TRANSFORMATIONS
A DARIAH Journal

Workflows: Introduction (Part 1)

Anne Baillot[1, 2], Françoise Gouzi[2], Toma Tasovac[2, 3]

1 Le Mans Université
2 DARIAH-EU
3 Belgrade Center for Digital Humanities

In simple terms, a workflow is a structured sequence of tasks or processes designed to carry out a research activity. Workflows are key to our daily research practice: they shape how we collect, process, interpret and manage information relevant to our work (Goble et al. 2020). Yet despite their centrality, there is no one standard way of describing a workflow in the humanities, as opposed to other areas of research (Atkinson et al. 2017). Lacking a common language (Crusoe et al. 2022), workflows are currently difficult to share, replicate and reuse.

This question, however, is not wholly overlooked. As the inaugural volume of *Transformations: A DARIAH Journal* proves, there is an active community of researchers in the humanities and social sciences that strives to foster a culture of workflow reusability.[1] The published articles showcase how projects from a variety of scholarly horizons and areas of research share a common interest in carefully drafting contextualised workflow descriptions to facilitate their reuse (see also Marongiu, McGillivray, and Khan 2024). And they do not simply shed light on the methodologies developed; they reflect on how to present these methodologies in a replicable manner and on what their significance is for scholarship at large. In the context of open science, digital tools are critical to the implementation of reusability schemes. As such, they play a central role in the articles we present here.

Depending on their scope, workflows require technological tooling of varying complexity. Still, rather than organising this volume by the size of workflows

1. We are publishing this first batch of articles ahead of the DARIAH Annual Event 2025 to provide early access to these contributions and foster timely discussion and exchange. The second batch of articles, which will complete this first volume, is planned for publication in the autumn of 2025.

EPI sciences

To cite this article: Baillot, A., Gouzi, F. & Tasovac, T. (2025). Workflows: Introduction (Part 1). *Transformations. A DARIAH Journal*, vol. 1, p. 1-5, https://doi.org/10.46298/transformations.15780, CC BY 4.0.

or the tools employed, we have chosen to group the articles according to their primary objects of study: metadata-based workflows, textual data-based workflows, and image-based workflows. In all three areas, the articles grapple with the challenge of formalising research methodologies in ways that are not only transparent and rigorous, but also meaningfully reusable. The opening article by **Isto Huvila, "Paradata Conveys Understanding of Workflows and Facilitates the Reuse of Research Data in the Arts and Humanities,"** offers a methodological apparatus and a generic approach to workflow reusability based on the documentation of research data management. The paper reflects on the gaps between the data management information typically provided by research projects and the information users need to actually re-implement the corresponding methodology. It also establishes a core theme running through all the articles: the empowering role of the FAIR Principles[2] in elevating workflows as valuable research outputs.

The first section is dedicated to workflows related to metadata, an essential aspect in ensuring the findability and accessibility of research data in line with the FAIR Principles. Catalogues play a key role in this section. The article by **Gustavo Candela, Cezary Rosiński and Arkadiusz Margraf, "A Reproducible Framework to Publish and Reuse Collections as Data: The Case of the European Literary Bibliography,"** sheds light on bibliographic databases as an example of digital collections curated by galleries, libraries, archives and museums (GLAM institutions) and relevant to humanities scholarship. The workflow this article presents focuses on a specific literary bibliographic database, the European Literary Bibliography, and proposes concrete steps to make the corresponding metadata reusable in computational projects. The interaction between catalogues and (literary) research is also central in the article by **Iiro Tiihonen and Kira Hinderks, "Genre Classification Workflow for the English Short Title Catalogue (ESTC)."** This paper offers a solution to improve genre classification in the extraction of metadata from a catalogue. Based on a large English-language corpus, it presents both literary challenges and extraction techniques, while also addressing important questions of representation bias. The article by **Till Grallert, "Adding Every Arabic Periodical Published Before 1930 to Wikidata,"** expands the discussion to non-English corpora, highlighting issues of multilingualism and information transfer across languages. Here too, it is crucial that the workflow bridge the different types of metadata—along with their formats, curation and management practices—whether created by scholars or generated by cataloguers in the GLAM sector. This contribution underscores the advantages of relying on Wikidata to implement the FAIR Principles in this context. In **"A Workflow to Publish Collections as Data: Looking Back at Europeana.eu and Forward to the Common European Data Space for Cultural Heritage,"** by **Gustavo Candela, Sally Chambers, Alba Irollo,**

2. See https://www.go-fair.org/fair-principles/.

Nuno Freire, Vicky Dritsou, Antoine Isaac, Agiatis Benardou, Vicky Garnett and Toma Tasovac, the authors consider the role of research infrastructures in leveraging sustainability. They present existing resources and EU-wide initiatives to facilitate data transfer from GLAM institutions to research projects, demonstrating the continuity of heritage data as both cultural and scholarly assets. In this section, both Candela, Rosiński and Arkadiusz, and Candela et al. showcase Collections as Data as a powerful tool to facilitate workflow reusability at the interface between heritage and research.

The second section addresses approaches to text-based data. Digital scholarship relies on textual information, since text is the basis for the provision and curation of computer-readable information. It extends from metadata and annotation to actual textual content. **"The Sixth Generation of the Perseus Digital Library and a Workflow for Open Philology,"** by **Gregory Crane, James Tauber, Alison Babeu, Lisa Cerrato, Charles Pletcher, Clifford Wulfman, Sergiusz Kazmierski and Farnoosh Shamsian,** offer insights into one of the most important humanities projects of the past decade. The Perseus Digital Library has not only pioneered the constitution and dissemination of rich digital corpora for humanities research; as this article demonstrates, the project is also able to refine its workflows iteratively, adapt to change and ensure reusability for a wide array of users. It combines technical detail, especially around exporting annotations, with reflections on how a massive project can evolve to improve quality and accessibility over time. **Emiliano Degl'Innocenti, Francesco Pinna, Alessia Spadi and Federica Spinelli** also address the management of textual collections, albeit at a smaller scale, in **"Creating a Scientific Workflow to Manage Old Italian Texts."** Their article showcases the complementarity of publishing a workflow description on the SSH Open Marketplace and contextualising it in a journal publication, providing both technical detail and general reflections on reusability. Following approaches to corpora in Ancient Greek (Crane et al.) and Old Italian (Emiliano Degl'Innocenti et al.), the article by **Adeline Clarke, Krista Lagus and Maria Valaste, "finnsurveytext: Analysis of Open-Ended Survey Responses in R,"** moves the focus to the social sciences and to lesser-resourced contemporary languages. The workflow at the heart of this paper concerns itself with the analysis of open-ended survey answers—e.g. free text—in languages other than English. In addition to detailed guidance on using a dedicated R package, it offers insight into the improvement of text quality through semantic enrichment tools. In this sense, it converges with Crane et al. in the effort to contextualise highly technical semantic tools and make them accessible to humanities and social science practitioners less familiar with computational tools—which is precisely what reusable workflows aim to achieve.

The final section is dedicated to image-based workflows. In this context, annotation and semantic enrichment remain essential but are layered with

additional challenges related to the generation, curation and analysis of images. In **"Improving Workflows in Digital Art History: Sharing Annotations for Cultural Heritage Image Segmentation and Object Detection," Léa Maronet and Alice Truc** explore the challenges of using computational methods in art history. Their analysis provides valuable insights into how image-related work can be enhanced through standardisation—a crucial step to achieve interoperability and reusability, guided by the FAIR Principles. Accessibility and reproducibility are also key in **"A Reproducible Workflow for the Creation of Digital Twins in the Cultural Heritage Domain"** by **Sebastian Barzaghi, Alice Bordignon, Federica Collina, Francesca Fabbri, Bruno Fanini, Daniele Ferdani, Bianca Gualandi, Ivan Heibi, Nicola Mariniello, Arcangelo Massari, Marcello Massidda, Arianna Moretti, Silvio Peroni, Sofia Pescarin, Maria Felicia Rega, Giulia Renda and Mattia Sullini.** This article tackles 3D reconstruction at the intersection of research and dissemination. It offers a thorough overview of the many standards used in heritage-based 3D reconstructions and proposes a workflow to facilitate the processing of the corresponding metadata. It then describes the creation of a digital exhibition—a digital twin of a real-life exhibition—and reflects on both the technical and museological challenges involved. Beyond its technical contributions to digitisation processes in a heritage context, the paper argues in favour of openness, accessibility and reusability.

The contributions in this issue provide contextualised workflow descriptions in a series of areas highly relevant to research in the humanities and social sciences. They pave the way to what has the potential to become a major research output in its own right: scholarly articles focused on research methodologies—or what might be called *workflow papers*. Positioned at the crossroads of open science, state-of-the-art technologies and reflexive approaches to major fields of inquiry, workflow papers offer new perspectives for humanities research. They increase data power and digital literacy while helping to shape scholarly ethics in line with the FAIR Principles. By dedicating this issue of *Transformations: A DARIAH Journal* to workflows, we intend to root workflows as a recognised publication format in the scholarly landscape of the arts and humanities.

References

Atkinson, Malcolm, Sandra Gesing, Johan Montagnat, and Ian Taylor. 2017. "Scientific Workflows: Past, Present and Future." *Future Generation Computer Systems* 75 (October): 216–27. https://doi.org/10.1016/j.future.2017.05.041.

Crusoe, Michael R., Sanne Abeln, Alexandru Iosup, Peter Amstutz, John Chilton, Nebojša Tijanić, Hervé Ménager, et al. 2022. "Methods Included: Standardizing Computational Reuse and Portability with the Common Workflow Language." *Communications of the ACM* 65 (6): 54–63. https://doi.org/10.1145/3486897.

Goble, Carole, Sarah Cohen-Boulakia, Stian Soiland-Reyes, Daniel Garijo, Yolanda Gil, Michael R. Crusoe, Kristian Peters, and Daniel Schober. 2020. "FAIR Computational Workflows." *Data Intelligence* 2 (1-2): 108–21. https://doi.org/10.1162/dint_a_00033.

Marongiu, Paola, Barbara McGillivray, and Anas Fahad Khan. 2024. "Multilingual Workflows for Semantic Change Research." *Journal of Open Humanities Data* 10(1): 15. https://doi.org/10.5334/johd.179.