

FAIR Paper: Applying FAIR to Academic Publishing

Wouter Beek^{1,*†}, Rick Mourits^{2,†} and Auke Rijpma^{3,†}

¹*Triply B.V.*

²*International Institute for Social History*

³*Universiteit Utrecht*

Abstract

The FAIR principles have a significant and lasting impact on the way in which research is performed in the Digital Humanities. However, the FAIR principles have not yet significantly impacted the way in which research papers are published and disseminated. This paper describes a new approach towards academic publishing called ‘FAIR Paper’. A FAIR Paper is an academic publication that lives on the Web, uses open standards, and is completely reproducible. We report on our findings based on an early Proof-of-Concept implementation of the FAIR Paper concept.

Keywords

Linked Data, FAIR Principles, Academic Publishing

1. Introduction

The FAIR principles have a significant and lasting impact on the way in which research is performed in the Digital Humanities. However, the FAIR principles have not yet significantly impacted the way in which research papers are published and disseminated.

The Common Lab Research Infrastructure for the Arts and Humanities (Clariah) project¹ makes extensive use of linked data principles and techniques. Within the Clariah context, insights are often communicated between researchers in so-called Data Stories: online publications that contain interactive query visualizations. The ability to communicate research insights in a visual and meaningful way, is essential for a project like Clariah, where many users have a less technical background.

With the appearance of an increasing number of elaborate Data Stories, the need arose to give these stories a more academic and professional appeal. The rest of this article details the results of this exploration, resulting in a Proof-of-Concept (PoC) implementation of our FAIR Paper approach.

SemDH 2024: First International Workshop of Semantic Digital Humanities co-located with ESWC 2024, Hersonissos, Greece

*Corresponding author.

†These authors contributed equally.

✉ wouter@triply.cc (W. Beek); rick.mourits@iisg.nl (R. Mourits); auke.rijpma@gmail.com (A. Rijpma)

🆔 0000-0003-0250-9655 (W. Beek); 0000-0002-2267-1679 (R. Mourits); 20000-0002-8950-8227 (A. Rijpma)

© 2024 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

See <https://www.clariah.nl>

2. Related work

The RASH Framework for academic publishing was introduced in [1, 2]. It introduces a new publication approach that relies on open web standards like HTML and CSS instead of PDF and Word. In addition, RASH publications can include RDF for metadata annotations. The RASH Framework has been successful in changing the practice of academic publishing in the Semantic Web research domain, where the main conferences now accept submissions in that format. In [3], the RASH approach was extended for non-textual entities like complex formulas and figures.

Kuhn et al. [4] observe that the vast majority of work under the heading of ‘semantic publishing’ has focused on adding semantic layers on top of existing academic publication approaches. They argue that ‘genuine semantic publishing’ must be more radical, and requires changes to the actual process of constructing and disseminating academic publications. Our FAIR Paper approach implements some, but not all requirements that Kuhn et al. enumerate under the heading of ‘genuine semantic publishing’. Specifically, a FAIR Paper can still be viewed and regarded as a traditional, narrative-based publication.

There are several attempts at creating academic publications that can be reproduced by rerunning – and thereby reproducing – the work of the original authors. An early attempt at this approach is [5]. Other approaches make use of an online notebook system, such as Jupyter Notebooks², where Python or R code can be executed on a web server. However, none of these approaches makes exclusive use of web standards, and relies on complex execution environments like virtual machines or servers that perform the heavy lifting.

The FAIR Principles were introduced in [6]. They have not yet had a lasting impact on the dissemination of academic papers.

3. The FAIR Paper approach

FAIR Papers make use of the following technology stack:

- One or more RDF triple stores, where research data is published according to shared data models.
- One or more data models, expressed in SKOS, RDFS, OWL, and/or SHACL.
- One or more SPARQL endpoints, through which data can be retrieved.
- Multiple SPARQL queries, in which research data is retrieved, processed, integrated, aggregated, computed, and visualized.
- One HTML page, that contains the structural and textual/narrative content of the academic publication.
- One CSS style sheet that implements the academic style of the venue for which the academic publication is prepared. This must include a web style and a print style.
- A modern web browser with a generic print feature.

²<https://jupyter.org>

Notice that the technology stack for FAIR Paper only uses open web standards and readily available software components. It does not require an arbitrary execution environment, as with notebooks or virtual machines. All computation is performed through standardized web APIs like SPARQL. This also means that FAIR Papers cannot do everything that a notebook or virtual machine can do. For example, a statistical test that cannot be implemented in contemporary linked data standards like SPARQL/SHACL/OWL, cannot be utilized.

The following subsections explain details of our FAIR Paper PoC implementation.

3.1. Data, queries, and stories

For storing data, the Clariah linked data environment is used³. This linked data environment currently contains 139 public datasets.

Besides data, the Clariah triple store also allows SPARQL queries to be stored. The ability to store SPARQL queries is not (yet) part of the SPARQL standard. In an earlier Clariah project, a metadata language for disclosing SPARQL queries was developed under the name GRLC. This approach is published in [7] and was implemented in the Clariah linked data environment⁴. The Clariah environment currently contains 453 publicly accessible saved queries.

In addition to data and queries, the Clariah linked data environment also allows Data Stories to be written. Data Stories are online publications that contain interactive query visualizations. The Clariah environment currently contains 39 public Data Stories.

3.2. Styling

In the FAIR Paper PoC, it is possible to apply academic styling in the web browser. We were able to re-implement a CSS version of the Lecture Notes in Computer Science (LNCS) style. This style is used by many journals and conferences in the computer science domain, and a CSS version was previously devised for RASH Papers as well. We were not able to implement academic styling for a venue in the (digital) humanities yet, since academic styles are not typically made available in a CSS format. This makes the creation of a CSS implementation labor-intensive or otherwise complex.

When this CSS style is applied to a FAIR Paper in the web browser, it is visually indistinguishable from a traditional LNCS publication that is created in \LaTeX , see Figure 1. This exercise makes it likely that other academic styles, at least to a very large extent, can also be expressed by using open web standards.

3.3. Printing

In the FAIR Paper PoC, it is possible to print papers from popular web browsers. This is implemented with CSS, where print styling is specified that deviates from web styling. This is crucial, since web content is not typically split across multiple pages, which is a requirement for print.






³The environment can be found at <https://druid.datalegend.net> and makes use of the TriplyDB triple store.

⁴See <https://triplify.cc/blog/2023-06-grlc> for a blog post on this topic.



The 'Spanish' Flu Epidemic in The Netherlands 1918-19



Ruben Schalk¹  [0000-0003-1991-5289](#),
Rick Mourits^{1,2}  [0000-0002-2267-1679](#),
Ingrid van Dijk^{2,3}  [0000-0001-6549-9090](#),
Auke Rijpma¹  [0000-0002-8950-8227](#), and
Richard Zijdeman⁴  [0000-0003-3902-3720](#)

¹Utrecht University

²Radboud University

³Lund University

⁴International Institute of Social History

Abstract. This datastory uses Dutch death certificates from 1910-20 to map the temporal, spatial and social distribution of the 'Spanish' flu epidemic that hit The Netherlands in 1918-19. Some regions, such as Groningen were notably more affected than others.

Keywords: Epidemics · FAIR Paper

I. The dataset

Thanks to the indexation efforts of archives and the [LINKS project](#), large parts of the Dutch civil registry ('Burgerlijke Stand') are now becoming available for historians. The death certificates used here are retrieved from [openarch.nl](#) (available [here](#)). From the individual death certificates files per archive, one combined dataset was created. One challenge of working with these certificates is that the same certificate may have been indexed by more

Figure 1: An example of the first page of a FAIR Paper in the web browser, that uses the LNCS style.

The result is shown in Figure 2, where the print and print preview features of regular web browsers are used. Our PoC was tested in recent versions of Chrome, Edge, Firefox, and Safari. It is possible to print to a PDF file and/or to physical paper.

3.4. Paper structure and metadata

Academic papers contain several structural and metadata elements that are unique to publishing, and that cannot be specified in a regular Data Story. Support for the following structural and metadata elements was added in the FAIR Paper PoC:

- List of authors, ORCID IDs, and affiliations

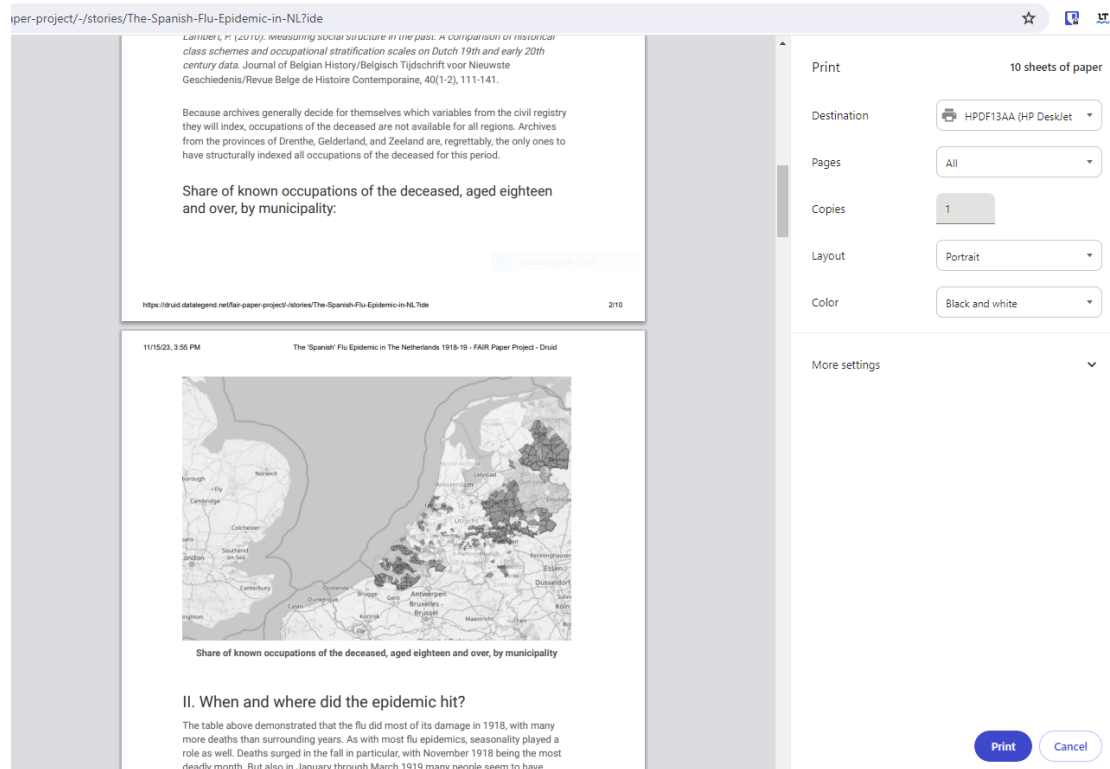


Figure 2: A FAIR Paper that is printed by using the standard print functionality in the web browser.

- Abstract
- Keywords
- Code blocks
- Figures, including live generated galleries, 2D/3D maps, timelines, chars, class diagrams, flow charts, or network visualizations, based on a SPARQL query.
- Tables, including live generated tables based on a SPARQL query.

The following structural and metadata elements were not added in the FAIR Paper PoC, due to a lack of time:

- References from running text to figures, tables, and code blocks.
- Bibliographic references.
- Footnotes and/or end notes.
- Bibliography styles such as APA or MLA.
- Captions for figures, tables, and code blocks.

In theory, some of the missing elements can be manually written with HTML tags, but in practice authors prefer an easier specification format for such elements.

4. User group test

During the FAIR Paper PoC, a user group of 5 domain experts in Clariah was formed. These experts have used several intermediate versions of the FAIR Paper PoC implementation. The domain experts are historians who work at the following research institutes: International Institute for Social History⁵, Cultural Heritage Agency⁶, University of Antwerp⁷, Leiden University⁸, and Utrecht University⁹. These domain experts all had prior experience with linked data, FAIR data, and/or data handling in general.

These technologically predisposed domain experts were asked to provide feedback and reflect critically on the FAIR Paper PoC. This resulted in many improvements that were incorporated during the PoC, in addition to several ideas for continued development. Furthermore, the domain expert were able to perform the following tasks with the final PoC version:

- Read an existing FAIR Paper.
- Modify an existing FAIR Paper.
- Create a new FAIR Paper from scratch.

This small user group test, with a limited and carefully selected group of users, indicates that FAIR Papers are already usable for domain experts with prior experience with Linked Data. Since the current implementation did not specifically optimize for user-friendliness, we hope that a future version of FAIR Paper will be accessible to a wide group of researchers in the humanities.

5. Conclusion and future work

Based on the results of this PoC, we conclude that it is possible to create dynamic and online academic publications that have the same professional features as static and offline publications. The technologies of the web (HTML, CSS, RDF, SPARQL) are strong enough to support such dynamic and online behavior. FAIR Papers have the benefit that they are published together with the data and queries. The tables and figures in a FAIR Paper can be recreated online, and/or can be adjusted by the reader to obtain different results over the same data. This gives readers of a FAIR Paper the opportunity to reproduce the research, and interact with the underlying data directly, so that publications can become truly FAIR.

Even though FAIR Papers can do many things that traditional papers cannot do, they are still consistent with the current practice of static and offline publishing. For example, they allow printing to PDF and/or physical paper.

Concrete examples of FAIR Papers can be found at <https://druid.datalegend.net/fair-paper-project>. This includes a tutorial that explains how users can create their own FAIR Papers.

⁵<https://iisg.amsterdam>

⁶<https://english.cultureelerfgoed.nl>

⁷<https://www.uantwerpen.be/en>

⁸<https://www.universiteitleiden.nl/en>

⁹<https://www.uu.nl/en>

Acknowledgments

We thank Clariah Work Packages 4 for funding the development of the FAIR Paper approach, and Work Package 5 for collaborating on the Data Stories and FAIR Paper concepts. We thank the domain experts that were part of the user test group. We thank the Triply software developers and data scientists who worked on this project.

References

- [1] A. Di Iorio, A. G. Nuzzolese, F. Osborne, S. Peroni, F. Poggi, M. Smith, F. Vitali, J. Zhao, The rash framework: enabling html+ rdf submissions in scholarly venues, ISWC 2015 Posters and Demonstrations Track, ISWC-P and D 2015 - co-located with the 14th International Semantic Web Conference, ISWC 2015 - 11 October 2015 (2015).
- [2] G. Spinaci, S. Peroni, A. Di Iorio, F. Poggi, F. Vitali, The rash javascript editor (raje): A wordprocessor for writing web-first scholarly articles, in: Proceedings of the 2017 ACM Symposium on Document Engineering, DocEng '17, Association for Computing Machinery, New York, NY, USA, 2017, p. 85–94. URL: <https://doi.org/10.1145/3103010.3103018>. doi:10.1145/3103010.3103018.
- [3] S. Mirri, S. Peroni, P. Salomoni, F. Vitali, V. Rubano, Towards accessible graphs in html-based scientific articles, in: 2017 14th IEEE Annual Consumer Communications & Networking Conference (CCNC), IEEE, 2017, pp. 1067–1072.
- [4] T. Kuhn, M. Dumontier, Genuine semantic publishing, *Data Science* 1 (2017) 139–154.
- [5] P. Van Gorp, S. Mazanek, Share: a web portal for creating and sharing executable research papers, *Procedia Computer Science* 4 (2011) 589–597.
- [6] M. D. Wilkinson, M. Dumontier, I. J. Aalbersberg, G. Appleton, M. Axton, A. Baak, N. Blomberg, J.-W. Boiten, L. B. da Silva Santos, P. E. Bourne, et al., The fair guiding principles for scientific data management and stewardship, *Scientific data* 3 (2016) 1–9.
- [7] A. Meroño-Peñuela, R. Hoekstra, grlc makes github taste like linked data apis, in: H. Sack, G. Rizzo, N. Steinmetz, D. Mladenović, S. Auer, C. Lange (Eds.), *The Semantic Web*, Springer International Publishing, Cham, 2016, pp. 342–353.