. 21 No. 14, pp.1–11. <https://doi.org/10.5334/dsj-2022-014>
- 31  
32  
33 Kendall, E.F., & McGuinness, D.L. (2019), *Ontology engineering*, Morgan and Claypool.  
34 [10.2200/S00834ED1V01Y201802WBE018](https://doi.org/10.2200/S00834ED1V01Y201802WBE018)
- 35  
36  
37 Kramper, P. (2019), *The Battle of the Standards. Messen, Zählen und Wiegen in*  
38 *Westeuropa, 1660-1914*, De Gruyter, Berlin / Boston.
- 39  
40 Krünitz, J.G. (1773-1858), *Ökonomische Enzyklopädie oder allgemeines System der Staats-*  
41 *, Stadt-, Haus- und Landwirthschaft, in alphabetischer Ordnung*, [http://www.kruenitz1.uni-](http://www.kruenitz1.uni-trier.de/)  
42 [trier.de/](http://www.kruenitz1.uni-trier.de/)
- 43  
44  
45 Lodi, G., Asprino, L., Nuzzolese, A.G., Presutti, V., Gangemi, A., Recupero, D.R., Veninata,  
46 C., & Orsini, A. (2017), "Semantic Web for Cultural Heritage Valorisation", in: Hai-Jew, S.  
47 (Ed.), *Data Analytics in Digital Humanities. Multimedia Systems and Applications*, Springer,  
48 Cham, pp.3-37. [https://doi.org/10.1007/978-3-319-54499-1\\_1](https://doi.org/10.1007/978-3-319-54499-1_1)
- 49  
50  
51 MacFarland, K., & Vokes, A.W. (2016), "Dusting off the data: curating and rehabilitating  
52 archaeological legacy and orphaned collections", *Advances in Archaeological Practice*, Vol.  
53 4 No. 2, pp. 161-175. 10.7183/2326-3768.4.2.161
- 54  
55  
56 Marlet, O., Zadora-Rio, E., Buard, P.-Y., Markhoff, B., & Rodier, X. (2019), "The  
57 Archaeological Excavation Report of Rigny: An Example of an Interoperable Logician  
58 Publication" *Heritage*, Vol. 2 No. 1, pp.761-773. <https://doi.org/10.3390/heritage2010049>
- 59  
60

1  
2  
3 Martin, K., & Quan-Haase, A. (2017), "'A process of controlled serendipity": An exploratory  
4 study of historians' and digital historians' experiences of serendipity in digital environments",  
5 *Proceedings of the Association for Information Science and Technology*, Vol. 54 No. 1,  
6 pp.289-297.

7  
8  
9 [McCartney, L. \(2018\), "The Evolution and Issues Encountered in Providing Meaningful and  
10 Useful Online Access to Oral History Recordings", \*International Association of Sound and  
11 Audiovisual Archives \(IASA\) Journal\*, Vol. 47, pp. 30–40. <https://doi.org/10.35320/ij.v0i47.53>](#)

12  
13  
14 McDonough, W., & Braungart, M. (2013), *The upcycle: Beyond sustainability--designing for  
15 abundance*. Macmillan.

16  
17 McKenzie, D.F. (1999), *Bibliography and the Sociology of Texts*. Cambridge University  
18 Press, Cambridge.

19  
20 Meroño-Peñuela, A. and Hoekstra, R. (2014), "What is Linked Historical Data?", Janowicz,  
21 K., Schlobach, S., Lambrix, P., & Hyvönen, E. (Eds.), *Knowledge engineering and  
22 knowledge management. 19<sup>th</sup> international conference, EKAW 2014, Linköping, Sweden,  
23 November 24-28, 2014. Proceedings*, Springer, Cham, pp.282–287.

24  
25 Mokyr, J. (2002), *The Gifts of Athena. Historical Origins of the Knowledge Economy*,  
26 Princeton University Press, Princeton. <https://doi.org/10.1515/9781400829439>

27  
28 Noback, C. & Noback, F. (1851), *Vollständiges Taschenbuch der Münz-, Maass-, und  
29 Gewichtsverhältnisse, der Staatspapiere, des Wechsels- und Bankwesens, und der  
30 Usanzen aller Länder und Handelsplätze*, Brockhaus, Leipzig.

31  
32  
33 [Ricci, S. \(2008\), "Saving, Rebuilding, or Making: Archival \(Re\) Constructions in Moving  
34 Image Archives", \*The American Archivist\*, Vol. 71, pp. 433–55.  
35 <http://www.jstor.org/stable/40294525>.](#)

36  
37  
38 Richardson, D., Beedham, K., & Schofield, M.M. (1992), *Liverpool Shipping and Trade,  
39 1744-1786. A computerised edition of the Liverpool plantation registers* (Unpublished  
40 manuscript). <https://sp.ukdataservice.ac.uk/doc/2923/mrdoc/pdf/guide.pdf>

41  
42  
43 Ronald K., Collins, L., & Skover, D.M. (1992), "Paratexts", *Stanford Law Review*, Vol. 44 No.  
44 3, pp.509–52. <https://doi.org/10.2307/1228974>

45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65  
66  
67  
68  
69  
70  
71  
72  
73  
74  
75  
76  
77  
78  
79  
80  
81  
82  
83  
84  
85  
86  
87  
88  
89  
90  
91  
92  
93  
94  
95  
96  
97  
98  
99  
100  
101  
102  
103  
104  
105  
106  
107  
108  
109  
110  
111  
112  
113  
114  
115  
116  
117  
118  
119  
120  
121  
122  
123  
124  
125  
126  
127  
128  
129  
130  
131  
132  
133  
134  
135  
136  
137  
138  
139  
140  
141  
142  
143  
144  
145  
146  
147  
148  
149  
150  
151  
152  
153  
154  
155  
156  
157  
158  
159  
160  
161  
162  
163  
164  
165  
166  
167  
168  
169  
170  
171  
172  
173  
174  
175  
176  
177  
178  
179  
180  
181  
182  
183  
184  
185  
186  
187  
188  
189  
190  
191  
192  
193  
194  
195  
196  
197  
198  
199  
200  
201  
202  
203  
204  
205  
206  
207  
208  
209  
210  
211  
212  
213  
214  
215  
216  
217  
218  
219  
220  
221  
222  
223  
224  
225  
226  
227  
228  
229  
230  
231  
232  
233  
234  
235  
236  
237  
238  
239  
240  
241  
242  
243  
244  
245  
246  
247  
248  
249  
250  
251  
252  
253  
254  
255  
256  
257  
258  
259  
260  
261  
262  
263  
264  
265  
266  
267  
268  
269  
270  
271  
272  
273  
274  
275  
276  
277  
278  
279  
280  
281  
282  
283  
284  
285  
286  
287  
288  
289  
290  
291  
292  
293  
294  
295  
296  
297  
298  
299  
300  
301  
302  
303  
304  
305  
306  
307  
308  
309  
310  
311  
312  
313  
314  
315  
316  
317  
318  
319  
320  
321  
322  
323  
324  
325  
326  
327  
328  
329  
330  
331  
332  
333  
334  
335  
336  
337  
338  
339  
340  
341  
342  
343  
344  
345  
346  
347  
348  
349  
350  
351  
352  
353  
354  
355  
356  
357  
358  
359  
360  
361  
362  
363  
364  
365  
366  
367  
368  
369  
370  
371  
372  
373  
374  
375  
376  
377  
378  
379  
380  
381  
382  
383  
384  
385  
386  
387  
388  
389  
390  
391  
392  
393  
394  
395  
396  
397  
398  
399  
400  
401  
402  
403  
404  
405  
406  
407  
408  
409  
410  
411  
412  
413  
414  
415  
416  
417  
418  
419  
420  
421  
422  
423  
424  
425  
426  
427  
428  
429  
430  
431  
432  
433  
434  
435  
436  
437  
438  
439  
440  
441  
442  
443  
444  
445  
446  
447  
448  
449  
450  
451  
452  
453  
454  
455  
456  
457  
458  
459  
460  
461  
462  
463  
464  
465  
466  
467  
468  
469  
470  
471  
472  
473  
474  
475  
476  
477  
478  
479  
480  
481  
482  
483  
484  
485  
486  
487  
488  
489  
490  
491  
492  
493  
494  
495  
496  
497  
498  
499  
500  
501  
502  
503  
504  
505  
506  
507  
508  
509  
510  
511  
512  
513  
514  
515  
516  
517  
518  
519  
520  
521  
522  
523  
524  
525  
526  
527  
528  
529  
530  
531  
532  
533  
534  
535  
536  
537  
538  
539  
540  
541  
542  
543  
544  
545  
546  
547  
548  
549  
550  
551  
552  
553  
554  
555  
556  
557  
558  
559  
560  
561  
562  
563  
564  
565  
566  
567  
568  
569  
570  
571  
572  
573  
574  
575  
576  
577  
578  
579  
580  
581  
582  
583  
584  
585  
586  
587  
588  
589  
590  
591  
592  
593  
594  
595  
596  
597  
598  
599  
600  
601  
602  
603  
604  
605  
606  
607  
608  
609  
610  
611  
612  
613  
614  
615  
616  
617  
618  
619  
620  
621  
622  
623  
624  
625  
626  
627  
628  
629  
630  
631  
632  
633  
634  
635  
636  
637  
638  
639  
640  
641  
642  
643  
644  
645  
646  
647  
648  
649  
650  
651  
652  
653  
654  
655  
656  
657  
658  
659  
660  
661  
662  
663  
664  
665  
666  
667  
668  
669  
670  
671  
672  
673  
674  
675  
676  
677  
678  
679  
680  
681  
682  
683  
684  
685  
686  
687  
688  
689  
690  
691  
692  
693  
694  
695  
696  
697  
698  
699  
700  
701  
702  
703  
704  
705  
706  
707  
708  
709  
710  
711  
712  
713  
714  
715  
716  
717  
718  
719  
720  
721  
722  
723  
724  
725  
726  
727  
728  
729  
730  
731  
732  
733  
734  
735  
736  
737  
738  
739  
740  
741  
742  
743  
744  
745  
746  
747  
748  
749  
750  
751  
752  
753  
754  
755  
756  
757  
758  
759  
760  
761  
762  
763  
764  
765  
766  
767  
768  
769  
770  
771  
772  
773  
774  
775  
776  
777  
778  
779  
780  
781  
782  
783  
784  
785  
786  
787  
788  
789  
790  
791  
792  
793  
794  
795  
796  
797  
798  
799  
800  
801  
802  
803  
804  
805  
806  
807  
808  
809  
810  
811  
812  
813  
814  
815  
816  
817  
818  
819  
820  
821  
822  
823  
824  
825  
826  
827  
828  
829  
830  
831  
832  
833  
834  
835  
836  
837  
838  
839  
840  
841  
842  
843  
844  
845  
846  
847  
848  
849  
850  
851  
852  
853  
854  
855  
856  
857  
858  
859  
860  
861  
862  
863  
864  
865  
866  
867  
868  
869  
870  
871  
872  
873  
874  
875  
876  
877  
878  
879  
880  
881  
882  
883  
884  
885  
886  
887  
888  
889  
890  
891  
892  
893  
894  
895  
896  
897  
898  
899  
900  
901  
902  
903  
904  
905  
906  
907  
908  
909  
910  
911  
912  
913  
914  
915  
916  
917  
918  
919  
920  
921  
922  
923  
924  
925  
926  
927  
928  
929  
930  
931  
932  
933  
934  
935  
936  
937  
938  
939  
940  
941  
942  
943  
944  
945  
946  
947  
948  
949  
950  
951  
952  
953  
954  
955  
956  
957  
958  
959  
960  
961  
962  
963  
964  
965  
966  
967  
968  
969  
970  
971  
972  
973  
974  
975  
976  
977  
978  
979  
980  
981  
982  
983  
984  
985  
986  
987  
988  
989  
990  
991  
992  
993  
994  
995  
996  
997  
998  
999  
1000

\*\*\*

\*\*\*

- 1  
2  
3 Schöch, C., Van Dalen-Oskam, K., Antoniak, M., Jannidis, F., & Mimno, D. (2020),  
4 “Replication and Computational Literary Studies”, *Digital Humanities Conference 2020*  
5 (*DH2020*), Ottawa, Canada. <https://doi.org/10.5281/zenodo.3893428>  
6  
7  
8 Schubert, C. (2021), “Digital Humanities auf dem Weg zu einer Wissenschaftsmethodik:  
9 Transparenz und Fehlerkultur”, *Digital Classics Online* Vol. 7, pp.39-53.  
10  
11 <https://doi.org/10.11588/dco.2021.7.82371>  
12  
13 Shiue, H.S.Y., Clarke, C.T., Shaw, M., Hoffman, K.M., & Fenlon, K. (2021), “Assessing  
14 Legacy Collections for Scientific Data Rescue”, Toeppe, K., Yan, H., & Chu, S.K.W. (Ed.s),  
15 *Diversity, Divergence, Dialogue: 16th International Conference, iConference 2021, Beijing,*  
16 *China, March 17-31 2011, Proceedings, Part II*, Springer, Cham, pp. 308–318.  
17  
18 [https://link.springer.com/chapter/10.1007/978-3-030-71305-8\\_25](https://link.springer.com/chapter/10.1007/978-3-030-71305-8_25).  
19  
20 Skare, R. (2020), “Paratext”, *Knowledge Organization*, Vol. 47 No. 6, pp.511-519.  
21  
22 <https://doi.org/10.5771/0943-7444-2020-6-511>  
23  
24 Skare, R. (2021), “The paratext of digital documents”, *Journal of Documentation*, Vol. 77 No.  
25 2, pp.449–460. 10.1108/JD-06-2020-0106.  
26  
27 Smith, G. N. (2022), “The Promise and Peril of the Data Deluge for Historians”, *Journal of*  
28 *Cognitive Historiography*, Vol. 6, No. 1-2, pp.277–287. <https://doi.org/10.1558/jch.21156>  
29  
30 Sommer, D. (2010), “VD16, VD17, VD18: Diversität und Integration”, *ABI-Technik*, Vol. 30  
31 No. 2, pp.120-128.  
32  
33 Thaller, M. (2017), “Historical Information Science: Is There Such a Thing? New Comments  
34 on an Old Idea [1993]”, *Historical Social Research, Supplement*, Vol. 29, pp.260-286.  
35  
36 <https://doi.org/10.12759/hsr.suppl.29.2017.260-286>  
37  
38 Thieberger, N. (2018), “Research methods in recording oral tradition: Choosing between  
39 evanescence of the digital and senescence of the analog”, Ievenberg, I., Neilson, T.,  
40 Rheams, D. (Ed.s), *Research Methods for the Digital Humanities*. Palgrave Macmillan,  
41 Cham, pp. 233-241. [https://doi.org/10.1007/978-3-319-96713-4\\_13](https://doi.org/10.1007/978-3-319-96713-4_13)  
42  
43  
44 Thomas III, William G. (2004), “Computing and the historical imagination”, Schreibman, S.,  
45 Siemens, R., & Unsworth, J. (Ed.s), *A companion to digital humanities*, Malden / Oxford /  
46 Carlton, Blackwell Publishing, pp.56-68. <https://doi.org/10.1002/9780470999875.ch5>  
47  
48 Van Hulle, Dirk (2022), *Genetic Criticism: Tracing Creativity in Literature*, Oxford University  
49 Press, Oxford.  
50  
51 Vearncombe, J., Riganti, A., Isles, D., & Bright, S. (2017), “Data upcycling”, *Ore Geology*  
52 *Reviews*, Vol. 89, pp.887-893. <https://doi.org/10.1016/j.oregeorev.2017.07.009>  
53  
54  
55 \*\*\*  
56  
57 Wilkinson, M., et al. (2016), “The FAIR Guiding Principles for scientific data management  
58 and stewardship”, *Scientific Data*, Vol. 3. <https://doi.org/10.1038/sdata.2016.18>  
59  
60

1  
2  
3 Zimring, C.A. (2016), "Upcycling in History: Is the Past a Prologue to a Zero-Waste Future?  
4 The Case of Aluminum", *RCC Perspectives*, Vol. 3, pp.45-52.

5  
6 <http://www.jstor.org/stable/26241375>

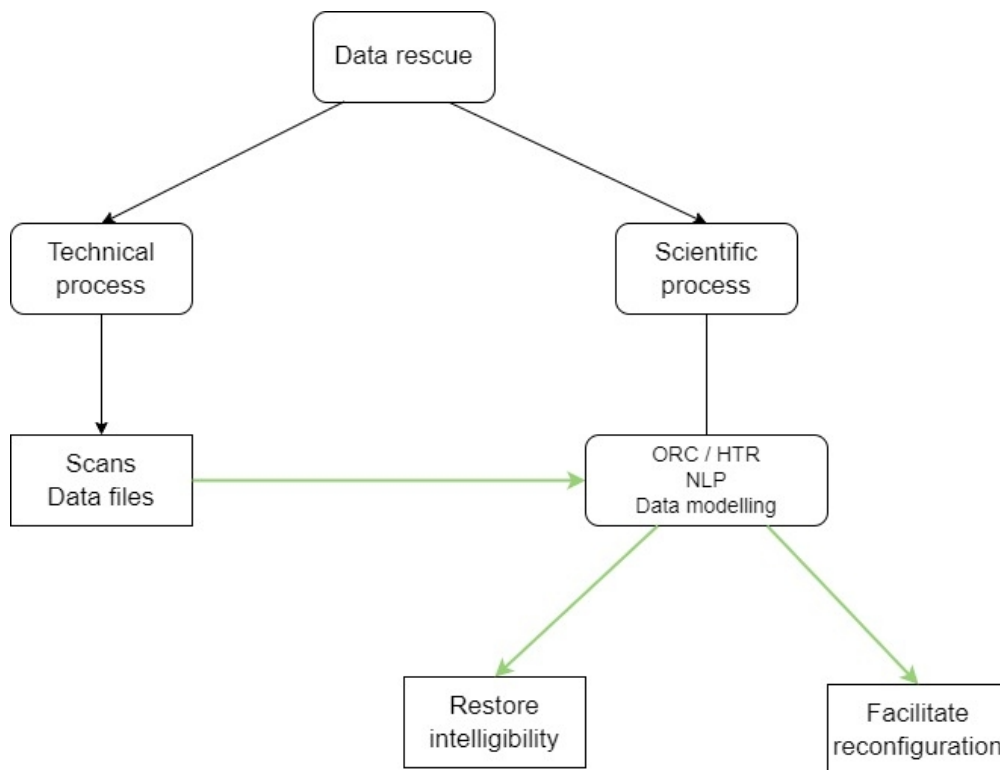
7  
8 Zupko, R.E. (1985), *A Dictionary of Weights and Measures for the British Isles: The Middle*  
9 *Ages to the twentieth century*. American Philological Society, Philadelphia.  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

Journal of Documentation

|                       |                             |                              |
|-----------------------|-----------------------------|------------------------------|
|                       | <i>Historical metrology</i> | <i>Maritime history</i>      |
| <i>Pre-1950 HDCs</i>  | Noback & Noback 1851        | Sound Toll Tables            |
| <i>Post-1950 HDCs</i> | Zupko 1985                  | Liverpool Shipping and Trade |
|                       | <i>Reference works</i>      | <i>Legacy collections</i>    |

Table 1: Overview of example HDCs in the text.

Journal of Documentation

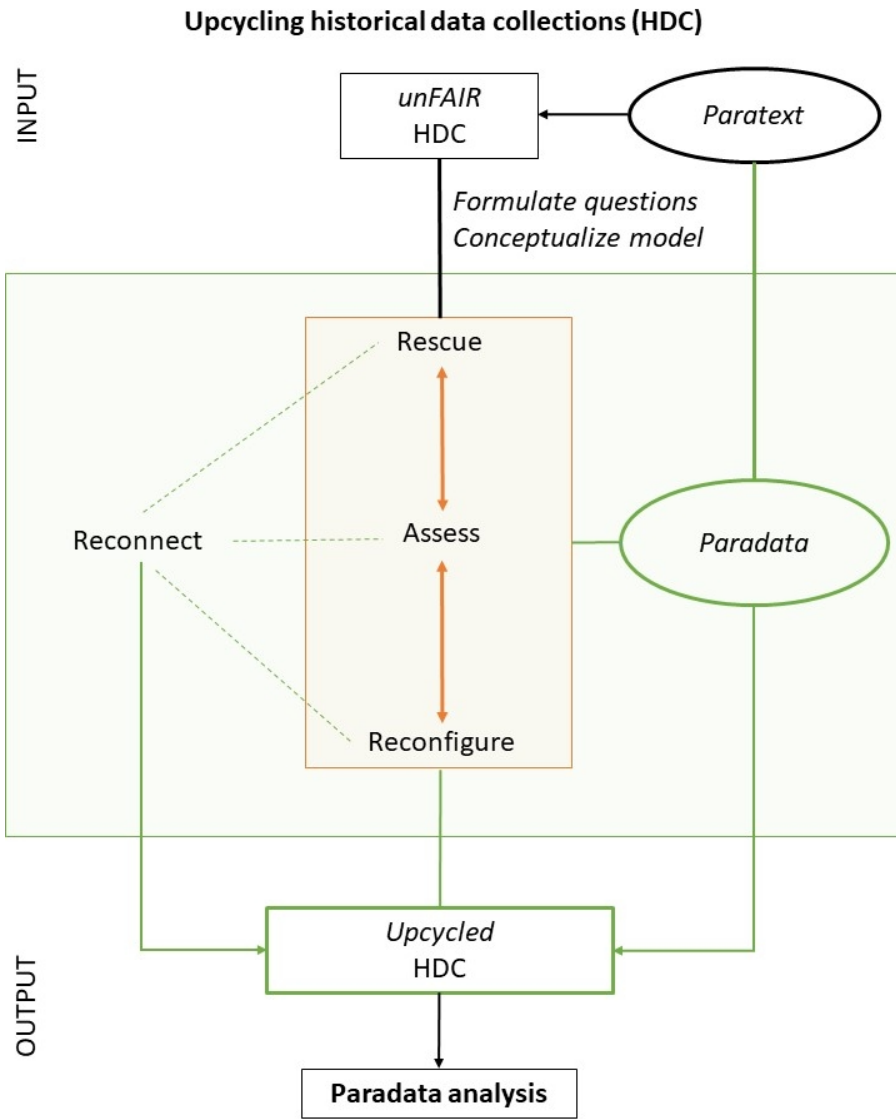


Value gains from data rescue

233x178mm (72 x 72 DPI)

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

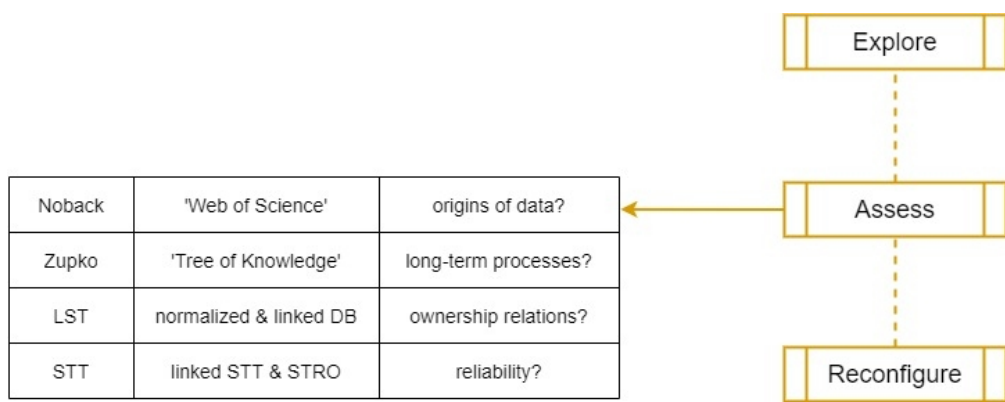
1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60



The upcycling workflow

190x254mm (96 x 96 DPI)

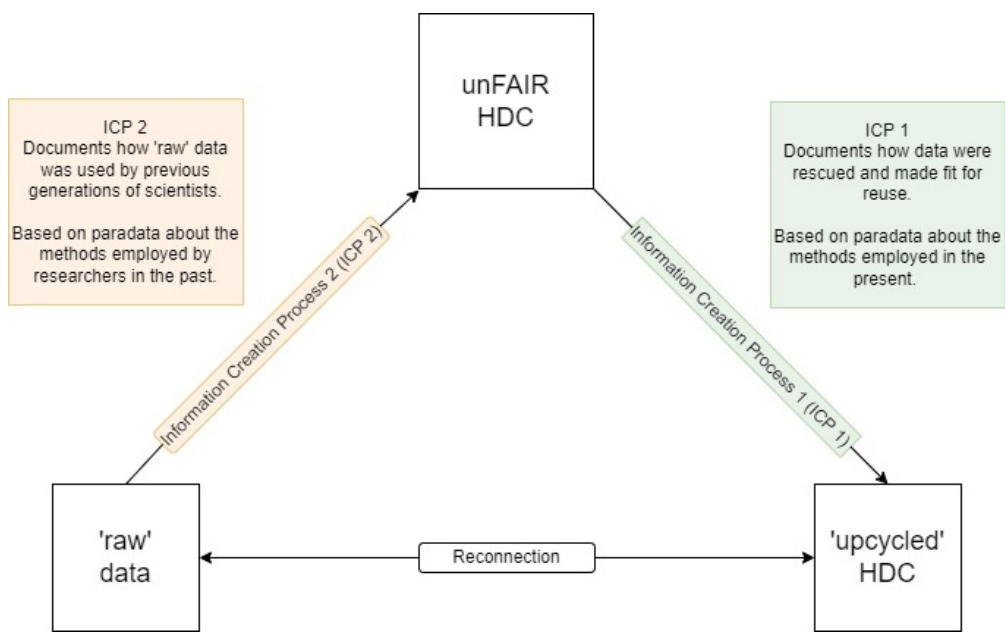
1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60



Applications of the scenario-based method

254x99mm (72 x 72 DPI)

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60



The study of information creation processes using paradata analysis

240x148mm (72 x 72 DPI)