Corpora Ethnographica Online:
Strategien der Digitalisierung kultureller Archive und ihrer
Präsentation im Internet

Rostock, 26–28 September 2012

From 26 to 28 September 2012, the University of Rostock held an international scientific symposium entitled Corpora Ethnographica Online. The purpose of the symposium was further explained by its subtitle, namely developing strategies for the digitization of cultural archives and their presentation on the Internet. The initiator of this conference was Christoph Schmitt, director of the Wossidlo archive in Rostock (city in the federal state of Mecklenburg-Vorpommern). The Wossidlo archive is an ethnographic and linguistic archive that will be fully digitized in the coming years, in the so-called Wossidlo project (DNA = digital archive), thanks to a substantial grant. The digitization project will be executed by the folklore department (Institut für Volkskunde) in cooperation with the university library and the computer science department of the University of Rostock. Representatives of the two last-mentioned departments who were speaking at the symposium were, among others, library director Robert Zepfand information scientists Holger Meyer and Alf-Christian Schering. The conference was attended by representatives from the humanities of several national and international universities, institutes, museums, archives and libraries (for which reason the official languages of the symposium were German and English).

The entire symposium was interlaced by a number of meaningful connecting threads, which will in succession, play a role of increasing importance in the years to come: the digitization of heritage, the importance of open access, the need for data enrichment, the value of crowdsourcing, the use of portals, the advancement of interconnectivity by means of harvesters and the need for large-scale standardization. Furthermore, international cooperation will have to be sought on an ever-growing scale. And last but not least, the scientific approach should not be neglected in the overall process.

The central topic of the symposium was the digitization of (both material and immaterial) heritage. Barbara Sosíl of the Slovene Ethnographic Museum, for instance, talked about the digitization of the folklore collection of Istra Orli (1903–62) and his team. Another example was Risto Järvi's presentation on the digitization of the Estonian Folklore Archives in Tartu. Sutti Kivell explained how the Finnish Sámi Archives use digital techniques for the disclosure and the preservation of the language and culture of the Sámi people. On a side note, what applies to the digitization of the Sámi collection, applies to many more collections in the humanities: the digitization process has only just begun and only a small
Humboldt (1769–1859) kept up correspondence with distinguished scholars like Georg Forster and Charles Darwin. At the time, there was a network of corresponding scholars, who sent and received letters that have ended up separated from each other, in numerous archives and libraries. Due to the digitization of all this correspondence, the Kalliope project will be able to bring together letters that respond to one another.

During the evening lecture at Rostock City Hall, the Australian researchers Philip Batty (Museums Victoria, Melbourne) and Jason Gibson (Australian National University, Canberra) gave an impressive example of how digitization is capable of regathering (online) Aboriginal heritage from all over the world. The project consists in an elaborate reconstruction of the collection of the Australian anthropologists Walter Baldwin Spencer (1860–1920) and Frances James Gillen (1855–1912), who did extensive fieldwork among the Aboriginal population of Australia between 1875 and 1903. Letters, (written) recordings, photographs, wax cylinders with sound recordings and artefacts are scattered all over the world, in museums, archives and libraries. Their anthropological work was a source of inspiration for scholars like Emile Durkheim and Sigmund Freud. The online regathering of this enormous ethnographic collection must be regarded not only as a virtual return of Aboriginal heritage, but also as a source of great scientific importance, for in its present coherent state, the collection allows an overview that was lost.

The digitization of folklore material could take advantage of the ever-increasing interactivity of the Internet. In other words, Web 2.0 should be exploited.

Ethnology could, for instance, benefit from a phenomenon like crowdsourcing: methods such as using selected, knowledgeable volunteers to transcribe archival materials and suchlike are already being applied successfully, as we speak. In addition, there are the so-called portals that allow large digital cultural collections to be brought together. One such portal, which was frequently mentioned at the symposium, is europeana (www.europeana.eu). The so-called harvesters take it one step further: these smart search engines are capable of performing searches in several databases on an international scale and of presenting results from several regions, countries, languages and cultures. This form of interconnectivity could be of great value to the folklore databases that are presently being constructed. It would be quite a scientific innovation if folklores could be retrieved simultaneously from several digital databases, like the Dutch Folklore Database, the Folklore database of the Woolsto archive, the Estonian folklore archive and the Georgian database of Eliaji Dadamashvili. Other candidates (not represented at the symposium) could join this initiative, such as the Dutch Folklore Database, the Archive of Portuguese Legends and BondCat: Catalan Folklore Search Engine. In order to be able to pass a harvester over all the participants, the databases must meet some standard requirements, like those set forth by the Dublin Core Metadata Initiative (DCMI), the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) or CLARIN. Several fields in the databases must be standardized and equalized. For images, there already exists an international classification standard.
like iconclass (which is primarily suitable for paintings, though). For text files, we will need standardization for the keyword system and some other classification categories. In addition, we may ask ourselves the question if our folktales will be sufficiently served by the ATU system and the Motif-Index alone. Ideally, an international group of folkloric researchers would take up an initiative like the "Type-Index of Legends" (a project that was suggested in the past, but never really got off the ground).

In conclusion, we ought to realize one thing. When we are dealing with a digitization operation of considerable proportions, at some point policymakers, politicians and taxpayers will bring up a vital question: before long, all your material will be digitized - how will what will happen next? Or, to quote Christiane Contissau and Erja Nunez-Charásia: "Da eine nutzen die schönsten Informationsysteme, die besten Rechtsberater und die geschäftstüchtigsten Quellenkritiker, wenn es keine inhaltlichen Fragen gibt, die an die Quellen benannt werden?" It will be up to the scientists to confront the source material with research questions, be it old questions that could be asked again, or entirely new ones. In my own presentation, I suggested some possible research approaches concerning folktales databases. The comparative research into variants in time and space has always been taken a central position in folkloric research. This research angle will soon enjoy the benefits of both a superabundance of available source materials and a more sophisticated set of techniques to detect variants and visualize their interconnections. We will shortly be able to record the specific mechanisms of the dynamics and variations in oral (and written) tradition, with much more precision than ever before. It goes without saying that the scientists themselves will remain responsible for the interpretation of the variations as far as their function and meaning are concerned. The researcher into historical and geographical distribution we have long been familiar with may soon enter into a promising liaison with Google Maps or Google Earth.

Thanks to the advantages of enrichment, stories, places, objects, persons and historical developments can get interconnected in meaningful ways. And now that we will soon have access to an incredible number of digitized stories, structuralist research can venture a renewed claim on our attention. Before long, it will be possible to take on the analysis of motif sequences as narrative building blocks of fairy tales, legends, jokes and contemporary legends with much larger corpora than Vladimir Propp could ever have imagined in his wildest dreams. By using large folkloric corpora, we could research the possible existence of a so-called narrative grammar, perhaps even a universal narrative grammar. It is up to the scientists to make creative use of the possibilities offered by the new digital collections and to continue discussions and collaborations with the specialists in the field of information technology.

In my perception, Corpora Ethnographica Online was a highly successful symposium, well worth repeating. After all, there is every indication that the twenty-first century will be one of large-scale digitization and also of computational research in the humanities; for there are signs that in this century, scientific research, too, will develop (at least in part) in the direction of the eHumanities.