

Scaling up Arts and Humanities: The DARIAH Approach to Data and Services

Aurélien Berra, Matej Durco, Chad Gaffield, Nicolas Larrousse, Paulin Ribbe, Mike Priddy, Carsten Thiel

▶ **To cite this version:**

Aurélien Berra, Matej Durco, Chad Gaffield, Nicolas Larrousse, Paulin Ribbe, et al.. Scaling up Arts and Humanities: The DARIAH Approach to Data and Services. DH 2017, Aug 2017, Montréal, Canada. hal-02152986

HAL Id: hal-02152986
https://hal.science/hal-02152986

Submitted on 11 Jun 2019

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.

Scaling up Arts and Humanities: The DARIAH Approach to Data and Services

Aurélien Berra
aurelien.berra@u-paris10.fr
Université Paris-Ouest Nanterre La Defense, France

Matej Durco
matej.durco@oeaw.ac.at
Austrian Center for Digital Humanities, ÖAW, Austria

Chad Gaffield
gaffield@uottawa.ca
University of Ottawa, Canada

Nicolas Larrousse
nicolas.larrousse@huma-num.fr
Centre National de la Recherche Sienctifique, France

Paulin Ribbe
paulin.ribbe@huma-num.fr
Centre National de la Recherche Sienctifique, France

Mike Priddy
DANS - KNAW, Netherlands
mike.priddy@dans.knaw.nl

Carsten Thiel
thiel@sub.uni-goettingen.de
SUB Göttingen, Germany

Introduction

The panel contributions will introduce and discuss several central aspects of DARIAH-EU aimed to support emerging practices in the Digital Humanities, to disseminate consolidated practices and to provide tools for navigation through the rich and changing landscape of the productions of research communities, be they research data, tools, methods or other assets.

Our focus point is how to boost research on a number of different levels: ensuring the sustainability of the infrastructure, its "scaling up" to enable better access to research data, community-building, linking teaching activities and weaving the network of affiliated initiatives.

Panel Overview

Chad Gaffield (Moderator)

Panel participants as listed below will introduce their topic for 5 minutes each, followed by a general discussion.

It has been much debated in the Digital Humanities communities whether scaling up – in terms of shared resources, formation of team science, cross-domain research – is at all appropriate for the way knowledge is produced in the various humanistic disciplines. Instead of tackling again this general problem, this panel takes a much more practical approach. There can be no doubt that research practices in the humanities and social sciences are profoundly changing (Wouters et al., 2013). The turn to digital methods happens in a relationship of mutual dependency and co-evolution with the infrastructures traditionally feeding humanities research: collections, libraries and archives. This development is most visible in large-scale research infrastructures such as CLARIN and DARIAH, and it is acknowledged by their respective funding.

This panel presents the efforts made within the DARIAH (Digital Research Infrastructure for the Arts and Humanities) community in its mission to further develop the Arts and Humanities research infrastructure, now that the Digital Humanities have long left the incubation stage or the phase of early adopters. The field is still growing but also starting to differentiate and specialise. The "scaling up" of the humanities refers to the growth of the community – as a social, cognitive and technological network –, the increasing availability of Big Data, the ongoing penetration of standards and the use of consolidated tools. While the mission is quite clear, its implementation is a far more complex process.

We will focus on the relevance of the **community** and **sustainability** as driving factors in building the research infrastructure**.** The contributions report about lessons learned, examine achievements and open questions, and aim to engage with the audience about their expectations, experiences and visions of what a research infrastructure (RI) for the Arts and Humanities should look like.

These two aspects are not arbitrary. Community-building responds to the motto of DARIAH: *services developed for researchers by researchers* and is relevant to formulate needs, to find appropriate technical solutions but also to encourage continuous (re)use by means of education and training. DARIAH itself is a pan-European infrastructure, driven by grass-rooted needs of researchers, and governed by national representatives, entailing processes of collaborative evaluation. In addition to community-

building within the RI, DARIAH has always aimed to expand and strengthen the external cooperations with a broader ecosystem involving both affiliated RI projects (EHRI, CENDARI, PARTHENOS) and sister initiatives like CLARIN, as well as with research infrastructures that specialize in providing basic services (eduGain, EGI or EUDAT). Research Infrastructures address both their users and producers with their immediate needs. At the same time they have to be stable, reliable and of good quality, which needs to be addressed continuously throughout its life. Another primary concern of DARIAH is the focus on reuse of research data and ensuring their longevity. This is where the sustainability of software and services comes into play, ensuring the continued availability of the outputs from shorter-lived projects in the long run.

Building a durable infrastructure

Nicolas Larrousse and Paulin Ribbe

During the last decades, digital made his way into the Arts and Humanities research world at every stage of research projects. As a consequence, the way researchers work today has considerably evolved. They can no longer work isolated and their work requires more and more sophisticated tools to deal with research material. There is a need to scale up and build infrastructures dedicated to Arts and Humanities. DARIAH is one possible answer to these new requirements.

DARIAH is organized as an European framework for research infrastructures based upon a consortium of states which agree to support the infrastructure in the long term. It was designed to establish sustainable digital services and develop common practices. Such objectives reflect a long-term vision which would not fit in a "classical" project funded for three years.

Within the consortium, ongoing efforts help to concretely build the infrastructure and to better answer those fundamental questions: what is an infrastructure for Arts and Humanities and how can we build it on the European level?

DARIAH contributions and the reference architecture they inhabit

Mike Priddy (Co-authors: Francesca Morselli, Lisa de Leeuw, Andrea Scharnhorst)

DARIAH constitutes itself as a pan-European infrastructure and relies on national contributions from its member states. 90% of these contributions are in-kind and represent the research outputs that each member country contributes to the infrastructure, showing the richness and diversity of the European research landscape. DARIAH is creating a way to collect, disseminate, review, assess and evaluate the contributions made to the infrastructure and community.

As researchers in the Arts and Humanities familiarise with the creation, management, storage and reuse of data, they need a technical infrastructure that can support such activities. On the other hand, a mere technical support seems insufficient to sustain their research activities, as research communities play a crucial role in the development of the infrastructure by developing processes and offering services and resources. An approach which supports the knowledge creation process as well as an agreement on standardised processes (e.g. data exchange) is therefore highly needed.

With a reference architecture we aim to create a common language among the European member countries per individual type of contribution to describe and communicate DARIAH contributions, as well as provide a reference for other projects and infrastructures in the Arts and Humanities.

Re-usability of tools and services

Carsten Thiel

The software tools and solutions powering the services provided by infrastructures are subject to their individual life cycles. While most projects realised through research grants focus on the initial phases – from inception and initial design, through user testing and design revisions, to beta releases –, delivering a "final" product by the project's end remains the primary objective. But the software's life does not end there. As researchers use the tools, new needs and requirements can arise that need to be addressed and implemented. At the same time the IT world evolves and technologies change. This leads to an increased need for adopting existing services and their underlying software for continued usability. Finding resources to sustain the services and to ensure their availability and accessibility requires a change already to the initial objectives in the development phase (see Doorn, Aerts, Lusher 2016). Designing software from the get-go with re-usability in mind ensures the long-term benefit of enabling future contributions and adoptions needed by the community as a whole.

Teaching: where and how?

Aurélien Berra

DARIAH's "Research and Education" programme has contributed or planned several services which emphasize how data are central in building up a sense of community and deliberate, open, collaborative research strategies.

Three projects will be discussed. The Digital Humanities Course Registry is a collaborative database aiming to provide up-to-date information on European courses to students looking for a programme and to teachers and researchers interested in evaluating the state of the field at a national or international level. The initiatives comprised in the "Integration of Training Material" project offer a framework for sharing and preserving Digital Humanities teaching materials (through the user-centred design of #dariahTeach, an open platform for extensible and translatable contents, http://dariah.eu/teach) as well as a philological toolbox based on the principle of re-use and adaption (Biblissima's BaOBab, a catalogue of online tools, guidelines and tutorials focused on manuscripts traditions, http://outils.biblissima.fr). The Master Classes scheduled for 2017 in Maynooth (Public Humanities), Florence (Evaluating Digital Scholarship), Berlin and Paris (Data Reuse in the Humanities) are designed as experiments in developing a format which will engage advanced users and foster a collective reflection on "data fluidity" (see Romary, Mertens & Baillot 2016).

DARIAH & affiliates (aka "DARIAH's outer network")

Matej Durco

Despite the vast diversity of the individual research agendas, at a certain level all research infrastructures and initiatives need the same kind of basic facilities represented by the lower levels of the architectural stacks – stable hosting and storage of data, processing and computational capacity, secure authentication to build trust, stable referencing of resources, etc. DARIAH has always embraced the idea of cooperation between the different initiatives to harness the synergies emerging from this shared needs and has strong ties both "horizontally" to affiliated RI-projects (EHRI, CENDARI, PARTHENOS) and to sibling infrastructures like CLARIN, as well as "vertically" to networks that specialize on the actual provision of these basic services like eduGain, EGI, or EUDAT. This is in line with DARIAH's self-image as a facilitator that speaks the language of the researchers, but also has the "global critical mass" and expertise to approach and negotiate with large technical infrastructures.

Bibliography

Borgman, C.L., Edwards, P.N., Jackson, S.J., Chalmers, M.K., Bowker, G.C., Ribes, D., Burton, M. (2013) *Knowledge Infrastructures: Intellectual Frameworks and Research Challenges*. Report. http://pne.people.si.umich.edu/PDF/Edwards_etal_2013_Knowledge_Infrastructures.pdf

Doorn, P., Aerts, P., Lusher, S. (2016) *Research Software at the Heart of Discovery*. Technical Report, DANS, NLeSC, 2016. https://www.escienceCenter.nl/pdf/Software_Sustainability_DANS_NLeSC_2016.pdf

Fihn, J., Gnadt, T., Hoogerwerf, M., Jerlehag, B., Lenkiewicz, P., Priddy, M., Shepherdson, J., *DASISH - Reference Model for Social Science and Humanities Data Infrastructures*. European Commission, 2016. https://sites.google.com/a/dans.knaw.nl/reference-model-for-ssh-data-infrastructure/

Romary, L., Mertens, M. et Baillot, A. (2016) *Data Fluidity in DARIAH – Pushing the Agenda Forward*, *Bibliothek Forschung und Praxis*, 39, 3, 2016, p. 350–357, https://hal.inria.fr/hal-01285917.

Wouters, P., Beaulieu, A., Scharnhorst, A., & Wyatt, S. (eds) (2013). *Virtual knowledge: Experimenting in the humanities and the social sciences*. Cambridge, MA: The MIT Press.