

Digital Humanities research on the Prior Nachlass

Henriette Roued-Cunliffe & Volkmar Engerer

Aalborg, 6th October, 2015

General remarks

We propose to supplement the project's endeavors to promote Prior's logic and philosophy with a digital humanities and information science perspective, directed at supporting the development of a digital presentation of the Nachlass to the Prior community and, as a long-term perspective, to a broader academically interested audience.

The heart of the present Prior Virtual Lab is a database containing pictures of Prior's hand-written manuscripts and correspondence waiting to be transcribed. These transcriptions, once made, are digitized representations of analog "originals" and can as such in a second step be organized to be searched, used and reused by the research community. Thus, there are, roughly speaking, two domains of organization where the Prior research community is engaged in the Nachlass,

- a) in transforming analog manuscripts into digitized texts, and
- b) in (re)using the transcribed documents in future research contexts.

While activities in the a)-domain are restricted in the sense that only registered Prior researchers have access to the database of pictures and transcriptions ("closed community"), the b)-domain will be an open site where all interested users can access the digitized texts.

These two areas of involvement are characterized by different environmental and legal settings (closed vs open community), diverging motivation and objectives of engagement (academic research vs more open and diffuse goals like information seeking, learning, etc.), and, last but not least, by different competence and educational profiles of potential users (logician-specialist vs non-specialist). These two spheres with their diverging characteristics pose interesting research questions and development possibilities seen from a digital humanities and information science perspective.

Transcription (a) - closed domain

Transcribing images of hand-written manuscripts into digital text is a complex task, demanding professional expertise of the topic, knowledge of the historical and cultural context in which the manuscripts came into being, and, last but not least, a profound motivation for the time and energy invested in this activity. Information science views the Prior Virtual Lab as a Closed Collaborative Community - CCC, where researchers in a closed platform engage in making Prior's unpublished material accessible for research. In order to facilitate this transformation process, the following questions will be addressed:

- How can digital tools support the editors' needs for effective transcription and communication with other editors?
- How can a virtual platform effectively offer tools that can aid editors attempting to solve interpretational dilemmas (e.g. online encyclopedias)?

- How can digital tools facilitate real-time knowledge sharing in the transcription process?
- What is the ultimate motivation for editors involved in this time-consuming activity and how can this be incorporated into a sustainable digital solution?
- And, quite basically, why do many scholars resort to working offline with a text-editing program, such as MS Word?

Representation (b) - open domain

Already in the digitizing and transcribing phase, the question can be asked how the resulting digital documents best are described and represented in order to become useful items in prospective scientific research. Modern research in indexing reliability and accurateness of indexing data suggests that representational items (subject descriptors, name indices, etc) best are generated and assigned by subject specialists - and not information specialists. Subject indexing is therefore, ideally, regarded as a part of the editing task itself.

More generally, the question of digital text representation occurs on two levels, the first on the level of text, i.e. the document itself (“In-text representation”), and the second on the level of the more traditional representation of documents, often called metadata or “bibliographic records”.

Traditional methods like specialist-driven subject indexing via subject descriptors or classification schemes is, in the case of the Prior data, not sufficient. Both academic correspondence as a chain of connected communications on academic topics on the one hand and the unpublished manuscript themselves on the other (standard definitions of knowledge organization rely on published information) constitute quite untypical objects for mainstream Library and Information Science.

An adequate representation of the items in the Prior Nachlass, both in-text and through metadata, must not only take in account the specifics of the data described, but also be flexible enough for unforeseeable research questions that may arise in future research. Decisions on indexing strategies of the Prior documents must therefore be based also on these theoretic and knowledge-epistemological considerations.

The reuse of textual material such as the Prior Nachlass is typically conducted through textual encoding with TEI (Text Encoding Initiative) formatted XML. This enables the option of publication through different channels, namely as open data through a web service API.

This poses another set of questions:

- Is the difference between making the transcribed texts and metadata available as real-time raw encoded data and publishing it on a static website clear to the scholars involved?
- How does the TEI format facilitate representations that meet the special requirements of (chains of) letters and unpublished texts as well as visualisation of linked data?
- How do other projects tackle the responsibility of encoding between scholars, information scientists and IT specialists?
- How are different stages of revision of Prior’s manuscripts linked to the “final”, published version?

- How have the contents of Prior's professional correspondence like controversies with colleagues about manuscripts to be represented in order make explicit their impact on different stages in manuscript revision?
- How can modern technologies of visualization be utilised to contextualize Prior's professional development in terms of, for example, historical events or other cultural impacts?

Our Contribution (tentative)

In order to answer these questions, we propose to:

- I. First conduct a review of literature and projects attempting to solve similar issues as those mentioned above. From an digital humanities point-of-view the technical issues faced by the Prior group are very similar to those faced by scholars editing other handwritten texts (e.g. Latin tablets or Buddhist manuscripts). Therefore, we will conduct this review to include a very wide range of different text editing projects and scholars.
- II. Secondly, we propose interviews with Prior researchers on how they experience and have experienced the ongoing technological development and how they incorporate information technology into their professional work with hand-written texts.
- III. Thirdly, we attempt to extract from our theoretical review (I) and empirical results from our interviews (II) critical developing needs for the Prior Virtual Lab and the growing open website, and propose practical improvements to be implemented in future.