

## Connecting the Humanities through Research Infrastructures

Sheena Bassett, Leon Wessels, Steven Krauwer, Bente Maegaard, Hella Hollander, Femmy Admiraal, Laurent Romary, Frank Uiterwaal

► **To cite this version:**

Sheena Bassett, Leon Wessels, Steven Krauwer, Bente Maegaard, Hella Hollander, et al.. Connecting the Humanities through Research Infrastructures. 4th Digital Humanities in the Nordic Countries (DHN 2019), Mar 2019, Copenhagen, Denmark. hal-02047512

**HAL Id: hal-02047512**

**<https://hal.inria.fr/hal-02047512>**

Submitted on 25 Feb 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.



# Connecting the Humanities through Research Infrastructures

Sheena Bassett<sup>1</sup>, Leon Wessels<sup>2,4</sup>, Steven Krauwer<sup>2,4</sup>, Bente Maegaard<sup>3,4</sup>,  
Hella Hollander<sup>5,7</sup>, Femmy Admiraal<sup>5,7</sup>, Laurent Romary<sup>6,7</sup>, Frank Uiterwaal<sup>8</sup>

<sup>1</sup>PIN Scrl, University of Florence, Piazza Ciardi 25, 59100 Prato, Italy

<sup>2</sup>Utrecht University, The Netherlands

<sup>3</sup>University of Copenhagen, Denmark

<sup>4</sup>CLARIN ERIC, Utrecht, The Netherlands

<sup>5</sup>Data Archiving and Networked Services (DANS), The Netherlands

<sup>6</sup>Institut National de Recherche en Sciences du Numérique (INRIA), France

<sup>7</sup>DARIAH ERIC, Paris, France

<sup>8</sup>NIOD Institute for War, Holocaust and Genocide Studies, The Netherlands

**Keywords:** cross-discipline, policies, language resources, Research Infrastructures, standards.

## 1 Introduction

Several Research Infrastructures (RIs) exist in the Humanities and Social Sciences, some – such as CLARIN, DARIAH and CESSDA – which address specific areas of interest, i.e. linguistic studies, digital humanities and social science data archives. RIs are also unique in their scope and application, largely tailored to their specific community needs. However, commonalities do exist and it is recognised that benefits are to be gained from these such as efficient use of resources, enabling multi-disciplinary research and sharing good practices. As such, a bridging project PARTHENOS has worked closely with CLARIN and DARIAH as well as ARIADNE (archaeology), CENDARI (history), EHRI (holocaust studies) and E-RIHS (heritage science) to identify, develop and promote these commonalities. In this paper, we present some specific examples of cross-discipline and trans-border applications arising from joint RI collaboration, allowing for entirely new avenues of research.

## 2 CLARIN: Language is not just for linguists

Linguists study language, but language is not just studied by linguists. As the European Research Infrastructure Consortium for language resources and technology, CLARIN offers digital language data and tools to scholars, researchers, students and citizen-scientists from all disciplines [1]. CLARIN's vision is to make all digital language data collections and tools from all over Europe and beyond accessible through a single sign-on online environment [2]. In this section we will highlight two initiatives particularly aimed at stimulating cross-border and cross-disciplinary collaboration using language resources and tools.

## 2.1 Resource Families

To provide user-friendly overviews of available corpora (i.e. collections of digital language data), including information on language, size, annotation, availability and licence, CLARIN has introduced the Resource Families initiative (<https://www.clarin.eu/resource-families>). Resource Families are data collections that display a high degree of maturity and homogeneity, that are available for most EU languages and are a rich source of social and cultural data [3]. Currently CLARIN offers overviews of seven Resource Families: language data collected from social media, historical corpora, second-language learner corpora, newspaper corpora, parallel corpora, parliamentary corpora and spoken corpora. Resource Families are relevant for multiple scientific disciplines in the digital humanities and social sciences. Parliamentary data for instance, is used by political scientists, historians, and linguists to study complex phenomena such as Euroscepticism and gender stereotypes [4].

## 2.2 Language Resource Switchboard

Equally important to language resources are the means to process them. The CLARIN Language Resource Switchboard (<https://switchboard.clarin.eu>) helps researchers to connect resources to the tools to process them [5]. It provides tools for different types of parsing and annotation, analysis, distant reading, etc. By providing uniform, manually-annotated overviews of corpora and a range of advanced, robust tools for processing and analysing these corpora, CLARIN enables researchers from various disciplines to find cross-national and cross-language patterns in language data far more quickly and easily, on a much larger scale than before.

## 3 DARIAH: The Standardization Survival Kit

Whereas some scholarly groups have been very active in defining and using standards in their practices, the project observed that there is a general lack of precise knowledge in the standardization landscape, especially from newcomers to digital methods [6]. A solution to this information gap is the Standardization Survival Kit (SSK), which assists scholars in finding information on which standards may play a role in their specific scholarly activities and is designed around research scenarios in line with the requirements gathered earlier in the project [7]. The current scenarios illustrate both the importance of standards in research processes, as well as the usefulness of designing an online environment which allows researchers to browse through scenarios and access relevant reference material [8]. Additionally, the project developed material (a leaflet and several videos) to make scholars aware of standards. The resulting digital environment was designed in close collaboration with digital specialists from the various communities and RIs represented in the PARTHENOS project with the support of DARIAH who initiated the standards work. The current scenarios cover all types of scholarly domains and methodologies: managing field surveys, creating a textual or a musical corpus, managing archival information as well

as using laser techniques for conservation practice in heritage science and includes four specific linguistic scenarios. An important expected future feature will be the possibility to couple the maintenance of validated scenarios with the freedom for scholars to publish their own scenarios through the platform.

#### **4 PARTHENOS: The need for common policies and implementation strategies**

As a cluster project, it has been PARTHENOS' aim to make optimal use of experiences gained throughout all disciplines. Shared best practices and formal policies make it possible to answer research questions in a more overarching way. Not only does this enable researchers from different universities and geographic areas to refer to the same policies in a shared language and mutual understanding. Also, researchers can more easily reuse each other's results, allowing them to build on each other's observations and knowledge. It is PARTHENOS' aim to work towards the integration of research data policies across the disciplines in the humanities, as well as to increase their interdisciplinary exchanges. This is a broad mission, for which one generic solution could never suffice, also due to the current, distributed and previously unmapped landscape of research policies. However, the FAIR Principles have been applied as an underlying framework and three important forms of project output provide solutions to this situation, helping all stake holders in making more research data more findable, accessible, interoperable and re-usable [9]:

- 1.) **The PARTHENOS Policy Wizard.** The Policy Wizard is a tool which helps researchers to discover which data policy applies best to their particular data. It allows the user to select an area of interest, for example History, and find a range of policies suggesting how data should be collected, processed, stored and shared with other researchers.
- 2.) **Data Management Plan template.** The PARTHENOS DMP template builds on the Horizon2020 version and has been tailored for the Humanities using expert input. It provides built-in guidance with some pre-filled options for completing a DMP.
- 3.) **PARTHENOS Guidelines.** Twenty high-level guidelines are offered as common recommendations to build bridges between different, although tightly interrelated, fields and stakeholders (researchers and data repositories) within the Humanities by the harmonisation of policy definition and their implementation.

#### **5 Conclusions**

Collaboration between RIs for the purpose of developing synergy on common aspects of research has helped initiate harmonisation and led to the development of useful

tools whilst also reducing barriers for cross-discipline and cross-border research. Other collaborative examples in PARTHENOS are shared international contacts and a Training Suite that addresses the need for knowledge about RIs as well as shared Humanities topics.

## References

1. Hinrichs, E., Krauwer, S.: The CLARIN Research Infrastructure: Resources and Tools for eHumanities Scholars. In: Proceedings of the Ninth International Conference on Language Resources and Evaluation, pp. 1525-1531. European Language Resources Association, Reykjavik (2014).
2. De Jong, F., Maegaard, B., De Smedt, K., Fišer, D., Van Uytvanck, D.: CLARIN: Towards FAIR and Responsible Data Science Using Language Resources. In: Proceedings of the Eleventh International Conference on Language Resources and Evaluation, pp. 3259-3264. European Language Resources Association, Miyazaki (2018).
3. Fišer, D., Lenardič, J., Erjavec, T.: CLARIN's Key Resource Families. In: Proceedings of the Eleventh International Conference on Language Resources and Evaluation, pp. 1320-1325. European Language Resources Association, Miyazaki (2018).
4. More information can be found on the PARTHENOS training module Boosting Digital Humanities Research with Parliamentary Data (<http://training.parthenos-project.eu/test-2/digital-humanities-research-questions-and-methods/researching-parliamentary-records-in-the-digital-humanities/boosting-digital-humanities-research-with-parliamentary-data/>).
5. Zinn, C.: The CLARIN Language Resource Switchboard. In: Abstracts of the CLARIN Annual Conference 2016, Aix-en-Provence (2016).
6. Illmayer Klaus and Marie Puren. 'How to work together successfully with e-Humanities and e-Heritage Research Infrastructures' (PARTHENOS Webinar). 3rd cycle. PARTHENOS eHumanities and eHeritage Webinar Series, Germany. 2018. <http://training.parthenos-project.eu/sample-page/ehumanities-eheritage-webinar-series/webinar-work-with-research-infrastructures/> (last accessed: 20 September 2018).
7. This is an example of the dialogue between the team which collected requirements and the team which builds a tool to fulfil the addressed needs (see under "an approach based on the needs of communities").
8. For additional background, please see: Laurent Romary, Emiliano Degl'innocenti, Klaus Illmayer, Adeline Joffres, Emilie Kraikamp, et al. Standardization survival kit (Draft). [Research Report] Deliverable 4.1 (2016); Laurent Romary, Piotr Banski, Jack Bowers, Emiliano Degl'innocenti, Matej Ďurčo, et al., Report on Standardization (draft). [Technical Report] Deliverable 4.2 (2017).
9. The FAIR principles provide an important frame of reference for the output PARTHENOS generates on the topic of research data management. See: Mark D. Wilkinson, Michel Dumontier, IJsbrand Jan Aalbersberg and others, 'The FAIR Guiding Principles for scientific data management and stewardship,' *Scientific Data* 3 (2016).