**NeDiMAH Taxonomy Working Group Meeting**

**DANS, The Hague, the Netherlands**      http://www.dans.knaw.nl

**Scientific Summary**

The NeDiMAH Network researches the practice of, and evidence for, advanced ICT methods in the arts and humanities across Europe. The Network brings together practitioners in a series of thematic Working Groups, that examine the use of formal computationally-based methods for the capture, investigation, analysis, study, modelling, presentation, dissemination, publication and evaluation of arts and humanities materials for research. This research is contributing to the classification and expression of ICT methods used the arts and humanities in a key output: an enhanced ICT Methods Ontology, that will formalize and codify the expression of work in the digital arts and humanities, give greater academic credibility to this work, and enable peer-reviewed scholarship in this area. It will contribute to the development of a methodological layer that allows arts and humanities researchers to develop, refine and share research methods that allow them to create and make best use digital methods and collections. Better contextualization of ICT Methods will also be of particular benefit for early stage researchers.

In order to take forward the development of the ICT Methods Ontology, a group of experts was convened at this second Working Group Meeting. This is a follow up to the first, highly successful meeting of the group that took place in London on November 14th, 2012.
The conclusion of this first meeting was that the taxonomy/ontology should be taken forward in collaboration with DARIAH, which is also working on this activity through their Virtual Competency Centre 2 (research and education). As there is considerable overlap and synergy between the two organisations, this meeting was intended to scope existing work developed to date, and  begin to investigate the way in which NeDiMAH and DARIAH analysis of practice can be situated into ongoing international developments into the development of ontologies. This meeting was an important opportunity to take the DARIAH-NeDiMAH collaboration on the ontology of scholarly methods to the next level.  This meeting addressed the needs, goals, and resources of each of the three parties involved in ontology construction (NeDiMAH, DARIAH-DE, and DARIAH-EU), identifying excellent opportunities for pan-European cooperation on an essential task.

**Abstract**

This meeting was of a cross-team working group, with representation from NeDiMAH activities as well as some invited experts in the field, and representatives from DARIAH.

The meeting identified common aspects of digital arts and humanities research with a view to developing a taxonomy for classification of digital arts and humanities practice – a key NeDiMAH output.

**Meeting Programme: presentations**

**1. Introductions and aim of the event (Lorna Hughes, National Library of Wales)**

There was an overview of NeDiMAH, and a note that the membership of so many MOs in the Network reflects the current high profile of digital humanities across Europe. More and more funding has been made available for humanities research with digital outputs, and for

humanities research that relies on digital tools and methods. The six NeDimAH Working Groups were introduced, and it was noted that the overall aim of the Working Groups is to document current methods of scholarly practice and impact of digital humanities. There are common methods used across the disciplines, and an ICT methods taxonomy will be a way of communicating this work broadly. It will help identify comparative (and comparable) work across disciplines and effect parity across projects disciplines and regions.

Better identification of the relationship between *content, methods, and tools* will contribute to the transfer of knowledge and information across the international digital humanities community, provide evidence of use, help evaluation and peer review of DH projects, and contribute to making their approaches more visible. This model is a logcal extension of the idea of the "scholarly primitives" (Unsworth, 2000) and the ideas first articulated as a "methodological commons": http://www.cch.kcl.ac.uk/tmp/sci6/dighum-0807.pdf

Key aspects of an ICT methods ontology will be the role it can play in increased visibility of "digital transformations" in the arts and humanities, by highlighting collaboration and communication around tools, methods and content; and new paradigms for enabling research that would previously have been impossible, as well as facilitating existing research. This will add value to digital collections and identify the communities of practice that are involved.

The suggestion made at the previous meeting, that the ICT taxonomy work should be developed under the auspices of DARIAH and modified through input from the NeDiMAH WGs was explored in some detail through concrete examples of how this could work in practice, bringing in WG3 (linked data and ontologies), and eCloud, who could help develop a user case (DCU Greece) around the use of tools and methods. There have also been conversations about working with Cesar Gonzalez Perez (INCIPIT).

**2. Overview of NeDiMAH working Group activities: example from NeDiMAH WG 5 (Rene van Horik, DANS, Netherlands)**

This presentation described the NeDiMAH working group and survey on large-scale textual data, prepared in collaboration with Karina Van Dalen-Oskam, The aim of the WG is to demonstrate examples of how IT tools and methods, such as information retrieval and extraction methods (including, for example, text and data mining), can reveal new knowledge from large amounts of textual data, extracting hidden patterns by analyzing the results and summarizing them in a useful format. This working group is investigating practices in this area, building on the work of corpus linguistics and related disciplines to develop a greater understanding of how large-scale text collections can be used for research. On of the issues is defining "large" (or "large-scale") in this context, and convincing scholars of the need for large textual collections, so examples of practice related to content are very important. Some commonalities have been identified among scholars using "large" corpora:

- All have a need for smart tools for analysis
- Legality regarding use and re-use is always an issue for scholars
- There is a need for scholars to be convinced of the benefit of this work has to be convinced
- Have to explain the "black boxes of tools"
- There is a need to document and understand new kinds of collaboration.

**3. Introduction to ontology work at DCU: Costis Dallas, DCU, Greece**

Existing ontologies that can be used in the humanities, for example, CIDOC CRM are unclear in terms of their development plans or goals as they relate to methods in the digital humanities. "Events" is a central concept to CIDOC CRM. The DCU work on ontologies is based on an attempt to understand digital curation and its place in scholarship. They developed the idea from activity theory: researcher that is addressing a specific research question, finds an object, and tries to interpret and annotate it in order to improve it for scholarship. This work has been developed by, for example, Ellis (Sheffield), Tibbo (UNC) and Unsworth (UVA) tried to develop this work.

DARIAH is developing a representation of scholarly activity, based on an empirical attempt to understand researcher behaviours. This work summarizes observations of scholars, via questionnaires and other survey carried out, for example, by EHRI. From this work, information becomes generalizable, and the concept of the ontology extends beyond what can be accomplished using CRM. The work is based on close understanding of information use, based on scholars' diaries about how they use a book, or a library, at a certain level of detail.

This approach is borrowed from Activity theory, and the notion of activity as nested activities: for example, we talk about studying a manuscript, which is a sypraset of many smaller activities (granularity). Based on this methodology, DCU developed a schema of a knowledge base that starts with an "actor":, research goal, research activity, procedure, method, tools services, formats, concepts, info object, resource type, finally from that actor, a proposition that takes methods and tools as goals: a method, in the digital humanities, is used on content, and means one has to chose a specific tool in a system that can help scholars make informed decisions about what they want to do. At what point does this stop being a knowledge base and start being an ontology?

**Development of DARIAH VCC2 work on ontologies (Erik Champion, Christof Schöch, and**

**Matt Munsen).**

This presentation was a discussion of potential NeDiMAH-DARIAH collaboration. DARIAH has a pressing need for an ontology for its DARIAH.DE portal in German. It was noted that there is a need for associated bibliographies and supporting tools, if people are to use an ontology provided. It must sit within an established scholarly frame of reference. An experiment could be conducted using ZOTERO, using different tools on the same types of documents across languages to see similarities and dissimilarities, working with a common set of reference corpora. There is a need to compare tools across languages to benchmark

**Discussion and next steps**

The key question is: Do we start from what we have already (e.g., arts-humanities.net, Digital Humanities at Oxford, DRAPIER), or do we develop something new? If so, one goal of DARIAH could be to sustain the database, we could have our researchers of experimental groups to use this taxonomy and use this as a collaborative thing (if web or database)…so called collaborative database, to experiment it and then maybe to develop it further. t was agreed that the DARIAH-DE portal group should keep in close contact and collaborate with this NeDiMAH WG.

The first step is to boil down arts-humanities.net into some *Urmethoden* - Could Erik do this in his comparison of the existing taxonomies, and produce a form of "taxonomy distillation". It was suggested that we should be adding to Tools Content Methods, an *arena* (where it is debated,

stored, where or whom it influences), and any exemplars because methodology means also *study and debate of* method. This debate is part of communities of practice BUT not yet part of content-method-tools ontology. DARIAH courses/researchers/students can add in this community of practice as and when time permits (see photos).

**Suggested workplan**

In order to move beyond the data gathering stage, it is necessary to develop a workplan. DARIAH VCC2 will be a mechanism to:

- Develop a research methodology and manual for conducting comparative research on current and emergent scholarly information practices and needs
- Perform baselines say 4-5 representative countries across Europe structured phone interviews and online questionnaire survey
- Perform a trans-European literature survey and create an online bibliography of scholarly information practices and needs in the arts and humanities
- Create and maintain an up to date knowledge base on scholarly practices and needs as well as an up to date registry of tools and methods etc.

Erik noted that Denmark can offer, collaborate on, or provide in-house expertise, including

- Contribution to user surveys
- PhD summer courses.
- At Sept conference could have NeDiMAH workshop in Best Practices and Paradigms in Digital Humanities at event in Denmark
- Denmark are building a website so can also like dariah.de start from arts-humanities.net
- Could write a paper on arts-humanities.net and oxford dh taxonomies for TPDL.
- Can also supervise PhD or postdoc projects.

The group should meet about 2 times a year, a formal working group that will contribute to the development of the NeDiMAH ontology.


**List of speakers/participants**


Lorna Hughes, National Library of Wales, UK

Erik Champion University of Arhus, Denmark

Christof Schöch, University of Wuerzburg, Germany

Costis Dallas, University of Toronto

Luise Borek DARIAH-DE

Matt Munsen, University of Göttingen, Germany

Rene Van Horik, DANS, Netherlands

Gabrielle Kfaft, DARIAH.DE

## Top whiteboard

METHODOLOGICAL COMMONS

Method
— place
— time

Tool

methodology
- collection of methods
- methods (potentials)
- study /debate of methods

in closing
· folksonomy = user tags
ontology also strictly defines
relationships between objects.
(IT often uses both).
ontology allows multiple perspectives/entries

contribution
1 — 30–60.000 DKK
user survey
2 — PhD summer course
funding
3 — host NeDiMAH
panel or workshop
at SEPTEMBER conf.
"BEST PRACTICES; PARADIGMS
IN D.H." → ROYAL LIBRARY
COPENHAGEN
← [WE ALSO ARE BUILDING
A WEBSITE.]
4 WRITE PAPER COMPARING
CURRENT ISSUES IN
DH TAXONOMIES (Digicon:)
(TPDL, MALTA 2013).

## Bottom whiteboard

Users — scholars

ontopedia — wouldbescholars

METHO
C

Content

Method
— place
— time

arch
impact    librarian

ARENA
Stored
Debated
Influence
exemplars

Tool

in c