Data Showcases: the Data Journal in a Multimodal World

Leen Breure

ScieMedia

Peter Doorn

DANS

Hans Voorbij

Poperinge

Abstract

As an experiment, the Research Data Journal for the Humanities and Social Sciences (RDJ) has temporarily extended the usual format of the online journal with so-called 'showcases', separate web pages containing a quick introduction to a dataset, embedded multimedia, interactive components, and facilities to directly preview and explore the dataset described. The aim was to create a coherent hyper document with content communicated via different media (multimodality) and provide space for new forms of scientific publication such as executable papers (e.g. Jupyter notebooks). This paper discusses the objectives, technical implementations, and the need for innovation in data publishing considering the advanced possibilities of today's digital modes of communication. The data showcases experiment proved to be a useful starting point for an exploration of related developments within and outside the humanities and social sciences. It turns out that small-scale experiments are relatively easy to perform thanks to the easy availability of digital technology. However, real innovation in publishing affects organization and infrastructure and requires the joint effort of publishers, editors, data repositories, and authors. It implies a thorough update of the concept of publication and adaptation of the production process. This paper also pays attention to these obstacles to taking new paths.

Keywords: coherence, data preview, data reuse, executable papers, hypertext, interactivity, multimedia, multimodality, supplementary files

International Journal of Digital Curation
2022, Vol. 17, Iss. 1, 24 pp.

1

http://dx.doi.org/10.2218/ijdc.v17i1.789
DOI: 10.2218/ijdc.v17i1.789

Introduction

An Exhibit of Datasets

The *Research Data Journal for the Humanities and Social Sciences* (RDJ) is a peer-reviewed journal, which is designed to comprehensively document and publish deposited datasets and facilitate their online exploration. RDJ is e-only and open access; it focuses on research across the humanities and the social sciences. RDJ was founded by DANS (Data Archiving and Networked Services – the major Dutch research data archive) in 2016, in collaboration with the academic publisher Brill. From 2021 onwards, it is also supported by CESSDA, the Consortium of European Social Science Data Archives.

The goal of starting RDJ was to make datasets more accessible to users, especially to those who are not experts in the research area from which the data originated. In addition, a publication in RDJ gives credits to the depositors and offers the opportunity of linking the dataset to related publications elsewhere. Data repositories document datasets through metadata and technical documentation, and some additionally have a limited data preview option. RDJ is set up to offer something more. Authors are also asked to describe the background of the dataset in the context of the research project and, if possible, to provide visuals: photos, videos, graphs, maps, and the like. Conclusions concerning a certain research problem as in an ordinary academic paper are permitted but not required. The intended length of the data paper was relatively short, although many authors exceeded the maximum length in practice.

RDJ is published according to the generally accepted format of a digital journal: a web page with static images while all other multimedia material is stored in repositories such as Figshare and is accessible via hyperlinks. Thus, although all supplementary files are available to the reader, this separation of text from relevant additions is unsatisfactory, for reasons that we will discuss in detail below. In short, we believe that integration of multimedia and other interactive content with the main text contributes to a more coherent representation of the data in the context in which they were created. By making it easier to preview the data and by presenting visualizations of different types, the user can directly explore the dataset itself in conjunction with reading the text.

That is why we conducted a short-term (two-year) experiment: we extended the data journal with an accompanying website, the **Exhibit of Datasets**[1], where a selection of datasets got their showcases. A *showcase* in this context is a separate web page containing a quick introduction to a dataset, longer than an abstract but limited to essential information taken from the data paper and compiled by an editor. Multimedia are embedded in the text and have interactive features where possible and appropriate. Tabular datasets can be explored in a linked online spreadsheet (read-only, but with the option of selecting and sorting), or via a link to the online project's database (if available), or through an embedded widget with a data sample. The showcases were not limited to typical technical data papers, but also turned out to be a good fit for papers that focussed more on the research itself conducted with the data set.
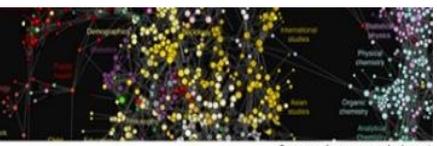
In the next section, we will discuss the objectives of the data showcases and, after that, the technical infrastructure required for integration. A discussion of the results and a conclusion follow at the end. The duration of the project was rather short for gauging user acceptance representatively. Therefore, we must suffice with the lessons learned and with the occasional responses of authors over time.

---

[1] Exhibit of Datasets: https://dansdatajournal.nl/rdp/ (retrieved: June 14, 2022). The basis of each showcase is the data paper with the corresponding dataset(s) deposited in the DANS data archive or in another trusted repository. Like the data papers, the showcases contain explicit references to the DOIs of the datasets.

Figure 1. Home page of the Exhibit of Datasets.

Objectives

Coherence

A conventional digital academic paper consists of text,[2] images and tables; all further elements belong to the category of supplementary files, which are stored elsewhere, sometimes at the publisher's server or in a digital repository. This is analogous to books in which all the illustrations are in a separate plate section rather than on the pages where they are discussed. If non-textual elements are significant to the argument, embedding them in the text creates more opportunities to clarify the story. Two examples: (1) An author writing about movies may want to comment on certain scenes. Then it is not convenient for the reader to be referred to a repository and scroll to the relevant fragment in the film with the help of a video timestamp. If the video is embedded in the text, the scene can be displayed automatically by a simple mouse click. The same applies to fragments in audio clips. (2) Static images are generally less suitable for making complex information visible: interactive charts can display hidden values when the reader moves the mouse over the curve; clickable maps are far easier to read than their overloaded counterparts in historical atlases; and panning and zooming are indispensable with crowded timelines.

Although more types of interactive multimedia can be embedded fairly easily in an HTML page, for technical reasons other content will still end up on separate web pages, which are accessible from the main text through hyperlinks. The reader should experience this network of nodes as a coherent whole, as a single *hyper document*, and not as an e-document with some non-descriptive material dangling somewhere in the cloud. Much research has been done on networked hypertext and comprehension; the results are inconclusive, but there is a strong indication that hypertexts with more nodes are more challenging to read than linear texts. Most studies show that the greater the number of hyperlinks, the greater the cognitive workload placed on the reader's working memory (Blom et al., 2018). However, coherence seems to be much more important, in particular for those readers with little domain knowledge (Herrada-Valverde & Herrada-Valverde, 2017).

The concept of text coherence was developed for linear text, but coherence-building is particularly relevant for hyper documents because of the inherent selective reading mode. Users may read the central node first, but are completely free in what they choose next. In the design of the showcases we have addressed this issue by consistently adding usability cues (Storrer, 2002):

1. *Local context cues*: hyperlinks in the form of screenshots of a data preview web page and other distinctive graphic components such as buttons to access supplementary files and deposited data. All of them give marked hints about the node to which the hyperlink connects and guide the user's expectations. Layout is important here. By placing these cues on the showcase page in a well-organized manner, the reader gets an impression of all the information available in the entire hyper document at a glance.

2. *Global context cues*, which help the user to identify a certain web page as part of the hyper document. This includes obvious things like a banner that makes it clear that the page is part of RDJ, an explanation of the content (in particular for visitors from outside the Exhibit), and navigation controls that return the user to the showcase itself.

Coherence also implies clear *pathways*, from a global introduction descending into more detailed content: a *layered structure*, at the top a *hub* such as a showcase, from which the user can go to deeper layers, for example to a full research paper about the project, to videos, and to the

---

[2] We consider elements such as references, hyperlinks, footnotes, acknowledgements, etc. as belonging to the text, and these are not particularly relevant for the discussion here. We pay attention to metadata in section *'Diversity of users and gaining trust'*.

dataset itself. The basic idea is not to overload the readers with information right away, but to allow them to choose it when they are prepared for it. Referring to one large appendix of separate files in which the users must find their way does not work; supplementary information has to be carefully linked to corresponding passages in the main text.

Multimodality

Standard academic publishing is in strong contrast to our daily communication: a Whatsapp chat is a mixture of words, smileys, photos, videos, and audio messages; video, audio, and interactive maps play a dominant role on any news site. In social media videos are cut and fragments are remixed, re-contextualized, and combined with emoji to reinforce the message and to better express the emotional charge (Esposito & Zollo, 2021). We live in a multimodal world; we use a complex of means or "modes" to make meaning. Multimodality refers to the interplay between different modes of representation, for instance, between images and written text, and multimodal literacy refers to awareness and effective use of this range of modalities (Bezemer & Kress, 2008; Jewitt, 2013; Korhonen, 2010; Kress & Van Leeuwen, 2001).

This cultural trend, supported by the wide availability of powerful mobile devices, is also impacting the practice of scientific research. In several disciplines, multimodal output has begun to emerge:

- Despite many reservations, some non-traditional multimedia dissertations have already been completed. Lamb (2019a) made a Ph.D. study about how learning spaces in higher education are being affected by the shift to the digital. He felt a strong need for multimodality: "The challenge I faced was to present audio and visual content in juxtaposition with written argumentation" (Lamb, 2019b). He solved the problem of coherence by inserting QR-codes in the PDF text. By pointing the smartphone camera at a specific code, the user can open the relevant multimedia on Vimeo or Thinglink.

- Visconti (2015) published an award-winning dissertation on James Joyce's *Ulysses* called *Infinite Ulysses*, fully online and interactive, exploring and demonstrating how this literary work can be annotated with infinite interpretations, questions, and contextualizations.

- Other examples are about music genres like hip-hop and rap or YouTube videos. One of the authors explained that she continuously communicated with visual imagery and wanted to practice this process in her Ph.D. (details about these dissertations in Tran, 2019).

Within the research areas covered by RDJ, there are some fields where the need for the combination of text and multimedia data is particularly strongly felt. Three examples follow.

Oral history

In oral history**,** transcripts are important for analysis and a quick review of interview content, but listening to, or even better, watching the interlocutors, allows for much more empathy and a better understanding. DANS holds a large collection of audiovisual witness stories from the second world war. We published a showcase about the story told by two brothers who were victims of a German *razzia* in Rotterdam in 1944. The concise showcase text contains hyperlinks to passages in the video, allowing the reader to switch from the summary to experiencing the corresponding part of the interview (de Jager, 2016). The UK Data service published a large collection of life story interviews with pioneers of social research as transcripts on its website and made the videos available on YouTube (Corti-Georgiou, 2019). They can easily be made accessible directly from the text via hyperlinks and, if required, even at the level of a specific video fragment by appending the corresponding timestamp to the end of the link.

Anthropology

Multimodality is making remarkable progress in anthropology projects. In 2020, the leading journal *American Anthropologist* launched a special multimodal section, curating material that moves beyond the written word and welcoming submissions that utilize diverse media forms, including (but not limited to) films, photo essays, graphic art, soundscapes, and social media threads (Astacio et al., 2020). This is nothing new to this discipline, but rather an explicit recognition of the phenomenon. Anthropologists have been experimenting with different forms of media technologies throughout the twentieth century, but legitimization still must be fought for (Collins et al., 2017). Until a decade ago, traditional ways of researching with a notebook and a pen dominated the research practice, which might explain the resistance to change, to trying novice approaches, and to showcasing multimodal research (Gill, 2019). Moreover, multimodal anthropology has also led to a critical, ethical discussion about the political and societal context in which the new technologies are embedded (Takaragawa et al., 2019).

Archaeology

Archaeological datasets generally consist of extensive reports and carefully compiled lists of finds that have been photographed, drawn, mapped, digitized, measured and described in detail. In this format, it is not easy both for outsiders and for fellow-archaeologists to form a concrete picture of the context (e.g. societal, environmental) from which the material originates. But archaeology also has a significant public face. The monumentality of archaeological sites and the artistic quality of major finds have always sparked public interest and imagination, and digital technologies have added an extra dimension (Huvila, 2014). A well-known example is the website *Jamestown Rediscovery*, which has brought to life the stories of early James Fort in Jamestown, Virginia, USA.[3] As archaeology is both a scientific discipline and an important segment of the cultural heritage sector, archaeologists want to do more than just showing and documenting artifacts. They strive to engage audiences using different digital media like video, 3D models, and augmented reality to let users experience the process of excavation and interpretation of finds (Interactive Digs, 2019; Unger et al., 2020; Wakefield, 2020). Multimodality has reached great heights in a system as VSim. It facilitates the real-time exploration of highly detailed, three-dimensional computer models in various educational settings. Sources and narrative text can be embedded in the 3D model exactly where they belong, which ensures optimal coherence.[4]

Commercial motives also play an important role. Tourists can immerse themselves in the past through augmented reality when a 3D reconstruction is projected over an existing excavation (Corallo et al., 2017). The general public is even better reached through *archaeogaming*, where cultural heritage knowledge and data are intertwined with video games. It has two research aspects: archaeology *in* video games, but also on a meta-level the *study of* the games themselves (Mol et al., 2017; Reinhard, 2018, pp. 2-3). Digital games are considered as archaeological sites; landscapes, artifacts, and the game spaces held within those media are also understood archaeologically as digitally built environments containing their own material culture.

However, all this does not mean that archaeological data papers have undergone a dramatic metamorphosis. The Journal of Open Archaeology[5] suffices with short, sparsely illustrated data papers, hardly more than announcements of a dataset with metadata. Other journals show a mixed picture, for example Internet Archaeology.[6] Most of the submissions are conventional papers with illustrations where required, but this journal also offers the possibility

---

[3] https://historicjamestowne.org/ (retrieved: June 15, 2022).
[4] VSim: [Video]. Institute for Digital Research and Education. YouTube. https://www.youtube.com/watch?v=MYYtDTucKL8 [See also the corresponding website: Institute for Digital Research and Education, UCLA (n.d.). VSim. https://idre.ucla.edu/research/active-research/vsim] (retrieved: June 15, 2022).
[5] Journal of Open Archaeology: https://openarchaeologydata.metajnl.com/ (retrieved: June 14, 2022).
[6] Internet Archaeology: https://intarch.ac.uk/ (retrieved: June 14, 2022).

to include embedded videos and interactive multimedia (Graham, 2020; Loy et al., 2021; Richards & Roskams, 2013).

Executable papers[7]

A next step in the direction of transparent research is providing full insight into the processing of the data through so-called *executable papers*, dynamic pieces of software that combine text, raw data and the program code used for the analysis, with which a reader can interact:

> 'In principle the idea is to enable the reader to reproduce each and every step taken to arrive at a publication's conclusions, from the raw data to the polished plot. Therefore, the paper should be linked to a repository containing the raw data and should support the necessary functionality to make the research process transparent and reproducible. Specifically, the executable paper should display nicely formatted text, including references and links (just like a journal article), as well as figures, plots, and possibly also videos and interactive elements. Additionally, the analysis code used for creating plots from data should be displayed and interpreted by the executable paper, also allowing for user input to modify the analysis. Lastly, the executable paper should be composed of completely Open Source components and hosted under a free license to be easy to share and reuse' (Lasser, 2020).

Executable papers are also known as *notebooks*, for instance, the *Jupyter Notebook*,[8] which is an open-source web application that allows the creation and sharing of documents that contain live code (in Python, R and many other languages), equations, visualizations, and narrative text. Microsoft offers a similar cloud solution in the form of Azure Notebooks, which is a 100% cloud implementation of Jupyter Notebooks and can import data from other sources like Google Drive and OneDrive. Google started the Colaboratory (Colab for short), which allows the author to write and execute Python directly in the browser.[9]

This form of research reporting is particularly popular in sciences with a substantial amount of computation, although this genre of literate programming, which combines code, text, and execution, is mainly criticized from the computer science side (Pimentel et al., 2019). Problems with the format, which can make it hard to reproduce a notebook, are partly met by a plugin as Jupytext, that can save notebooks as Markdown files and export scripts in many languages.[10] Executable papers are also penetrating the social sciences, for example, through the statistical package RStudio, which provides its own notebook (Xie et al., 2021). The executable paper does not fit well with conventional academic journals because of its dynamic and interactive nature, but a few publishers have started or intend to experiment with it (Brandt, n.d.). Currently, the challenge of dynamic storage is solved by using version control in repositories like GitHub, which is projected to have 8 million notebooks archived in 2021 (O'Shea, 2019).

However, it is far from easy to make a notebook's programming language work locally on someone's machine. Fortunately, online hosting platforms like *JupyterHub*[11] and *Binder*[12] are available for interpreted programming languages like Python. GitHub provides a simple interface with Binder, so that readers can execute the code without downloading any data or installing additional software (Lasser, 2020). Compared to conventional publications, the executable paper has the great advantage that users can explore the data themselves with different parameters and thus assess the researcher's argument much more critically. In our Exhibit of Datasets, we could not yet realize a dynamic working experience with the few notebooks submitted. Instead, we fell back on a simpler, static alternative: a notebook converted

---

[7] We graciously acknowledge the contribution to this section by Dirk Roorda (DANS).
[8] Jupyter project: https://jupyter.org/ (retrieved: June 14, 2022).
[9] Google's Colaboratory: https://colab.research.google.com/notebooks/intro.ipynb (retrieved: June 14, 2022).
[10] Jupytext: https://jupytext.readthedocs.io/en/latest/ (retrieved: June 14, 2022).
[11] JupyterHub: https://jupyter.org/hub and https://tljh.jupyter.org/en/latest/ (retrieved: June 14, 2022).
[12] Binder: https://gke.mybinder.org/ (retrieved: June 14, 2022).

to an HTML page, linked to the corresponding data showcase (e.g., Baciocchi et al., 2019; Roorda, 2018).

Figure 2. Predicted number of Jupyter Notebooks in Github in 2021 (O'Shea, 2019).

Exploring the dataset

Diversity of users and gaining trust

Data consumers are a diverse group that overlaps only partially with the data depositors. They involve students, researchers, museum curators, employees of private companies, government employees, et cetera (Borgman et al., 2019; Doorn, 2020; Gregory et al., 2020). Therefore, it seems a good idea to differentiate data interfaces and to tailor them to the interest, knowledge, and skills of the target audience. A document such as a data showcase may be useful for non-experts, possibly without academic background and with less profound interests, but also for researchers from different disciplines, who will appreciate a brief overview with a practical explanation of the data first, before going into detail.

Digital research data and related outputs must be FAIR, i.e., Findable, Accessible, Interoperable, and Reusable. This implies three basic conditions – researchers must be able to:

1. *locate* data sources (the "F" of FAIR);

2. *retrieve* the relevant data (the "A" of FAIR); and

3. *evaluate* the data quality in relation to their own research requirements and methodological criteria. In other words, potential users must gain *trust* in the data. This implies access to contextual information to understand the research problem, data gathering and recording. Also, the data must be available in a format that is compatible with a researcher's tools, techniques, and analysis applications (the "I" and "R" of FAIR) (Rieger, 2019; Sun & Khoo, 2016).

The "FAIRness" of a dataset depends not only on the characteristics of the data itself but also on the quality of its metadata.[13] There are numerous approaches to assessing the FAIRness of a dataset. Some of these are based on automated procedures, focusing on the formal properties of the data and metadata, such as the F-UJI tool,[14] developed in the FAIRsFAIR

---

[13] How to FAIR: https://howtofair.dk/how-to-fair/ (retrieved: June 13, 2022).
[14] F-UJI tool: https://www.f-uji.net/

project and FAIRsharing's FAIR Evaluation Services.[15] Unfortunately, the current state of affairs is that different automated tools provide widely diverging FAIR assessment scores, as a recent report has demonstrated (European Commission, Directorate-General for Research and Innovation, 2022).[16] In our view, several of the FAIR principles are intrinsically qualitative and cannot (yet?) be evaluated automatically. Examples are (*italics* by the authors, indicating qualitative aspects of the principles):

- Findability 2: Data are described with *rich* metadata;

- Reusability 1: (Meta)data are *richly* described with a plurality of *accurate and relevant* attributes;

- Reusability 1.3: (Meta)data meet *domain-relevant* community standards.[17]

In our opinion, what is "rich" (meaning extensive in this context), "accurate", and "(domain) relevant" cannot be objectively established, but require a judgement which can at best be intersubjective when multiple peer reviewers inspect the (meta)data. In practice, of course, this is easier said than done, mainly because as yet there seems to be both little expertise and little experience among reviewers in how to judge such things. For the time being, RDJ experimented with a "manual" and qualitative approach, using a structured FAIR assessment form (further explained in the section *'From data preview to data review'*). Probably, a combination of automated and manual FAIR-evaluation approaches is the way forward here for the future.

These considerations have no bearing on the Data Showcases themselves. The metadata belonging to the data described in the showcases are stored and made available in a trustworthy repository. The assessment of the data and metadata is (or should be) part of the reviewing process of data papers in the Data Journal.

Some researchers trust datasets commonly used in their fields or those used in peer-reviewed journals, others will highly value easy access to information about data collection and analysis. However, minor errors in the data or in the accompanying documentation will make researchers suspicious. So, if errors and problems can be recognized quickly, trust is more likely to be created (Gregory et al., 2020). Easy access to data and a smooth start are critical to successful data reuse. Access may be a greater barrier for new or novice researchers with little or no experience in data reuse. Unfamiliar data formats, software, or the requirement of special analytical programs may also hinder reuse (Pasquetto et al., 2019; Yoon, 2016, 2017). Therefore the data preview facility in the showcase, with the option to explore at will, can help overcome these hurdles and complement automatically generated (limited) previews offered by certain repository systems.

Data preview
The implementation of a preview facility in our showcases strongly depends on the type of data:

1. For *spreadsheet* data, we have resorted to cloud computing: a screen capture in the showcase with a link to an online version being a Google[18] or Zoho spreadsheet.[19] The user is then able to experiment with the dataset without permanently changing anything.

2. Relational *databases* are usually deposited as an SQL dump, a very large text file that cannot reasonably be consulted in that form. If the database with query facilities is

---

[15] FAIR Evaluation Services: https://fairsharing.github.io/FAIR-Evaluator-FrontEnd/#!/

[16] The supplementary materials belonging to this report contain further details, especially section 5.4. "Considerations about FAIR assessment tools" of *European Research Data Landscape Study Report* (deliverables 3.2, 4.2, 5.2), p. 152 ff. (Nov. 23, 2022), https://doi.org/10.5281/zenodo.7351121 (retrieved: Dec. 1, 2022).

[17] See https://www.go-fair.org/fair-principles/ for a full list of the principles.

[18] Google spreadsheets: https://www.google.com/sheets/about/ (retrieved: June 14, 2022).

[19] Zoho spreadsheets: https://www.zoho.com/sheet/ (retrieved: June 14, 2022).

available online, we added a hyperlinked screen capture. If not, we have loaded the data into a desktop DBMS (e.g., the popular SQLite[20]), created queries based on the data paper, and published the results as a spreadsheet or as a scripted HTML listing embedded in the showcase (e.g., van Nieuwkerk et al., 2020).

3. A special category is *multimedia* data in, for instance, archaeological projects. Images and videos insofar as directly relevant to the storyline of the paper were embedded in the showcase using standard HTML code or − with larger numbers of images − using a compact component as a slideshow. In this way, the voluminous image material was moved to a second layer below the main text.

However, some datasets are so complex that an outsider cannot immediately oversee and explore them, and a simple data preview is not feasible. In such cases, we visualized the most important characteristics of the data, for example, through charts, graphs, and data models, usually in collaboration with the author.

From data preview to data review

Creating a data preview can help improve data review. Assessment of the dataset has been a standard part of the RDJ data paper review, but in practice most of the reviewers' attention appeared to be focused on the text. This is not surprising, because of the extra effort, both technical and in terms of content, required for a thorough examination of the data. We have experimented with a FAIR data review form,[21] which however focused more on the formal characteristics of the data (and metadata) than on their value to the research area.

In making data explorable, we discovered that reusing downloaded datasets is not always as straightforward as one may expect. From time to time, we encountered flaws and hitches that required extra technical knowledge that the average user will not normally have (e.g., character encoding problems and other quirks of relational database systems). To help the user get off to a smooth start, we decided to include the necessary tips in a short *Quick Start* section on data reuse where appropriate.

Creating these Quick Start tips inevitably raised questions like "what is the data structure?", "is there an explicit data model?"", "how is the data model implemented?", "how consistent and careful did the author / data creator work?". So, unintentionally, an informal data review has crept in through the back door. The current preview is no more than a *browse* function; it is worth considering extending the preview facility further with a limited number of *queries*. This would not only support new users, but could also facilitate the data review. In many cases, such questions can be deduced from the data paper itself, but it may be a better idea to ask the author to submit a few representative sample questions, if possible in the form of queries.

Technical Aspects − Look and Feel

To present a data showcase well, it should be easy to read on screens of all sizes, from a wide desktop monitor to a small smartphone (i.e., meet the requirements of responsive design), and attract and retain attention by offering an appealing view. Apart from the usual HTML and CSS code, the Exhibit's technical infrastructure is based on the Bootstrap[22] framework and JavaScript for interactivity, animation, and website management. To have maximum freedom to experiment with page layout, we avoided a web content management system.

---

[20] SQLite: https://www.sqlite.org/ (retrieved: June 14, 2022).
[21] FAIR data review form:
https://docs.google.com/forms/d/e/1FAIpQLSd8_pd2r2SnjCVfCC3CHhEUHZzv2MTRC3RTh0S2YTvbVJj87Q/viewform (retrieved: June 14, 2022).
[22] Bootstrap: https://getbootstrap.com/ (retrieved: June 14, 2022).

The showcase consists of two main parts: (1) an opening section, intended to grab the attention of the reader (cf. the 'lead' of a news story) and showing an image that reflects the content as a kind of signboard; (2) the rest of the showcase containing more detailed information about the essentials of the research project and the dataset together with the data preview and Quick Start section.

As mentioned in the *'Introduction'* of this paper, every showcase was made by an editor, who based his text selection on the data paper published in RDJ. To simplify the authoring process and ensure consistency in coding, we created a 'wizard', a screen form into which texts can be entered into Markdown; it generates the HTML code for a *basic* version of the showcase. The *completion* is manual work: adding images, charts, maps, et cetera, and taking care of the layout of the final version as a whole. This finishing touch has also been made less labour-intensive by defining standard JavaScript code for each type of embedded multimedia (videos, animated statistical chart, interactive map, etc.), which requires a comprehensive set of de-facto standard software tools. An overview can be found on the page *Data Visualisation Demos*[23] of the Exhibit. Each of the images in this gallery corresponds to a fully working demo in the page style of the showcase. The corresponding program code can be easily copied and inserted into the showcase with new data making the addition of multimedia illustrations quite easy. It worked well in an experimental setting but should be further automated when scaling up.

Figure 3.   Form for creating a basic showcase developed for the Exhibit of Datasets (for editorial use only).

Discussion

Taking a few steps back to get a better overview of this experiment, we have several more questions that need to be addressed:

1. The **format** of the showcase: to what extent is this essential?

4. **Scalability**: should every data paper get a showcase? What does scaling mean in practice? To what extent is the showcase relevant for the collection of a data archive?

5. What should be the **role of the author** of the data paper (depositor of the dataset)? What is the optimal division of tasks between the author (depositor) and the editor of the showcases?

6. What about the **durability** of all the unconventional resources mentioned?

7. **Innovation**: How to respond to the described innovative trends? Are they bizarre examples, or are they harbingers of new standards?

Format

The format of a showcase can take various alternative forms. We opted for concise showcases as useful complements to data papers of the size used in RDJ. There are several data journals with shorter papers than RDJ, some of which are little more than announcements that a dataset is available; they could also be upgraded to showcases without much effort. The idea can be applied equally well to long data papers, where, for example, collapsed text may be used to create layers by temporarily hiding detailed digressions. In all cases, a well-arranged page layout is of great importance, providing the user an overview of the various directions in which to continue reading. If the objective is to generate interest, then the text should be visually appealing and hold attention through meaningful illustrations and good typography.

Scalability

Our experiment included 31 data papers, 18 of which were showcased. It left ample room for manual editing, which might not be possible with larger numbers. An inevitable question is how many showcases are necessary, desirable, and financially feasible. This depends on the strategic goal one has in mind: what aspect of the showcase does one want to build on?

- *The shop window aspect:* A data archive may use the showcase concept in the same way as museums or public archives use online exhibitions: show how interesting the collection is, draw attention to masterpieces, and thus persuade people to employ what is on offer. Digital repositories and data archives in the humanities and social sciences contain many datasets that appeal to a broader audience and that can be employed for educational purposes. The metadata catalogues of these archives are OK for searching, but hardly present datasets in an attractive way – on the contrary, they are rather dull and uninspiring catalog records. The showcase is also suitable for focusing on one or more representative datasets in a larger collection. Then, the production of a limited number of showcases per year will suffice.

- *The hub aspect:* Scaling up the number of showcases in their role of stepping stones to more detailed information requires an adjustment of the production process, with more automation and possibly a greater role for the authors. For example, they could fill in the text in an online form (something like our wizard) and upload accompanying multimedia material. However, this process still requires some degree of editorial

control. It should be noted that some (justifiably long) data papers do not lend themselves to a drastic showcase-style summary, but these are exceptions. They include projects with complex datasets and/or data operations that require detailed explanations. If one nevertheless wants a showcase for this category, it is best to ask the author to write a new short promotional text.

Role of the author

In our experimental setup, we did not ask the authors formal approval of the showcases. However, there was regular informal contact: authors knew that a showcase was being created and we often approached them for additional multimedia content. This worked fine in the RDJ context, but when scaling up, a more transparent solution would be desirable. Now it can be somewhat confusing that the main text of the showcase, which is almost always taken verbatim from the published accompanying data paper, is followed by a Quick Start section by the editor. A clearer division of roles and tasks between authors and editors would rectify this and make the process more cost effective as well.

Once the decision is made to have more multimodal publications, authors will play an important role in providing all the material in the appropriate technical format. Just as we now ask that the text of the data paper adheres to, for instance, a certain citation style such as APA, guidelines are needed for multimedia and interactive materials. Organizing this process properly may not be easy, but it is a necessary step to avoid time-consuming manual interventions during the publication production process.

Durability

A conventional academic paper is a stable document, sustainably managed by the publisher. Once published, nothing changes to the content, which is also necessary because of the scientific reasoning and conclusions. However, one may doubt whether this golden rule is equally applicable to data papers. The data production itself is often an ongoing process, with regular new versions, which may require a slight updating of the data paper. As mentioned in the section *'Executable papers'* version control is a standard concept with executable papers.

In the case of a hyper document with multimodal content that is distributed over several web servers and partially stored in the cloud, we need to be aware of technical hitches and less certainty of unlimited availability. The long-term survival of websites is less guaranteed than the permanency of journals. The website with the online database of a researcher may be poorly maintained or even disappear with a job change or when project funding ends. From this perspective, we have to accept that supplementary materials have a more dynamic and volatile character – as is the case with the internet as a whole – in contrast with the permanency of the basic information such as a data paper, a full research paper, and a deposited dataset. The showcases also belong to this more volatile category, but would become obsolete once data papers are transformed into full-fledged hyper documents.

The challenge of innovation

Much of what we have discussed above is less new than it seems. In the foregoing, we have dealt with data publishing from the point of view of modern communications and current technology. However, some issues have been addressed from a different perspective much earlier, for example in Elseviers' 'Article of the Future' project (Aalbersberg et al., 2012).

Bringing data journal and data repository closer together

One may wonder whether there is still something that needs to be improved in terms of data access if the publication refers to the dataset via a persistent identifier such as a DOI. However, the DOI usually does not point directly to the data but to the description on a landing page.

That is not a problem if the set consists of a small number of files. However, it becomes different when there are dozens or even hundreds of files involved, as in archaeological excavations. From a coherence perspective, it would be helpful if there were more granularity in the references, which the DOI suffix system also allows for.

There are a few more comprehensive examples of bringing publication, data, and supplementary files closer together:

Dryad

Dryad is a community-driven, curated general-purpose data repository.[24] Since 2019 it has had a partnership with Zenodo[25], which is based at CERN and offers the possibility of software archiving. The collaboration between the two allows for exposing the relationship between software, data, and other citations, so users can find all related work (Ioannidis, 2020). Dryad has a discovery service that looks like a data journal with short papers, which is realized by basic checks, rigid curation standards, and detailed guidelines for metadata (including methods, data usage, and related work). In many cases, the papers are more than the average descriptions in data repositories. A typical entry is a summary with on the right a column containing a link to the dataset, a full research paper, and associated software (if available), and metrics about views and downloads. Dryad accepts data in any file format from any field of research, but in practice has its focus on life sciences. The emphasis is clearly on data sharing; downloading is direct and straightforward without extra mouse clicks, which contributes to coherence. However, Dryad has no data preview option.

Dataverse and Open Journal Systems (OJS)

The Dataverse[26] plugin for journals based on OJS[27] makes it easy to upload the dataset at the same time as submitting a paper (Castro & Garnett, 2014). The widely used Dataverse data repository system is also interesting in this context because of its so-called *integrations*[28], which include tools for data exploration, visualization, and analysis, the aforementioned Binder for Jupyter notebooks, and software for handling geospatial data.

ARPHA

The ARPHA[29] acronym stands for Authoring, Reviewing, Publishing, Hosting, and Archiving, and is the name of Pensoft's innovative publishing platform, particularly tuned to biodiversity research. It supports the full life cycle of a manuscript within a single online collaborative environment and comprises various facilities, all offered as Software as a Service (SaaS). Pensoft introduced a novel article format that includes data and software. This concept of "integrated data publishing" is best demonstrated in the *Biodiversity Data Journal*.[30] It allows for direct import and export of data into and from the text in a structured form (Penev et al., 2017, particularly pp. 10-12). The ARPHA workflow is integrated with Dryad for deposition, publication, and permanent linking between data and data papers and interacts with special data platforms such as the Global Biodiversity Information Facility (GBIF)[31].

Managing supplementary files

The standard academic publication consisting of text, images, and tables, does not imply that authors do not want to add more varied material to their papers. On the contrary, over the

---

[24] Dryad: https://datadryad.org/ (retrieved: June 14, 2022).

[25] Zenodo: https://zenodo.org/ (retrieved: June 14, 2022).

[26] Dataverse: https://dataverse.org/ (retrieved: June 14, 2022).

[27] Open Journal Systems (part of Public Knowledge Project): https://pkp.sfu.ca/ojs/ (retrieved: June 14, 2022).

[28] Dataverse integrations: https://guides.dataverse.org/en/latest/admin/integrations.html (retrieved: June 14, 2022).

[29] ARPHA platform: https://arphahub.com/about/platform (retrieved: June 14, 2022).

[30] Biodiversity Data Journal: http://bdj.pensoft.net/ (retrieved: June 14, 2022).

[31] GBIF: https://www.gbif.org/ (retrieved: June 14, 2022).

past decade, journal editors have repeatedly complained about the growing number of supplementary files of all kinds. Editors generally agree that supplementary files are necessary because articles count limited numbers of pages. However, there is a risk of unmanageable proliferation (Greenbaum et al., 2017; Kwon, 2019; Marcus, 2009; NISO-NFAIS, 2013; Pop & Salzburg, 2015). To summarize the main challenges mentioned in these publications:

1. The supplementary appendix as a long list of items becomes an unstructured grab-bag.

8. Authors who work with new media, tend to mix essential and non-essential information.

9. There are excesses, for example, Science published a 6-page article plus 165 pages of supplementary material and another 5-page article with 144 pages of supplementary material plus a spreadsheet with six additional tables (Pop & Salzberg, 2015, p. 2).

10. Literature references within supplementary material do not get tracked by citation indices.

11. Supplementary materials are rarely thoroughly reviewed.

12. Supplemental tables are commonly in PDF format, which hampers users to extract data for reuse.

13. The long-term preservation, curation and maintenance costs, and accessibility and presentation of supplemental materials are not always taken care of.

The combination of all of these presents a huge trust issue. Recent research shows that many researchers appear not to share their supplementary data at all, even if they stated they would according to a journal's data availability policy (see Gabelica, Bojčić & Puljak, 2022). Several solutions for the management of supplementary data have been proposed: some journals (e.g., the *Journal of Neuroscience*, see Maunsell, 2010) no longer allowed authors to include supplemental material (which seems a step backwards); setting limits on the size of supplementary content is considered as arbitrary – a better alternative is a restriction to specific categories and to ensure that each supplemental item is associated with a figure or table in the main paper (Marcus, 2009; NISO-NFAIS, 2013; Pop & Salzberg, 2015). Greenbaum et al. (2017) proposed a hierarchical organization for supplements with some parts paralleling and "shadowing" the main text and other elements branching off from it, together with specific formatting to make this structure explicit. The primary text would figuratively sit on the top of the supplementary hierarchy. Local links point to more detailed descriptions of methods and data located within the supplemental materials.

Facing the innovator's dilemma

'They speak of electronic town meetings and virtual communities. Commerce and business will shift from offices and malls to networks and modems. And the freedom of digital networks will make government more democratic.
Baloney. Do our computer pundits lack all common sense? The truth in [sic!] no online database will replace your daily newspaper, no CD-ROM can take the place of a competent teacher and no computer network will change the way government works.'

Clifford Stoll, 1995.

The above-mentioned proposals on structuring supplementary material come close to the concept of a hyper document as a layered network of content nodes with clearly marked hyperlink paths. So, this idea itself is not so new, but the perspective of multimodality certainly

is. A weakness of the solutions mentioned above is that they continue to hold on to the traditional publication concept that goes back to the printed article with an appendix.

But if editors themselves also realize that innovation is necessary due to unstoppable technological developments, why has so little changed, and Pensoft, for example, is still a rarity? Some cynical answers can be found on the internet, such as that of Wikert (2015), who observed a lot of conservatism and anxiety with publishers being afraid of doing anything that might be perceived as a threat to the key retailers. They rely too much on the industry leader, Amazon, to also serve as an innovation leader. Although there are many creative people in-house, this industry cannot escape the so-called "innovator's dilemma", which states that any successful business will fail to innovate because it needs to sustain its current business model and serve existing customers (Evans, 2017).

However, it does not seem fair to point to publishers alone; data archives and authors also have a responsibility when it comes to innovation. Innovation is rarely welcome and is often misunderstood; this is why the default state of every new idea is non-adoption. Henry Ford was credited with the quote: "If I had asked the public what they wanted, they would have said a faster horse." If we had asked our authors if they wanted multimodal data papers, they might have replied: "rather put your energy into shorter time-to-publication and into more open access!"

It has always been difficult to properly assess the disruptive power of new developments. It is worth reading Stoll's (1995) anathema against the World Wide Web in Newsweek twenty-five years ago, from which the quote at the beginning of this section is taken – read it, not to joke about it, but to see why it was acceptable reasoning at the time. Rejection is generally based on a common sense view of the market and poorly understood *temporary* limitations of a new product. There are numerous examples. When Stephen Wozniak and Steve Jobs had completed their Apple I, Wozniak tried to interest Hewlett-Packard, his employer, in making personal computers, but HP doubted there was much of a market for such a machine, and refused (Augarten, 1984, pp. 277-279).

If we want to persuade people to embrace a new concept, we can learn, for example, from the edible insect industry. Most people will recoil if you offer them a handful of crickets as a snack, but there certainly appears to be a market for cricket flour processed in chocolate cookies, waffles, and pancakes (Nurmohamed, 2016). By analogy, we can also downscale a new publishing concept to lighter innovations in existing processes, but then it must be based on the joint effort of all parties involved, with each taking its share: publishers, data archives and authors. The question remains who will take the lead and which subsidy will initiate the process.

A Conceptual Model

The conceptual model below is a rough visual summary of where the previous discussion has taken us. It is "conceptual" because it makes no assumptions about actual implementation and organization, implying that it will not fit exactly every situation that arises in practice. It does not show the data paper or showcase as a hyper document in the centre of a hypertext network but focuses on the roles of publisher, data archive, and authors. We also omitted linked full research papers and project websites for the sake of clarity.

In this setup, the supplementary files are distributed over locations that are most appropriate for display. They are supposed to be connected with hyperlinks to the corresponding passages in the data paper as suggested in proposals for the management of media above. This does not mean that this material should *only* be stored there. Videos may be placed on YouTube (if you allow Google to monetize the content) but should also be deposited in a certified repository to guarantee permanency.
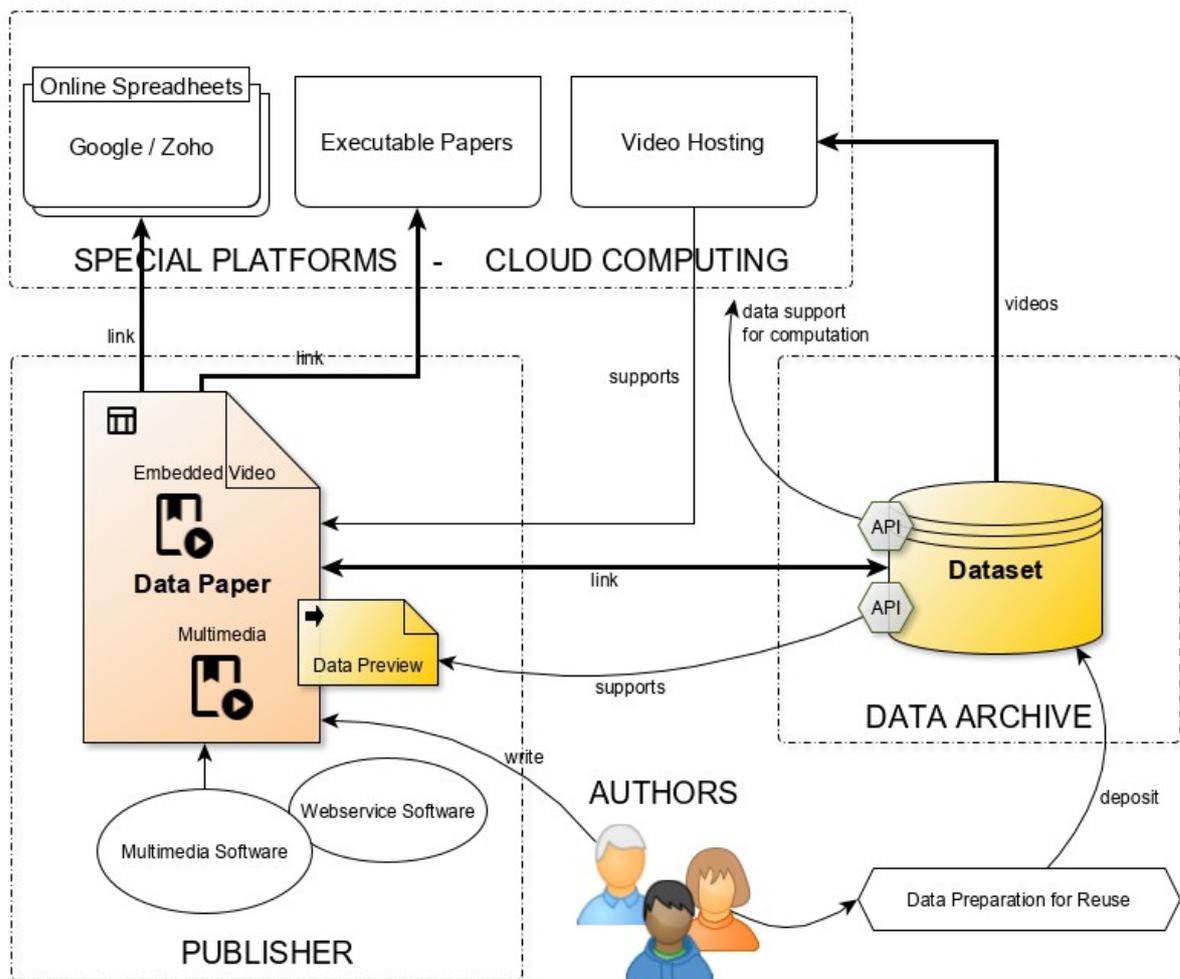
Figure 4.   Conceptual model for data publishing.

Some parts of the model deserve further explanation:

Publisher

Embedding more than just static images on an HTML page is not rocket science. Web service software as Bootstrap comes already with ready-made constructions for multimedia. Interactive maps, charts, and timelines that depend on a small dataset required in our experiment a few lines of JavaScript code, but this can be simplified by automation via scripting and asking authors to provide the data in an extractable format such as a spreadsheet. A good example is Knightlab's TimelineJS,[32] which reads its data from an online Google spreadsheet and requires minimal coding in the web page itself. If embedding is not possible, at least a screenshot should be included as a stub linked to the video or other medium elsewhere on the web.

Data preview

It is difficult to make general statements about the data preview, because this strongly depends on the organizational form (e.g., Dryad-Zenodo, ARPHA) and the nature of the datasets. As mentioned in the section *'Exploring the data set'* data preview serves several purposes: it supports the data description, provides trust by enabling exploration, arouses interest among

---

[32] TimelineJS: https://timeline.knightlab.com/ (retrieved: June 14, 2022).

non-experts, and is also a helping hand for the reviewers. These criteria are by no means always applicable. A separate data preview may be superfluous if the dataset has been very carefully prepared and is in a clear and comprehensible format. On the other hand, some datasets are too complicated for showcasing and/or are only intended for use among fellow experts. A report of an archaeological find that has yielded hundreds of potsherds does not have to count on widespread public interest, but a Late Iron Age farm or an excavated Roman ship certainly deserves a more popular presentation. In any case, an editor or data curator/archivist must take a critical look at the dataset, and this is best done by trying to browse through (a part of) the data. An appropriate API (if available) will help the journal embed or ingest more data.

Authors

Authors need to realize that a dataset they have worked with easily for many years is not always immediately comprehensible to outsiders. They should therefore think about data previews, provide examples for this, together with instructions for a quick start to explore the data. Of course, the journal and/or data archive should provide guidelines for this – it will not happen spontaneously, as it requires additional work for data depositors.

Cloud computing

Software as a service (SaaS) will become the inevitable mainstay of data publishing as a hyper document. We mentioned already read-only online spreadsheets that allow the reader to experiment. The Binder platform allows researchers to quickly create the computational environment needed to interact with research code and data shared online. To interact with someone else's work, users simply click a URL, which takes them directly to a live environment where they can run the code in the cloud (Jupyter Project, 2017; Google's Colaboratory).

Videos can be hosted on a regular website, but streaming video requires a dedicated hosting service that takes into account the characteristics of the target device and the quality of the connection (e.g., YouTube, Vimeo, and many specialized companies). Various smaller community-driven websites can also play a critical role in the hyper document network. For example, *The Amsterdam Time Machine* (ATM)[33] is an RDF-based public research resource on the history of Amsterdam with a systematic linkage of datasets from various projects and can provide academic context to a data paper.

Conclusions

The Exhibit of Datasets experiment successfully demonstrated how an online data publication could be transformed into a hyper document: a network with a showcase as the central hub, with embedded multimedia and hyperlink paths to in-depth content, clearly marked by local context cues in the form of buttons, screenshots, and clickable maps. This could be achieved with relatively simple means such as open-source software, cloud services, and some JavaScript code. The chosen format also made it possible to accommodate occasional requests to publish more additional documents than the data paper format itself allowed, without any hassle.

Scaling up does not necessarily imply following the complex, though sophisticated Pensoft model; changes can also be made incrementally, but cooperation between the various stakeholders is necessary. The publisher needs to make its production system more flexible and offer more features; the data archive can be helpful by facilitating data exploration more; authors need to realize that easy access to their data does not equal easy reuse; and finally, editors need to be willing to actively cooperate in this reuse direction and draw attention to these new facilities and encourage their use. RDJ was started to make datasets more accessible to users, especially non-experts. Not every dataset is suitable to be brought into the limelight and

---

[33] Amsterdam Time Machine – projects: https://amsterdamtimemachine.nl/category/projects/ (retrieved: June 14, 2022).

can attract some broader interest, but those that are suitable, should be treated as consumables and be provided with quick-start tips.

The main question remains whether anything should change if we can assume that most publications will still follow the conventional format for a long time to come. Moreover, in our domain, the problem of supplementary files seems to be limited to occasional cases, at least for the time being. In the short term, the main driver will not be market demand but should be the pursuit of more quality in line with current trends in communication. Much depends on how one sees one's mission: does a journal want to be a market leader that keeps pace with innovation (e.g., the *American Anthropologist* – see section *'Multimodality''*), or is it satisfied with an established, authoritative position? Does a data archive want to resemble a busy staple market with a drop-off and pick-up of datasets, or does the status of a certified "last resting place for data" suffice? Lack of innovation will drive authors demanding more advanced functions away, to other journals and platforms, such as those accommodating executable papers.

It is the innovator's dilemma: managers may be aware of disruptive technologies, but fail to value innovations properly because they are trying to apply them to their existing customers and product architectures. Moreover, these technologies are often too new and weak for the more advanced value networks for which the managers are responsible (Thrasyvoulou, 2014). IBM protected their mainframe business and never fully embraced the PC division, which was eventually sold to Lenovo. Microsoft was late to the internet, focused too much on operating systems, and was also late to cloud computing. That lateness has had irreparable consequences for their market positions. One should not wait until one sees the handwriting on the wall, because then the renewal has already taken place without gaining a significant position in the new constellation.

References

Aalbersberg, IJ. J., Heeman, F., Koers, H., & Zudilova-Seinstra, E. (2012). Elsevier's Article of the Future: Enhancing the user experience and integrating data through applications. *Insights*, *25*(1), 33–43. http://dx.doi.org/10.1629/2048-7754.25.1.33U (retrieved: June 14, 2022).

Astacio, P. A., Dattatreyan, G., & Shankar, A. (2020, November 8). Multimodal anthropologies section competition. *American Anthropologist*. http://www.americananthropologist.org/2020/11/08/multimodal-anthropologies-section-competition/ (retrieved: June 14, 2022).

Augarten, S. (1984). *Bit by bit. An illustrated history of computers*. Unwin Paperbacks.

Baciocchi, S., Beauguitte, L., Blavier, P., & Lambert, N. (2019). Documenting the diffusion of the 2016 French 'Nuit Debout'. *Research Data Journal for the Humanities and Social Sciences*. https://doi.org/10.1163/24523666-00401005 [Showcase in: *Exhibits of Datasets*, https://dansdatajournal.nl/rdp/exhibit.html?showcase=baciocchi2019] (retrieved: June 14, 2022).

Bezemer, J., & Kress, G. (2008). Writing in multimodal texts: A social semiotic account of designs for learning. *Written Communication, 25*(2), 166-195. https://doi.org/10.1177%2F0741088307313177 (retrieved: June 14, 2022).

Blom, H., Segers, E., Knoors, H., Hermans, D., & Verhoeven, L. (2018). Comprehension and navigation of networked hypertexts. *Journal of Computer Assisted Learning*, 1–9. https://doi.org/10.1111/jcal.12243 (retrieved: June 14, 2022).

Borgman, Ch. L., Scharnhorst, A., & Golshan, M. S. (2019). Digital data archives as knowledge infrastructures: Mediating data sharing and reuse. *Journal of the Association for Information Science and Technology (ASIS&T)*, *70*(8), 888-904. https://doi.org/10.1002/asi.24172 (retrieved: June 14, 2022).

Brandt, S. R. (n.d.). *Where can I publish executable papers and notebooks?* [Blog]. Software Sustainability Institute. https://www.software.ac.uk/resources/guides/where-can-i-publish-executable-papers (retrieved: June 14, 2022).

Castro, E., & Garnett, A. (2014). Building a bridge between journal articles and researchdata: The PKP-Dataverse Integration Project. *International Journal of Digital Curation, 9*(1), 176-184. http://dx.doi.org/10.2218/ijdc.v9i1.311 (retrieved: June 14, 2022).

Collins,S. G., Durington, M., & Gill, H. (2017). Multimodality: An invitation. *American Anthropologist 119*(1), 142-146. https://doi.org/10.1111/aman.12826 (retrieved: June 14, 2022).

Corallo, A., Esposito, M., Lazoi, M., Marra, M., & Sammarco, M. (2017). Innovating cultural heritage promotion through virtual and interactive technologies. *Proceedings of IMEKO, International Conference on Metrology for Archaeology and Cultural Heritage Lecce, Italy, October 23-25, 2017*, 504-509.

Corti-Georgiou, C. (2019). Pioneers of social research: A life story interview collection. *Research Data Journal for the Humanities and Social Sciences*. https://doi.org/10.1163/24523666-00401008 [Showcase in: *Exhibits of Datasets*, https://dansdatajournal.nl/rdp/exhibit.html?showcase=georgiou2019] (retrieved: June 14, 2022).

de Jager, E. J. (2016). 'Reis van de Razzia' (Victims of the German raid in Rotterdam 1944). *Research Data Journal for the Humanities and Social Sciences*. https://doi.org/10.1163/24523666-01000001 [Showcase in: *Exhibits of Datasets*, https://dansdatajournal.nl/rdp/exhibit.html?showcase=jager2016] (retrieved: June 14, 2022).

Doorn, P. K. (2020). Archiving and managing research data: Data services to the domains of the humanities and social sciences and beyond: DANS in the Netherlands. *Der Archivar*, 73(1), 44-50. Archiving and Managing Research Data: data services to the domains of the humanities and social sciences and beyond: DANS in the Netherlands (retrieved: June 14, 2022).

Esposito, E., & Zollo, S.A. (2021). "How dare you call her a pig, I know several pigs who would be upset if they knew": A multimodal critical discursive approach to online misogyny against UK MPs on YouTube. *Journal of Language Aggression and Conflict*, *9*(1), 44-75. https://doi.org/10.1075/jlac.00053.esp (retrieved: June 15, 2022).

European Commission, Directorate-General for Research and Innovation (2022), *European Research Data Landscape : final report*. Publications Office of the European Union. https://data.europa.eu/doi/10.2777/3648

Evans, B. (2017). The ideas paradox: why publishers fail to innovate. *LinkedIn.* https://www.linkedin.com/pulse/ideas-paradox-why-publishers-fail-innovate-rebecca-evans (retrieved: June 15, 2022).

Gill, H. (2019, March 28). Multimodality and the future of anthropological research and scholarship. *Arts and International Affairs.* https://theartsjournal.net/2019/03/28/multimodality-anthropology/ (retrieved: June 15, 2022).

Gabelica, M., Bojčić, R., & Puljak, L. (2022). Many researchers were not compliant with their published data sharing statement: mixed-methods study. *Journal of Clinical Epidemiology.* https://doi.org/10.1016/j.jclinepi.2022.05.019 (retrieved: June 15, 2022).

Graham, S. (2020). An approach to the ethics of archaeogaming. *Internet Archaeology, 55.* https://doi.org/10.11141/ia.55.2 (retrieved: June 15, 2022).

Greenbaum, D., Rozowsky, J., Stodden, V., & Gerstein, M. (2017) Structuring supplemental materials in support of reproducibility. *Genome Biology, 18*(64). https://doi.org/10.1186/s13059-017-1205-3 (retrieved: June 15, 2022).

Gregory, K. M., Cousijn, H., Groth, P., Scharnhorst, A., & Watt, S. (2020). Understanding data search as a socio-technical practice. *Journal of Information Science, 46*(4), 459-475. https://doi.org/10.1177%2F0165551519837182 (retrieved: June 15, 2022).

Herrada-Valverde, G., & Herrada-Valverde, R. I. (2017). Factors influencing hypertext reading comprehension. *Ocnos, 16*(2), 7-16. https://dx.doi.org/10.18239/ocnos_2017.16.2.1287 (retrieved: June 15, 2022).

Huvila, I. (2014). *Perspectives to archaeological information in the digital society.* Uppsala Universitet.

Interactive digs (2019). Archaeological Institute of America and Archaeology Magazine. https://www.interactivedigs.com/ (retrieved: June 15, 2022).

Ioannidis, A. (2020, March 10). Dryad & Zenodo: Our path ahead. *Zenodo Blog.* https://blog.zenodo.org/2020/03/10/dryad-and-zenodo-our-path-ahead/ (retrieved: June 15, 2022).

Jewitt, C. (2013). Multimodal methods for researching digital technologies. In S. Price, C. Jewitt, & B. Brow (Eds.), *The SAGE handbook of digital technology research* (Ch. 17). Sage.

Jupyter Project (2017, November 30). Introducing Binder 2.0 – share your interactive research environment. *eLife.* https://elifesciences.org/labs/8653a61d/introducing-binder-2-0-share-your-interactive-research-environment (retrieved: June 15, 2022).

Korhonen, V. (2010). *Dialogic literacy: A sociocultural literacy learning approach.* In A. Lloyd, & S. Talja (Eds.), *Practising information literacy. Bringing theories of learning, practice and information literacy together* (pp. 211-226). Elsevier.

Kress, G., & Van Leeuwen, T. (2001). *Multimodal discourse.* Edward Arnold.

Kwon, D. (2019, April 19). The push to replace journal supplements with repositories. *The Scientist*. https://www.the-scientist.com/news-opinion/the-push-to-replace-journal-supplements-with-repositories--66296 (retrieved: June 15, 2022).

Lamb, J. (2019a). *Space, sociomateriality, sound. The learning spaces of higher education*. The University of Edinburgh. https://hdl.handle.net/1842/36556

Lamb, J. (2019b, December 19). *What a digital thesis can look and sound like*. https://www.james858499.net/blog (retrieved: June 15, 2022).

Lasser, J. (2020, August 19). Creating an executable paper is a journey through Open Science. *Nature. Communications Physics*, *3I*. Article 143. https://www.nature.com/articles/s42005-020-00403-4 (retrieved: June 15, 2022).

Loy, M. P. A, Stocker, S. R., & Davis, J. L. (2021). From archive to GIS: Recovering spatial information for Tholos IV at the palace of Nestor from the notebooks of Lord William Taylour. *Internet Archaeology*, *56*. https://doi.org/10.11141/ia.56.5 (retrieved: June 15, 2022).

Marcus, E. (2009, October 2). Taming supplemental material. [Editorial]. *Cell 139*. https://doi.org/10.1016/j.cell.2009.09.021 (retrieved: June 15, 2022).

Maunsell, J. (2010, August 11). Announcement regarding supplemental material. [Editorial]. *Journal of Neuroscience*, *30*(32), 10599-10600. https://www.jneurosci.org/content/30/32/10599 (retrieved: June 15, 2022).

Mol, A. A. A., Ariese-Vandemeulebroucke, C. E., Boom, K. H. J., & Politopoulos, A. (Eds.). (2017). *The interactive past*. Sidestone Press. https://www.sidestone.com/books/the-interactive-past (retrieved: June 15, 2022).

NISO-NFAIS (2013, January). *Recommended practices for online supplemental journal article materials*. http://www.niso.org/standards-committees/supplemental (retrieved: June 15, 2022).

Nurmohamed, S. (2016, April 1). Selling a crazy idea: How radically new products can gain traction. *Knowledge @ Wharton*. https://knowledge.wharton.upenn.edu/article/samir-nurmohamed-research-int/ (retrieved: June 15, 2022).

O'Shea, K. (2019, May 16). *Estimating and predicting the number of Jupyter Notebooks on Github* [Blog]. https://kyso.io/KyleOS/nbestimate (retrieved: November 20, 2021).

Pasquetto, I. V., Borgman, C. L., & Wofford, M. F. (2019). Uses and reuses of scientific data: The data creators' advantage. *Harvard Data Science Review, 1*(2). https://hdsr.mitpress.mit.edu/pub/jduhd7og/release/7 (retrieved: June 15, 2022).

Penev, L., Georgiev, T., Geshev, P., Demirov, S., Senderov, V., Kuzmova, I., Kostadinova, I., Peneva, S., & Stoev, P. (2017). ARPHA-BioDiv: A toolbox for scholarly publication and dissemination of biodiversity data based on the ARPHA publishing platform. *Research Ideas and Outcomes*, *3*. Article e13088. https://doi.org/10.3897/rio.3.e13088 (retrieved: June 15, 2022).

Pimentel, J. F., Murta, L., Braganholo, V., & Freire, J. (2019). A large-scale study about quality and reproducibility of Jupyter Notebooks. *IEEE/ACM 16th International Conference on Mining Software Repositories (MSR), 2019* (pp. 507-517). http://dx.doi.org/10.1109/MSR.2019.00077 (retrieved: June 15, 2022).

Pop, M., & Salzberg, S.L. (2015). Use and mis-use of supplementary material in science publications. *BMC Bioinformatics 16*(237) . https://doi.org/10.1186/s12859-015-0668-z (retrieved: June 15, 2022).

Reinhard, A. (2018). *Archaeogaming. An introduction to archaeology in and of video games*. Berghahn.

Richards, J., & Roskams, S. (2013). Burdale: An Anglian settlement in the Yorkshire Wolds [Data Paper]. *Internet Archaeology*, *35*. https://doi.org/10.11141/ia.35.8 (retrieved: June 15, 2022).

Rieger, O.Y. (2019, May 8). *The usability of research data: If we curate, will they reuse it?* [Blog]. ITHAKA S+R. https://sr.ithaka.org/blog/the-usability-of-research-data-if-we-curate-will-they-reuse-it/ (retrieved: June 15, 2022).

Roorda, R. (2018). Coding the Hebrew Bible. *Research Data Journal for the Humanities and Social Sciences*. https://doi.org/10.1163/24523666-01000011 [Showcase in: *Exhibits of Datasets*, https://dansdatajournal.nl/rdp/exhibit.html?showcase=roorda2018] (retrieved: June 15, 2022).

Stoll, C. (1995). Why the Web won't be Nirvana. *Newsweek*, February 26. https://www.newsweek.com/clifford-stoll-why-web-wont-be-nirvana-185306 (retrieved: June 15, 2022).

Storrer, A. (2002). Coherence and text and hypertext. *Document Design*, *3*(2), 156-168.

Sun, G., & Khoo, C. S. G. (2016). Social science research data curation: issues of reuse. *Libellarium, 9*(2), 59-80. http://dx.doi.org/10.15291/libellarium.v9i2.291 (retrieved: June 15, 2022).

Takaragawa, S., Smith, T. L., Hennessy, K., Astacio, P. A., Chio, J., Nye, C., & Shankar, S. (2019). Bad habitus: Anthropology in the age of the multimodal. *American Anthropologist*, *121*(2), 517-524. https://doi.org/10.1111/aman.13265 (retrieved: June 15, 2022).

Thrasyvoulou, X. (2014). Understanding the innovator's dilemma. *Wired*. https://www.wired.com/insights/2014/12/understanding-the-innovators-dilemma/ (retrieved: June 15, 2022).

Tran, A. (2019). *Multimodal dissertations: Opportunities for multimodality in higher education*. https://ir.lib.uwo.ca/etd/6228 (retrieved: June 15, 2022).

Unger, J., Hemker, C., Lobinger, C., & Jan, M. (2020). VirtualArch: Making archaeological heritage visible. *Internet Archaeology*, *54*. https://doi.org/10.11141/ia.54.2 (retrieved: June 15, 2022).

van Nieuwkerk, M. M., Salters, L., Helmers, R. M., & Kisjes, I. (2020). Operatic productions in the Netherlands, 1886-1995: From printed annals to searchable performance data. *Research Data Journal for the Humanities and Social Sciences*. https://doi.org/10.1163/24523666-00502007 [Showcase in: *Exhibits of Datasets*, https://dansdatajournal.nl/rdp/exhibit.html?showcase=nieuwkerk2020b] (retrieved: June 15, 2022).

Visconti, A. (2015). *"How can you love a work if you don't know it?": Critical code and design towards participatory digital editions.* http://dr.amandavisconti.com/ (retrieved: June 15, 2022).

Wakefield, C. (2020). Digital public archaeology at Must Farm: A critical assessment of social media use for archaeological engagement. *Internet Archaeology*, *55*. https://doi.org/10.11141/ia.55.9 (retrieved: June 15, 2022).

Wikert, J. (2015). Whatever happened to innovation in the publishing industry? *DisruptorFest*. https://jwikert.typepad.com/the_average_joe/2015/12/whatever-happened-to-innovation-in-the-publishing-industry.html (retrieved: June 15, 2022).

Xie, Y., Allaire, J. J., & Grolemund, G. (2021). *R Markdown: The definitive guide.* Ch. 3.2: Notebooks. https://bookdown.org/yihui/rmarkdown/notebook.html (retrieved: June 15, 2022).

Yoon, A. (2016). Red flags in data: Learning from failed data reuse experiences. *Proceedings of the Association for Information Science and Technology, 53*(1), 1-6. https://asistdl.onlinelibrary.wiley.com/doi/full/10.1002/pra2.2016.14505301126 (retrieved: June 15, 2022).

Yoon, A. (2017). Role of communication in data reuse. *Proceedings of the Association for Information Science and Technology, 54*(1), 463-471. https://doi.org/10.1002/pra2.2017.14505401050 (retrieved: June 15, 2022).