



HAL
open science

HumaReC: Continuous Data Publishing in the Humanities

Claire Clivaz, Sara Schulthess, Anastasia Chasapi

► **To cite this version:**

Claire Clivaz, Sara Schulthess, Anastasia Chasapi. HumaReC: Continuous Data Publishing in the Humanities. ERCIM News, 2017, 111, pp.21-22. hal-02519359

HAL Id: hal-02519359

<https://hal.science/hal-02519359v1>

Submitted on 27 Mar 2020

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.



Distributed under a Creative Commons Attribution 4.0 International License

With the maturation of a base model, creation of a declarative mapping tool and language, a generalising query function and a model and method for managing integration processes, we believe that the key elements for meeting these challenges now lie in place.

Links:

- [L1] <http://www.cidoc-crm.org/>
- [L2] <http://www.researchspace.org/>
- [L3] <http://www.parthenos-project.eu/>

References:

- [1] ISO: ISO 21127: 2014, Information and documentation – a reference ontology for the interchange of cultural heritage information, 2nd edn., 2014.
- [2] N. Minadakis, Y. et al.: “X3ML Framework: An effective suite for supporting data mappings, Extending. Proceedings of the Workshop on Extending, Mapping and Focusing the CRM co-located with 19th International Conference on Theory and Practice of Digital Libraries

- (2015), Poznań, Poland, September 17, 2015”, CEUR Workshop Proceedings 1656, 1-12, 2016.
- [3] K. Tzompanaki & M. Doerr: Fundamental Categories and Relationships for Intuitive querying CIDOC-CRM based repositories, Technical Report, 2012.

Please contact:

George Bruseker
 Centre for Cultural Informatics
 ICS-FORTH, Greece
 bruseker@ics.forth.gr

HumaReC: Continuous Data Publishing in the Humanities

by Claire Clivaz, Sara Schulthess and Anastasia Chasapi (SIB)

HumaReC, a Swiss National Foundation project, aims to test a new model of digital research partnership: from original document source to publisher. Its object of study is a trilingual, 12th century, New Testament manuscript, which is used by HumaReC to test continuous data publishing.

HumaReC [L1] is a Vital-DH@Vital-IT project funded by the Swiss National Foundation. Under the leadership of Claire Clivaz, it started in October 2016 and has been running for two years. The project is based at Vital-IT (Lausanne, CH), under the guidance of group leader Ioannis Xenarios from the Swiss Institute of Bioinformatics. The team is composed of a mixture of humanities and computing scholars; Sara Schulthess is the main researcher.

The aim of HumaReC is to investigate how humanities research is reshaped and transformed by the digital rhythm of data production and publication; it also aims to establish the best practices for Digital Humanities research. As a test-case, the study is focussed on a

unique, trilingual New Testament manuscript: Marciana Gr. Z. 11 (379), written in Greek, Latin and Arabic. In the spirit of the OA2020 initiative [L4], continuous data publishing is being tested in partnership with all the research network stakeholders: the Marciana library (Venice, Italy), the Edition Visualization Technology (EVT) [L2], the Transkribus [L3] research teams, and the publisher Brill.

Rhythm is a central notion in the structure of HumaReC and we have chosen this key-concept, based notably on Meschonnic analysis [1], to observe the changes happening in digital humanities research which has been premised on printed culture for a long time. A two to three year research project in humani-

ties has traditionally been characterised by the writing, editing and publication of a final, printed book, often delayed to a certain date after the end of the project. This delay was even considered proof of authentic, high-level research in humanities, certified by an established book series. The digital transition is creating a completely new research paradigm in part due to the publishing of formats such as videos, short messages or draft papers, social media and blogs, all before the research is even completed and peer-reviewed. How is it possible to develop certified and continuous data publishing digital research in the humanities?

As the project’s first step, a virtual research environment was created for HumaReC, allowing the research process and results to be made continuously available. It provides a manuscript viewer in fully open access that includes quality images of the manuscript and three columns of transcriptions (see Illustrations below). The manuscript viewer is based on EVT open source technology and is a development of a previous project’s viewer [2]. The improved viewer offers additional features such as linking between text and image. In addition to this, an annotation system will allow users to directly comment on the manuscript viewer. Secondly, Transkribus, a

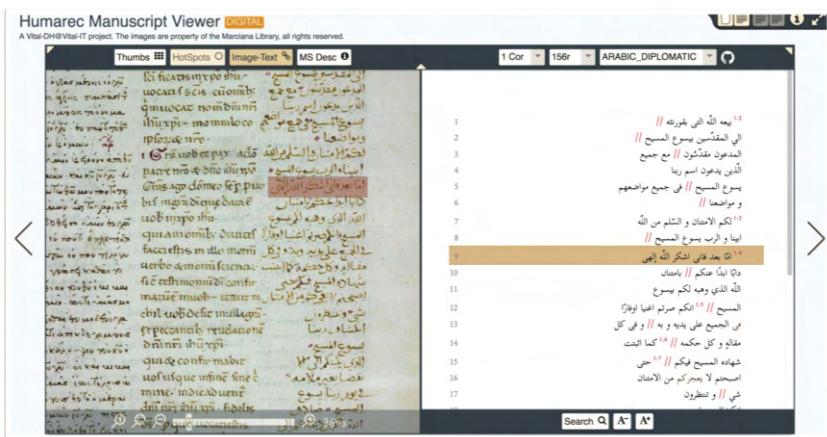


Figure 1: Manuscript viewer f. 156r © Marciana Library all rights reserved.

Handwritten Text Recognition tool, will also be tested by HumaReC: a certain number of words transcribed by hand allows a learning machine to be trained to recognise specific writing in a manuscript. The results are forthcoming: we will be able to compare the time taken by purely hand transcription with results produced by a human/machine team.

Finally, three publication formats were chosen for the continuous dissemination of the data:

- The virtual research environment itself; it received an ISSN (2504-5075) from the Swiss National Library: all the published material associated with the project can be referred to with this number. An international editorial board is providing project feedback and input on its research results.
- The research blog is an important interactive continuous publishing process. We regularly update the blog

about the development of the project and the research results. We also encourage discussions by being present on social media (Facebook and Twitter).

- The web-book, continuously written in open access, summarises the research in a long, structured text, similar to a conventional monograph but related to the data. It is produced in partnership with the publisher Brill. At the end of the project, it will be peer-reviewed, and hopefully published with its own ISSN by Brill.

We are confident that HumaReC will establish a new research and publishing model, including potential commercial developments for all interested publishers.

Acknowledgement

Many thanks to Harley Edwards for his English proof-reading, to the Marciana library for the image copyright, to the

reviewers for their useful remarks, and to Martial Sankar and Ioannis Xenarios for their collaboration and support in our Vital-DH projects@Vital-IT (SIB).

Links:

- [L1] <https://humarec.org>;
- <http://p3.snf.ch/project-169869>
- [L2] <http://evt.labcd.unipi.it/>
- [L3] <http://transkribus.eu>
- [L4] <http://oa2020.org>

References:

- [1] C. Clivaz et al.: “Editing New Testament Arabic Manuscripts in a TEI-base: fostering close reading in Digital Humanities”, JDMDH 3700, 2017.
- [2] H. Meschonnic: “Critique du rythme. Anthropologie historique du langage”, 1982.

Please contact:

Claire Clivaz, Swiss Institute of Bioinformatics, Switzerland, +41216924060, claire.clivaz@sib.swiss

A Data Management Plan for Digital Humanities: the PARTHENOS Model

by Sheena Bassett (PIN Scrl), Sara Di Giorgio (MIBACT-ICCU) and Paola Ronzino (PIN Scrl)

Understanding how data has been created and under which conditions it can be reused is a significant step towards the realisation of open science.

PARTHENOS [L1] is a Horizon 2020 project funded by the European Commission that aims at strengthening the cohesion of research in the broad sector of linguistic studies, cultural heritage, history, archaeology and related fields through a thematic cluster of European research infrastructures. PARTHENOS is building a cross-disciplinary virtual environment to enable researchers in the humanities to have access to data, tools and services based on common policies, guidelines and standards. The project is built around two European Research Infrastructure Consortia (ERICs) from the humanities and arts sector: DARIAH [L2] (research based on digital humanities) and CLARIN [L3] (research based on language data), along with ARIADNE [L4] (digital archaeological research infrastructure), EHRI [L5] (European Holocaust research infrastructure), CENDARI [L6] (digital research infrastructure for historical research),

CHARISMA [L7] and IPERION-CH [L8] (EU projects on heritage science) and involves all the relevant integrating activities projects.

Since 2016, the Horizon 2020 Programme has produced guidelines on FAIR data management [L9] to help Horizon 2020 beneficiaries make their research data findable, accessible, interoperable and reusable (FAIR) [L10]. Funded projects are requested to deliver an implementation of DMP which aims to improve and maximise access to and reuse of research data generated by the projects. This is in line with Commission’s policy actions on open science to reinforce the EU’s political priority of fostering knowledge circulation. Open science is in practice about “sharing knowledge as early as practically possible in the discovery process” and because DMPs gather information about what data will be created and how, and outline the plans for sharing

and preservation, specifying the nature of the data and any restrictions that may need to be applied, these plans ensure that research data is secure and well-maintained during a project and beyond, when it might be shared with others. DMPs are key elements to knowledge discovery and innovation and to subsequent data and knowledge integration and reuse.

Special attention has been paid to the development of a PARTHENOS data management plan which builds on the Horizon2020 DMP template. This has resulted in a template (draft) which aims to address the domain-specific procedures and practices within the humanities, taking into consideration standards and guidelines used in data management that are relevant for PARTHENOS specific research communities, which includes archaeologists, historians, linguists, librarians, archivists, and social scientists.