

Title

Exploring the reuse of data in the humanities by means of asynchronous collaboration and authorship in nodegoat.

Abstract

nodegoat is a web-based research environment that facilitates an object-oriented form of data management with an integrated support for diachronic and spatial modes of analysis. This research environment has been designed to allow scholars to determine and design custom relational database models. The environment can be used in self defined collaborative configurations with varying clearance levels for different groups of users. Due to the focus on relations and associations between heterogeneous types of objects, the platform is equipped to perform analyses spanning multitudes of objects.

In order to facilitate a practice of reuse of datasets in the humanities, an ecosystem has to exist in which scholars can publish their datasets, correctly attribute this data according to the roles played by each author, share these datasets and allow for various scenarios of reuse. Current publication channels do not allow for complex authorship attribution (Nyhan and Duke-Williams, 2014). In this paper we will explore reuse scenarios by means of an object-oriented referencing system in which datasets, data selection, entities and records are all referenceable objects with uniquely identifiable authors. Once a reference has been made to any of these objects, a citation is automatically determined based on the position of the referenced object in the network and all their corresponding authors. This object-oriented referencing system paves the way for various scenarios of reuse and processes of asynchronous collaboration.

Publishing research data transcends traditional citation practices on three levels. Firstly, publishing data may happen before any synthesised text is in sight. Secondly, research outcomes in the form of data can have an extended life cycle that stretches far beyond the reach of a static text. Thirdly, research data that would not have been included in the final syntheses can still be published as data and find its way to a wider audience. These opportunities show the potential of publishing data in the humanities. Still, a number of challenges has to be overcome to arrive at the position in which scholars in the humanities will directly publish their data.

One of the most prominent challenges we still face is the awarding of academic credit for publishing datasets (Nowviskie, 2011). As Claudine Mouline has stated, we need a 'change of publication cultures and recognition of these new publication cultures as equal to traditional ones'. Next to the monograph and the article, results and achievements in the form of the database, data visualisation, the scientific blog and micropublications in different forms should be recognised as well (Mouline, 2013).

In 2014, Dutch research institute Huygens ING together with the University of Amsterdam (UvA), the Free University (VU), the Royal Dutch Institute in Rome (KNIR) and LAB1100 led by Charles van den Heuvel ran a project that relied on asynchronous collaboration.¹ For this project, 'Mapping Notes and Nodes in Networks', multiple existing

¹ <https://www.huygens.knaw.nl/mapping-notes-and-nodes-in-networks/?lang=en>

datasets were brought together and manually enriched in order to map meaningful relationships between artists and intellectuals by combining biographical data with relevant contextual information for the history of the creative industry. Three complementary, but heterogeneous datasets Biographical Reference Works (Huygens ING), Ecartico (UvA) and Hadrianus (KNIR) were integrated in nodegoat.²

In the course of the project a number of researchers carried out individual research projects within the research environment that contained the three datasets. This led to a productive form of asynchronous collaboration as all the biographical data about artists and intellectual available in the existing datasets was used as context for new research questions. This prosopographical information was subsequently enriched with information about society membership in Italy (the Accademie). By adding this data, research questions regarding weak ties between these societies could be explored.

Lisa Spiro has developed a comprehensive overview of collaborative practices in the digital humanities (Spiro, 2012, 2009). She has identified three scenarios in which collaboration takes place: “(1) communicating and exchanging knowledge through participatory online environments; (2) building digital collections of primary and/or secondary scholarly resources; and (3) developing computational methods for analyzing humanities data” (Spiro, 2012. p. 45). In her work, she has mainly focused on synchronous forms of collaboration in which research groups or participatory projects work together on a set of resources. Although these challenges are closely related to the concept of asynchronous collaboration, they only apply on a closed environment in which the project team, project data and collaborators all work together. We propose a form of asynchronous collaboration that is platform independent. Platform independency ensures the sustainability of the datasets and fosters an extensive applicability of the data.

The effective reuse of the data functions as the dividing line between asynchronous collaboration and traditional citation practices. Whereas traditional citation practices also reference to other scholarly resources and in doing so extend their lifespan and validity, the underlying data is never reused. Although we can cite *The Waning of the Middle Ages* of Dutch historian Johan Huizinga, we will never *reuse* his research notes or card catalogue. Since the emergence of digital research tools, historians and other scholars in the humanities have the ability to create digital card catalogue systems (databases). Asynchronous collaboration aims to open up these vast resources of rich data in order to establish an ecosystem of reuse and multiple forms of authorship.

In traditional forms of scholarship in the humanities, the claim on authorship is closely connected to the composition of a narrative in which the syntheses of the research project are brought together. We propose new forms of asynchronous authorship that are connected to the publication of datasets. These forms of authorship are in essence hybrid as the creation process of a dataset is often a collaborative process. Moreover, once reuse of these datasets takes place, new forms of authorship emerge that can span multiple layers of conceptualisation, creation, selection and publication processes. The process of

² <http://www.biografischportaal.nl/>, <http://www.vondel.humanities.uva.nl/ecartico/>, <http://hadrianus.it/>.

asynchronous collaboration is to be regarded as an additional collaborative methodology for the humanities and poses new opportunities for scholarly communication.

References

MOULINE, Claudine, 2013, Je t'aime, moi non plus. Career, Financing and Academic Recognition in the Digital Humanities (#dhiha5) [online]. 12 June 2013. [Accessed 6 November 2014]. Available from: <http://annotatio.hypotheses.org/303>

NOWVISKIE, Bethany, 2011, Where Credit Is Due: Preconditions for the Evaluation of Collaborative Digital Scholarship. *Profession* [online]. 2011. p 169–181. DOI 10.1632/prof.2011.2011.1.169. Available from: <http://www.mlajournals.org/doi/pdf/10.1632/prof.2011.2011.1.152>

NYHAN, Julianne and DUKE-WILLIAMS, Oliver, 2014, Is Digital Humanities a collaborative discipline? Joint-authorship publication patterns clash with defining narrative [online]. 10 September 2014. [Accessed 6 November 2014]. Available from: <http://blogs.lse.ac.uk/impactofsocialsciences/2014/09/10/joint-authorship-digital-humanities-collaboration/>

SPIRO, Lisa, 2009, Collaborative Authorship in the Humanities [online]. 21 April 2009. [Accessed 6 November 2014]. Available from: <http://digitalscholarship.wordpress.com/2009/04/21/collaborative-authorship-in-the-humanities/>

SPIRO, Lisa, 2012, Computing and Communicating Knowledge: Collaborative Approaches to Digital Humanities Projects in : *Collaborative Approaches to the Digital in English Studies*. Old Main Hill : Computers and Composition Digital Press. p 44-82. ISBN: 9780874218879. Available from: <http://ccdigitalpress.org/cad/CollaborativeApproaches.pdf>