

## **Dissertations from Uppsala University 1602-1855 on the Internet**

*Dr. Anna Fredriksson, Uppsala University Library*

At Uppsala University Library, a long-term project is under way which aims at making the dissertations, that is theses, submitted at Uppsala University in 1602-1855 easy to find and read on the Internet. The work includes metadata production, scanning and OCR processing as well as publication of images of the dissertations in full-text searchable pdf files. So far, approximately 3,000 dissertations have been digitized and made accessible on the Internet via the [DiVA](#) portal, Uppsala University's repository for research publications. All in all, there are about 12,000 dissertations of about 20 pages each on average to be scanned. This work is done by hand, due to the age of the material. The project aims to be completed in 2020.

### **Why did we prioritize dissertations?**

Even before the project started, dissertations were valued research material, and the physical dissertations were frequently on loan. Their popularity was primarily due to the fact that generally, studying university dissertations is a great way to study evolvments and changes in society. In the same way as doctoral theses do today, the older dissertations reflect what was going on in the country, at the University, and in the intellectual Western world on the whole at a certain period of time. The great mass of them makes them especially suitable for comparative and longitudinal studies, and provides excellent chances for scholars to find material little used or not used at all in previous research.

Swedish older dissertations including those of today's Finland specifically are also comparatively easy to find. In contrast to many other European libraries with an even longer history, collectors published bibliographies of Swedish dissertations as far back as 250 years ago. Our dissertations are also organized, bound and physically easily accessible. Last year the cataloguing of the Uppsala dissertations was completed according to modern standards in LIBRIS. That made them searchable according to subject and word in title, which was not possible before. All this made the digitization process smoother than that of many other kinds of cultural heritage material. The digital publication of the dissertations naturally made access to them even easier for University staff and students as well as lifelong learners in Sweden and abroad.

### **How are the dissertations used today?**

In actual research today, we see that the material is frequently consulted in all fields of history. Dissertations provide scholars in the fields of history of ideas and history of science with insight into the status of a certain subject matter in Sweden in various periods of time, often in relation to the contemporary discussion on the European continent. The same goes for studies in history of literature and history of religion. Many of the dissertations examine subjects that remain part of the public debate today, and are therefore of interest for scholars in the political and social sciences. The languages of the dissertations are studied by scholars

of Semitic, Classical and Scandinavian languages, and the dissertations often contain the very first editions and translations of certain ancient manuscripts in Arabic and Runic script. There is also a social dimension of the dissertations worthy of attention, as dedications and gratulatory poems in the dissertations mirror social networks in the educated stratum of Sweden in various periods of time. Illustrations in the dissertations were often made by local artists or the students themselves, and the great mass of gratulatory poems mirrors the less well-known side of poetry in early modern Sweden.

### **Our users**

The users of the physical items are primarily university scholars, primarily our own University, but there is also quite a great deal of interest from abroad. Not least from our neighboring country Finland and from the Baltic States, which were for some time within the Swedish realm. Many projects are going on right now which include our dissertations as research material or which have them as their primary source material; Swedish projects as well as international. As Sweden as a part of learned Europe more or less shared the values, objects and methods of the Western academic world as a whole, to study Swedish science and scholarship is to study an important part of Western science and scholarship.

As for who uses our digital dissertations, we in fact do not know. The great majority of the dissertations are written in Latin, as in all countries of Europe and North America, Latin was the vehicle for academic discussion in the early modern age. In the first half of the 19<sup>th</sup> century, Swedish became more common in the Uppsala dissertations. Among the ones digitized and published so far, a great deal are in Swedish. As for the Latin ones, they too are clearly much used. Although knowledge of Latin is quite unusual in Sweden, foreign scholars in the various fields of history often had Latin as part of their curriculum. Obviously, our users know at least enough Latin to recognize if a passage treats the topic of their interest. They can also identify which documents are important to them and extract the most important information from it. If the document is central, it is possible to hire a translator.

But we believe that we also reach out to the lifelong learners, or the so-called “ordinary people”. The older dissertations examine every conceivable subject and they offer pleasant reading even for non-specialists, or people who use the Internet for genealogical research. The full text publication makes the dissertation show up, perhaps unexpectedly, when a person is looking for a certain topic or a certain word. Whoever the users the digital publication of the dissertations has been well received, and far beyond expectations. The first three test years of approximately 2,500 digitized dissertations published resulted in close to one million visits and over 170,000 downloads, i.e. over 4,700 per month. Even if we don’t – or perhaps *because* we don’t – either offer or demand advanced technologies for the use of these dissertations.

### **The digital publication and the new possibilities for research**

The database in which the dissertations are stored and presented is the same database in which researchers, scholars and students of Uppsala University, and other Swedish universities, too,

currently register their publications with the option to publish them digitally. This clears a path for new possibilities for researchers to become aware of and study the texts. Most importantly, it enables users to find documents in their field, spanning a period of 400 years in one search session. A great deal of the medical terms of diseases and body parts, chemical designations, and, of course, juridical and botanical terms are Latin and the same as were used 400 years ago, and can thus be used for localizing text passages on these topics. But the form of the text can be studied, too. Linguists would find it useful to make quantitative studies of the use of certain words or expressions, or just to find the words of interest for further studies. The usefulness of full-text databases are all known to us. But often one as a user gets either a well-working search system *or* a great mass of important texts, and seldom both. This problem is solved here by the interconnection between the publication database DiVA and the Swedish National Research Library System LIBRIS. The combination makes it possible to use an advanced search system with high functionality, thus reducing the Internet problem of too many irrelevant hits. It gives direct access to the digital full text in DiVA, and the option to order the physical book if the scholar needs to see the original at our library. Not least important, there is qualified staff appointed to care for the system's long-term maintenance and updates, as part of their everyday tasks at the University Library. Also, the library is open for discussion with users.

### **The practical work within the project and related issues**

As part of the digitization project, the images of the text pages are OCR-processed in order to create searchable full-text pdf files. The OCR process gives various results depending on the age and the language of the text. The OCR processing of dissertations in Swedish and Latin from ca. 1800 onwards results in OCR texts with a high degree of accuracy, that is, between 80 and 90 per cent, whereas older dissertations in Latin and in languages written in other alphabets will contain more inaccuracies. On this point we are not satisfied. Almost perfect results when it comes to the OCR-read text, or proof-reading, is a basic requirement for the full use and potential of this material. However, in this respect, we are dependent upon the technology which is available on the market, as this provides the best and safest product. These products were not developed for handling printing types of various sorts and sizes from the 17<sup>th</sup> and 18<sup>th</sup> centuries, and the development of these techniques, except when it comes to “Fraktur”, is slow or non-existing.

If you want to pursue further studies of the documents, you can download the documents for free to your own computer. There are free programs on the Internet that help you merge several documents of your choice into one document, in order for you to be able to search through a certain mass of text. If you are searching for something very particular, you could of course also perform a word search in Google. One of our wishes for the future is to make it possible for our users to search in several documents of their specific choice at one time, without them having to download the documents to their computer.

So, most important for us today within the dissertation project:

- 1) Better OCR for older texts
- 2) Easier ways to search in a large text mass of your own choice.

### **Future use and collaboration with scholars and researchers**

The development of digital techniques for the further use of these texts is a future desideratum. We therefore aim to increase our collaboration with researchers who want to explore new methods to make more out of the texts. However, we always have to take into account the special demands from society when it comes to the work we, as an institute of the state, are conducting – in contrast to the work conducted by e.g. Google Books or research projects with temporary funding.

We are expected to produce both images and metadata of a reasonably high quality – a product that the University can ‘stand for’. What we produce should have a lasting value – and ideally be possible to use for centuries to come.

What we produce should be compatible with other existing retrieval systems and library systems. Important, in my opinion, is reliability and citability. A great problem with research on digitally borne material is, in my opinion, that it constantly changes, with respect to both their contents and where to find them. This puts the fundamental principle of modern science, the possibility to control results, out of the running. This is a challenge for Digital Humanities which, with the current pace of development, surely will be solved in the near future.