Background

The aim of the project "Greek Manuscripts in Sweden" is to publish an internet-based catalogue of all the Greek manuscripts in Swedish libraries and archives. This will also entail a full digitization of the manuscripts, some 36 000 pages in total. The project is hosted by Uppsala University Library and is funded by The Central Bank of Sweden Tercentenary Foundation. Libraries that will contribute material to the database include Gothenburg University Library (4 MSS), Linköping Diocesan Library (28), Lund University Library (5), The National Gallery, Stockholm (1), The Royal Library, Stockholm (5), The Skokloster Castle Library (3), and Uppsala University Library (74), in all 120 manuscripts.

The Greek manuscripts, in the form of bound parchment and paper volumes, include a rich and diverse collection of texts from antiquity and the Byzantine period. They originate mainly from the Byzantine cultural area from the tenth century onwards, but some are Renaissance or early modern manuscripts from Western Europe. The existing nineteenth-century catalogue is outdated and in many cases incorrect. Furthermore it does not cover all manuscripts. It is therefore essential to create a new catalogue according to modern principles, including detailed codicological descriptions. Combined with a comprehensive digitization of the manuscripts this will facilitate and encourage new research on the material among Swedish and international scholars.

The decision to use the TEI manuscript description module for the catalogue was taken on the grounds that it affords an XML encoding system which is open source, well supported, and already implemented in other web-based catalogues. However, since manuscript research and codicology is a relatively young and sprouting field of research, we consider it necessary to modify and develop some aspects of this module further. One crucial issue is to pay more attention to the codicological units from which most manuscripts are composed. A manuscript may have been written and composed, put together, and perhaps taken apart, in several stages and processes, and the description needs to reflect this historical stratigraphy of production and use. A recent publication emphasizing this is La syntaxe du codex (Andrist et al. 2013). This approach was employed for the catalogue of Greek manuscripts in the Burgerbibliothek in Bern (Andrist 2007), but that catalogue started out as a printed catalogue and only a few items have been rendered into TEI encoding afterwards. The encoding mainly follows the setup of the printed catalogue.

Implementation

In the present project the manuscript descriptions are encoded in TEI P5 using a customised ODD schema which, to a certain extent, was influenced by the ENRICH schema (http://projects.oucs.ox.ac.uk/ENRICH). However, some modifications have been necessary, which means that our schema and the ENRICH schema are not completely compatible. Our schema also provides

bstractcataloguing guidelines with examples, and documentation of the elements and attributes used. In order to eliminate inconsistencies in the encoding, we enforce strict limits regarding the attributes and their values. We also make use of Schematron rules to check that the content is valid in certain elements. The manuscript descriptions are structured around the notion of codicological units: the intellectual content, physical description, and history, where applicable, of each unit is described in separate <msPart> elements, whereas information common to all units, e.g. the binding, provenance, and bibliography, is described outside the <msPart> elements, directly under the <msDesc> element. In each manuscript description the items will be linked to specific folios of the digitized manuscripts using the @facs attribute in the <locus> element. This will enable image retrieval from the image server.

The Uppsala University Library digital repository, ALVIN, will host the TEI-files, schema, style sheets, and images of the digitized manuscripts. The images will be stored both in the original TIFF-format, and in the JPEG2000-format for display on the web. The JPEG2000 images will be saved in a multi-resolution tiled pyramidal format to enable fast zooming and panning capabilities in the image viewer.

The manuscript descriptions are published continuously on the project web site as the project advances (http://www.manuscripta.se). The web interface for the catalogue is built with the eXist-db open source software (http://exist-db.org) which provides powerful search and browsing capabilities for the catalogue. To facilitate user-friendly orientation, the images are displayed adjacent to the manuscript description in the interface.

TEI-files, schema, style sheets, and source code for the eXist-db interface are also stored in the project GitHub repository (https://github.com/manuscripta). This enables sharing and re-use of the material, and makes it possible for anyone to download all files in bulk. We strongly believe in the idea of open access and open source to achieve transparency and encourage scholarly cooperation.

Acknowledgements

Bibliography

• Andrist, P., Canart, P. & Maniaci, M. (2013), La syntaxe du codex: essai de codicologie structurale, Turnhout

• Andrist, P. (2007), Les manuscrits grecs conservés à la Bibliothèque de la Bourgeoisie de Berne – Burgerbibliothek Bern. Catalogue et histoire de la collection, Dietikon-Zurich

- Enrich schema. URL: http://projects.oucs.ox.ac.uk/ENRICH

- Project web page. URL: http://www.manuscripta.se

- eXist database. URL: http://exist-db.org

- Project source code. URL: https://github.com/manuscripta