



HAL
open science

In Search of a Sustainable Model for Digital Heritage Repositories: A Case Study

Nathalie Fargier

► **To cite this version:**

Nathalie Fargier. In Search of a Sustainable Model for Digital Heritage Repositories: A Case Study. ELPUB 2018, Jun 2018, Toronto, Canada. hal-01816815v1

HAL Id: hal-01816815

<https://hal.science/hal-01816815v1>

Submitted on 15 Jun 2018 (v1), last revised 21 Jun 2018 (v2)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

In Search of a Sustainable Model for Digital Heritage Repositories: A Case Study

Nathalie Fargier

Introduction

- 1 A wide range of initiatives for developing research and data infrastructures have been funded in recent years. There is a growing concern amongst the academic community to maintain the resources invested beyond the period of the original research funding. If technical progress has been made to preserve the data themselves, few thinking and operational solutions exist for the institutions that create, disseminate, curate and preserve the data. How to ensure their existence over the medium or the long-term? This paper is a case study: it addresses the sustainability issues faced by *Persée*, a French platform dedicated to digitized documentary heritage that was launched in 2003. Through this example, the aim is to present, in practical terms, how an organization has to adapt and to change to sustain over time. Sustainability is a multi-faceted concept with no agreement on its means. For this paper, we define sustainability as

“(a) the continued operation of an organization that offers data collections and services, (b) where operations relates to technology, preservation, users, institutional relations, business models, and other facets, (c) in the face of on-going internal and external challenges which may or may not be resolved, (d) where stakeholders recognize continuity in the mission, data sources, and value of the data repository, (e) while keeping in mind that each of those elements might evolve over time in response to (c)” (Eschenfelder and Shankar, 2017).
- 2 We will give some background elements about *Persée* (historical and institutional origins, main achievements). Then, we will examine the different and complementary aspects of the strategy that was implemented in 2011 after a major crisis. *Persée* responded to this funding and institutional challenge by enhancing its technical framework, by fostering the engagement of the research community and the collaboration with other platforms

and infrastructures and by reshaping its strategy with the imperative to have reliable revenue sources to maintain its Open Access repositories. This experience shows that sustainability is not a steady situation but rather a continuous and resource intensive process for any organization.

Background elements

- 3 *Persée* was launched by the French Ministry of Higher Education and Research and Innovation (MESRI) in 2003 in order to develop a complete platform for the digitization, the documentation, the online publishing and the long-term preservation of back files of scientific journals in the field of humanities and social sciences. In 2005, the *Persée portal* (www.persee.fr) opened with 7 journals. Currently, 184 partnerships exist with academic publishers and 740.000 full texts documents and 570.000 illustrations are available in open access.
- 4 The original purpose of *Persée* was expanded according two directions: a disciplinary openness (earth and life sciences in addition to HSS) and the offer of a new range of services. *Persée* provides software tools, its digital platform and expertise to the researchers and, leads some projects in the field of the Digital Humanities and Open Data.
 - *Perseids* is the generic name of a range of websites developed by *Persée* in partnership with laboratories and heritage institutions as part of research projects. At the heart of a *Perseid*, there is a corpus that can be composed of primary sources, archives, grey literature, and published documents then, this content is digitized, precisely encoded and disseminated through a dedicated website with specialized index and tools for text analysis¹. Emphasis is given to the standardization of item description and document structuring to ensure the use and the reuse of the data from the digitized sources.
 - *Data Persée* (data.persee.fr) is the triplestore of *Persée*. All the metadata produced by *Persée* - through the *Persée portal* or the *Perseids* - are available by exploring a graph. Some collections have been enriched with data from external repositories (idRef, GBIF, Geonames, BnF, Cairo Gazetteer, etc.) and the pairings are certified by the *Persée* team.
- 5 As far as the content is concerned, *Persée* focuses on the documentary heritage. As defined by the Memory of the Word Programme (UNESCO): a document is composed of the information content (signs, codes, sounds, images) and the support in which it is recorded (it can have a large variety of form, it can be conserved, reproduced and transported). These characteristics exclude elements that are part of a fixed structure such as a building or natural site. *Persée* deals mainly with the “collections of old, rare and valuable documents housed by the libraries” (Desgraves, 1982) that is to say printed document (in particular, but not exclusively, academic journals, books, proceedings), archives, grey literature, maps and iconography. Researchers’ communities were closely associated to the definition of indexing methods and searching and visualization tools (Fargier, Néouze, 2008). Therefore, the originality of the *Persée portal* relies on an accurate structured description at the article level and a set of tools that are similar to those associated to born-digital publications. Researchers can search full-text documents, browse through tables of content, articles layouts or lists of illustrations, use a cited-by linking service, export citations or explore different indexes.

Action plan in favor of sustainability

Point of crisis

- 6 *Persée* was launched as a project after a call for bids. It lasted beyond the start-up phase and entered an operational phase because it demonstrated the proof-of-concept and met the researchers' expectations. During this period, from a financial point of view, *Persée* was supported by the French MESRI and a host institution (a university). In 2011, this latter argued that it was not its mission to support a national platform as *Persée*, began to complain about the cost of such a support and cut funding. At the time, this university was covering approximatively 40% of *Persée* costs. There were two responses to this threat: an emergency one with actions taken by others outside of *Persée* and a long-term one with an internal change of strategy and governance. A consortium was set-up with the French MESRI and three academic institutions². Complementary pathways were identified to ensure *Persée* sustainability:
- the transformation from a digital library to a data and services infrastructure;
 - a better inclusion within the existing landscape of research infrastructures, data repositories and digital platforms;
 - the diversification of the financing resources.
- 7 This approach considered that various factors had to be combined (technical, social, financial, organizational) with reciprocal influence, to achieve sustainability (Chowdury, 2014).

Technical actions

- 8 *Persée* relies on open standards, open source technologies, open access and open data in order to warranty data storage, retrieval and preservation.
- 9 Considering that "technical sustainability is directly related to the standards and best practices followed when creating the digital files" (Fyffe and Warner, 2005), *Persée* developed an integrated open source software named jGalith. It manages every step of the process to convert printed documents to electronic format: from digitization, through documentation, quality control, XML schemas compliance check, to online publishing and long term preservation. Printed documents are collected from publishers then digitized in high quality resolution (400 dpi, bitonal and grayscale), post-processed (OCR, image quality improvement), encoded (toc, author, abstract, keywords, illustrations, footnotes, annexes...) and linked (crossRef). Each document (article, review, preamble...) is individually described and Author indexes created. The process ends with four outputs: the *Persée* portal or the *Perseids* (bibliographic data + image mode + TEI + enriched PDF + index), the triplestore (RDF), the OAI-PMH (METS, MODS, DC) and the long term archive (OAIS model with PNG and XML files).
- 10 Furthermore, *Persée* has established partnerships based on interoperability: Open Edition (link between the past issues available on the *Persée* portal and the current ones on Open Edition Journals), Erudit (mutual access), HumaNum (Isidore).
- 11 All this accurate documentation and complete process contribute to enhance *Persée* brand image which is based on quality and reliable service.

Involving users' community

- 12 In 2011, the adequacy with researchers' needs was a key argument for maintaining *Persée* activity (the *Persée Portal* welcomed 17M visits in 2011 et 40M in 2017). The current phase is a further step to build a community that values, and not only uses, *Persée* services. It consists in setting up a large network of experts and a user committee. They will intervene at various levels: prioritization of materials to be digitized, feedback on new services and shared discussion about next steps. Users can be researchers or librarians and archivists from heritage institutions. In this way, *Persée* aims to show that its digitized content and services are clearly valuable and relevant for the academic community.
- 13 Another aspect to consider is the ability of *Persée* to develop relationships with other data repositories or infrastructures. Since the very beginning, a technical interoperability was established with the major francophone publishing platforms (Open Edition, Erudit and Cairn). The ongoing issue is to be interlinked with repositories and thesaurus more specialized³.

Strengthening institutional supports

- 14 Since 2011, *Persée* has been strengthening its lobbying. In 2013, *Persée* was created as a joint unit service supported by three major French universities and research organization. In 2016, *Persée* has been included in the French roadmap for the Research Infrastructures (RI) with *CollEx* (Collections of Excellence from academic and heritage libraries). This new infrastructure called *CollEx-Persée* aims at increasing the visibility, improving the access and promoting the use by researchers of the large collections held by research libraries and heritage institution. *Persée* is the key operator of this RI both for coordinating the digitization programs with researchers and managing scholarly enhancement and online dissemination of digitized corpora. In 2017, *Persée* was acknowledged as a French contributor to the European consortium DARIAH.

Towards a business model?

- 15 Developing a reliable revenue is the most challenging issue as *Persée* faces a twofold difficulty: building and maintaining, with an extensive level of curation, OA repositories of digitized documentary heritage without any payment-for-use.
- 16 We examine different funding models for OA digital data repositories (Kitchin, Collins and Frost-2015) and we identified four funding streams for *Persée*:
- Institutional revenue
The French MESRI provides the core operational costs through an annual subvention. And the members of the joint unit service pool human resources and funding, and facilitate capacity building.
 - Philanthropic revenue
Persée applies for endowment for private companies.
 - Research grants
Persée applies to research grants from national and European sources, with overheads to fund core services.

- White-label development

With the *Perseids*, *Persée* provides a customized service for the research laboratories and the heritage institutions. Each *Perseid* is powered by *Persée* (hardware, software and expertise) and, it is tailored and branded for a specific need and entity. The researchers pay one-off fee for subscription and includes maintenance and update costs.

- 17 In addition, *Persée* looks at another possibility with the fundraising on a voluntary basis including crowdsourcing.

Conclusion

- 18 Launched in 2003, *Persée* is a French platform dedicated to the digitized documentary heritage. It produces, curates and disseminates online large amounts of digitized content for researchers' communities. It underwent different threats of funding by capitalizing on its technical ability and by changing its strategy. If *Persée* managed to exist beyond the first phase of its creation, its sustainability depends of a mix of technical, social, financial, marketing and political mechanisms strongly interconnected. Different fronts are open-up simultaneously. The current situation can be described as a "temporary accomplishment" that is to say that "no institution is indefinitely sustainable (...) and sustainability is not a state of grace". *Persée*, as any institution, is always making adjustments to changing conditions and has more or less sustainable practices as far as funding, business planning, governance systems, organizational structure, etc. are concerned (Eschenfelder, Shankar, 2017).

BIBLIOGRAPHY

References

- Chowdury, G. (2014). "Sustainability of digital libraries: a conceptual model and a research framework." In *International Journal of Digital Libraries*, 14: 181–195.
- Eschenfelder K., Shankar K., Willimas R., Lanham A., Salo D., Zhang M. (2016). "What are we talking about when we talk about sustainability of digital archives, repositories and libraries?". *ASIST*, 14–18.
- Eschenfelder K. and Shankar K. (2017). "Organizational Resilience in Data Archives: Three Cases Studies in Social Scien Data Archives." In *Social Science Data Archives. Data Science Journal*, 16: 13, 1–12. <https://doi.org/105334/dsj-2017-012>
- Fargier, N., Néouze V. (2008). "Persée, for the promotion of French language scientific output in the human and social sciences." In: *IFLA2008 proceedings*.

French *General Code of Public Entity Properties*, L2112-1 10è. Available online: <https://www.legifrance.gouv.fr/affichCodeArticle.do?idArticle=LEGIARTI000006361198&cidTexte=LEGITEXT000006070299>

Kitchin R., Collins S., Frost D., (2015). "Funding models for Open Access digital data repositories", *Online Information Review*, vol. 39 Issue: 5: 664–681. <https://doi.org/10.1108/OIR-01-2015-0031>

KnowledgeExchange.Sustainability Index (2012). Available online: http://repository.jisc.ac.uk/6223/1/OA_Sustainability_Index.pdf

Maron, N. L., Pickle, S., & Marcum, D. (2013). *Searching for Sustainability Strategies from Eight Digitized Special Collections*. Ithaca, NY, 49p. <https://doi.org/10.18665/sr.22647>

Rapport au directeur du livre et de la lecture sur le patrimoine des bibliothèques. In *Bulletin des bibliothèques de France (BBF)*, 1982, n 12, 657–688. Available online: <http://bbf.enssib.fr/consulter/bbf-1982-12-0657-001>

Unesco, *Principes directeurs pour la sauvegarde du patrimoine documentaire*, rédigés par Ray Edmondson, édition révisée 2002. Paris : UNESCO, 2002. 79 p.

NOTES

1. Example of a Perseid: athar.persee.fr
 2. ENS de Lyon, CNRS and Université de Lyon
 3. The Cairo Gazetteer (cairogazetteer.fr/invisu/) or GBIF (www.gbif.org)
-

INDEX

Keywords: digitized documentary heritage, Practitioner's paper, data repositories, open access, open data, sustainability

AUTHOR

NATHALIE FARGIER

ENS de Lyon, Persée, France
nathalie.fargier@persee.fr
(corresponding author)