NeDiMAH Working Group Meeting: Methods Taxonomies for the Digital Arts and Humanities

Wilkins Garden Room, Bernard Katz Building, University College, London
November 14th, 2012

Background

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 taxonomy of methods for the digital arts and humanities (ideally expressed as an ontology) that will formalize and codify the expression of work in the digital arts and humanities, give greater academic credibility to this work, and support 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 of digital collections. Better contextualization of ICT Methods will also be of particular benefit for early stage researchers.

This meeting will scope existing work in the field, including taxonomy and classification work developed by libraries and digital humanities centres and organisations to date. They will begin to investigate ways in which NeDiMAH can carry out and analysis of practice from which a classification of digital arts and humanities can emerge, and how this can be situated into ongoing international developments into the development of ontologies. It will also build a connection with DARIAH Virtual Competency Centre 2, which has a remit to investigate the classification of digital humanities practice.

Aims and Objectives

This event will begin the formation of a cross-team working group, with representation from NeDiMAH activities and invited experts in the field, to scope existing work in the field and identify areas of new research to be carried out through NeDiMAH. The event will discuss common aspects of digital arts and humanities research with a view to developing a taxonomy/ontology for classification of digital arts and humanities. Outputs of the meeting will include a report, and a workplan for future activities.

Programme

09:30   Coffee and arrival
10:00   Welcome
10:10   Lorna Hughes: 'Overview of NeDiMAH: Aims, objectives and outputs'
10:30   Torsetn Reimer: 'Developing ICT methods taxonomies within the digital humanities: work to date'
11:00   Coffee
11:20   Anne Welsh: 'Hard Drugs and Soft Terms: Lessons learned from a European Taxonomy design project '
11:50   Ruth Reiche and Christof Schöch, 'DARIAH and methods classification in the digital humanities'
12:20   Discussion
13:00   Lunch

13:45   NeDiMAH Working groups and methods identified: discussion
14:30   Discussion and next steps: a workplan for future activities
16:00 Close

Participants

Lorna Hughes, National Library of Wales, (UK)
Anne Welsh, University College, London (UK)
Torsten Reimer, JISC (UK)
Christoph Schöch, University of Wuerzburg (Germany)
Ruth Reich, University of Darmstadt (Germany)
Matt Munsen, University of Gottingen (Germany)
Christian Emil Ore; University of Oslo (Norway)
Leif Isaksen, University of Southampton (UK)
Orla Murphy, University College, Cork (Ireland)
Tobias Blanke, King's College, London (UK)

Presentations and discussion

Torsten Reimer presented a review of arts-humanities.net (www.arts-humanities.net), and described the development of the methods taxonomy that underpins this resource. This was originally developed as a taxonomy of ICT methods for digitization by the UK's Arts and Humanities Data Service (AHDS) to describe digitization methods. This was developed further by the AHRC ICT Methods Network as a taxonomy of digital method for digital humanities, and became the underlying taxonomy for arts-humanities.net. The taxonomy was subsequently refined in collaboration between the Methods Network, the Digital Humanities Observatory in Dublin, and Oxford University e-Research Centre. An updated version of the taxonomy is now the basis for mapping digital humanities projects at Oxford (http://digital.humanities.ox.ac.uk/).

In the discussion of the taxonomy, attention was drawn to the tagging system developed by Torsten for arts-humanities.net, which could be developed into a folksonomy approach to tag projects according to their descriptions by users. A combination of community generated content and a closely edited model was felt to be a useful approach, and one that would address the question of whether a taxonomy for the digital humanities should e descriptive or prescriptive, and the importance of a useful ontology being 'self knowing'. This is at the core of what needs to be achieved by a functional ontology for the digital humanities: it must be firmly based on the documentation practice: the development of the taxonomy by the Methods Network, Oxford and the DHO was possible because the resource was based around the description of a critical mass of projects with digital outputs funded by the AHRC in the UK.

Anne Welsh introduced the difference between a taxonomy – with a clear hierarchical structure, and an ontology, which has a wider reach and is associated with meaning and, ideally, future longevity. Working with "softer' terms is harder than working with hard scientific terminology, which tends to have been pre-defined. Bibliometrics works much better in the sciences than the humanities. It is in the use of soft terms that the development of a taxonomy becomes an art, not a science. Understanding how to dissolve the boundaries in descriptions is part of this art – the example of drug classification was used, explaining how re-classification of drugs causes taxonomies of drug classification to collapse; similarly, terms like "prevention" have different meanings in different jurisdictions in the context of drug use. The use of the **Simple Knowledge Organization System** (**SKOS**) allows users to do more intuitive searching. This is of particular interest in Digital Humanities, an interdiscipline, where the use of terminologies accurately can break this down into things that people will understand, working around mind maps. One way to focus on this is to scope and identify the core concepts that are central to the digital humanities, and are consistent across disciplines and cultures. This would also work within the

folksonomy approach, where core concepts can be 'nodes' around which words can be clustered.

Ruth Reiche presented the blueprint for the DARIAH.DE portal, providing the 'analysis' component of DARIAH.EU, or research, analysis, and interpretation. This raised the interesting discussion of how we define "digital humanities', and the scope of the science/research/wissenschaft that it comprises, and therefore the question of how NEDIMAH could create elements that might become the 'methodological layer' of DARIAH. NeDiMAH is in many ways a meta aspect of many things in DARIAH, and outcomes of NeDiMAH could create research that DARIAH could build on.

This raised the question of how we define a 'digital humanities method'? Is it a method that is used by more than one discipline? For example, methods like ground penetrating radar, or remote sensing with LIDAR are used in archaeology, and not in other disciplines – does this make them "digital humanities" methods? There are relationships across methods – but are they paternal or fraternal? This problem indicates that what is required is an ontology, rather than a taxonomy. We could represent these disciplinary and methodological 'kinships' as a network map.

Christoph Schöch discussed his work in developing a bibliography of terms in Zotero, with over 2000 entries, which he has based around the "Scholarly Primitives" described by John Unsworth and built from this. These are based around 'activities'. It would be useful to integrate this bibliography with the NeDiMAH taxonomy activities. The collating of the bibliography is a cross-group initiative, working across disciplinary groups.

Leif Isaksen and **Orla Murphy** presented the work of NeDIMAH working groups 1 and 2, outlining what they have done, their outputs, and how these outputs could feed into the development of taxonomy related activities. Their conclusions were that what is most needed is for work at a more practical than theoretical level, and that what would be most constructive would be funding activities aimed at populating a taxonomy as a practical outcome, rather than concentrating on theoretical abstracts. They also presented a useful survey developed at the recent WG1 and 2 working group meeting in Romania, to create a body of data about use of digital methods, to uncover research practice in a more empirical fashion. It would be useful to think about how this data can be analysed and represented, and feed into the NeDiMAH database and taxonomy.

Christian-Emil Ore discussed the work of the NeDiMAH Ontology Working Group, which has been engaging with the Open Annotation community, and creating recommendations about developments in the field, focusing on taking the 'semantic web' beyond the phase of hype into a scalable development, where linked data approaches can be effectively communicated across the disciplines. They key is in comparing ontologies as conceptual models and not data models, looking to the work of initiatives like CDOC-CRM. The success of this initiative shows the importance of including archives, libraries, and museums in this research. This presentation again highlighted the fact that while a great deal has been invested in digitization, there has been very little focus on using digital content for research, and there is an important opportunity to address this gap in research.

Tobias Blanke discussed the work being carried out in methods engineering by the research lab of Cesar Gonzalez-Perez in Spain, which overlaps with attempts to formalize methods and is described in a recent paper for CAA. Also relevant is the DARVISH project to build tools for the social sciences and humanities.

Conclusions

Discussion at the event raised the question of how we define a 'digital humanities method'? Is it a method that is used by more than one discipline? For example, methods like ground penetrating radar, or remote sensing with LIDAR are used in archaeology, and not in other

disciplines – does this make them "digital humanities" methods? There are relationships across methods – but are they paternal or fraternal? This problem indicates that what is required is an ontology, rather than a taxonomy. We could represent these disciplinary and methodological 'kinships' as a network map.

There is, however, a need for a communal space to develop this type of activities, so that they may feed into the foundational aspects of NeDiMAH, and work must be progressed to find the best ways to take this work forward. We should explore potential partnerships (for example with DARIAH, and with Cesar Gonzalez-Perez) and develop a workplan to develop a skeleton ontology and flesh it out.

There must be a clearly defined need and purpose for the taxonomy, whether it is sharing expertise, communicating information about projects, or allowing critical assessment and peer review of work in the field: there could be a role for this resource as a tool for research councils in review and classification of DH research, or in classifying publications in the field, or in supporting younger researchers in developing work that can be identified as part of the wider DH field. Identifying the objective of the resource will incentivize its development as a vital component of the digital humanities infrastructure. A taxonomy must be based on "practice", and embedded into the documentation of digital humanities in practice. The NeDiMAH working groups are an important opportunity to gather data about digital humanities methods in practice. The current focus of the existing taxonomy hosted by Oxford is based around content, methods, and tools as the digital humanities "workspace" – are these still valid as core concepts?

With this in mind, we need to develop a coherent workplan for how the taxonomy can be taken forward, implemented, and sustained. This sort of work will evolve over time, so it needs to seen as an ongoing initiative.
Any output developed by NeDiMAH can be embedded in DARIAH for sustainability, for curating and sustaining the resource as an "institution" or resource over the long term. It must have a home that will sustain and develop it for it to be relevant, as digital humanities tools and methods are changing all the time. A successful example of how a useful piece of the DH infrastructure can be sustained in this way over the long term is the TEI, where Special Interest Groups work together to keep the resource current. Could DARIAH be the framework to sustain the taxonomy in this way, by facilitating and supporting community building activities through national DARIAH initiatives as well as DARIAH.eu.

An important consideration is striking a balance between depth and breadth of coverage. The most cutting edge in the DH is still contentious, so there is a need to strike a balance between having a small group not prescriptively describing activity versus casting net very wide and engaging with those with a wide range of experience. If adopting folksonomy methods, what is the tipping point – or traction – between user engagement as a means for identifying activities and trends, and developing an authoritative resource? Are there methods that have emerged from the use of folksonomies in earlier projects, like arts-humanities.net that can be useful, like normalization of data gathered, serendipitous reading, and multilingual approaches? We should also explore how the back end for this could be open linked data.

Next steps

The group agreed that Matt would work on integration of the Oxford ontology with work that has been carried out by his group through DARIAH (results attached to this report), and that the group should meet again in approx. 6 months to review this work in more detail.

We agreed that we should identify the key user of the taxonomy: for example, scholars, funders, students, by thinking about a couple of topics and identifying their core user groups. Examples could be: "social network analysis"; "GIS"; "Topic Modelling" and "Text Mining". Working through a concrete example of a distinct method would enable greater

research into the community of practice that uses the method, and understand how a taxonomy would benefit them. It could also tease out the issue of nodes and outputs versus the hierarchical approach. This can be developed through a series of joint DARIAH and NeDiMAH working group meetings, specifically convening a meeting with DARIAH Vcc2 representatives as soon as possible.

Developing the bibliography will also be key to all this, as a source of references and people should be directed to the biography through NeDiMAH. Working group outputs need to feed into this in a more formal manner.

The group agreed to set up a mailing list, and to discuss this at future DARIAH meetings and community engagement workshops.

We also need to develop a practical workplan that will enable the group to gather data on DH practice to abstract information that will populate a taxonomy. The survey developed by NeDiMAH WG 1 and 2 may be useful in this respect. Interviews could also be useful way of gathering data, possibly in collaboration with DARIAH. They key thing will be to ensure disambiguation – especially with regard to making a distinction between methods of creating digital humanities projects, as opposed to using their outputs via the use of DH tools and methods. It would also be interesting to carry out some research on how the development of a folksonomy might produce a 'view' of how the DH community sees itself this time. There are interesting opportunities to capture the tags for the annual "day of digital humanities", as has been carried out by Julianne Nyhan at UCL, and the emerging taxonomy used for the DH conference reviewing system. We also need a more formal process for capturing the outputs of the NeDiMAH working groups, and ensuring that the working groups collaborate on the 'foundational' aspects of NeDiMAH, like the taxonomy. We need a clearer template for developing and generating deliverables that will feed into this work.

The group needs to identify other participants – form DARIAH and beyond to take this forward in three key phases: data gathering; the integration of this data to populate an ontology; and publication of the outputs as Linked Open Data/RDF. This should be planned at a future meeting of the group early in the new year. With participation from DARIAH, especially DARIAH VCC2 and in consultation with the DARIAH leadership.