Poster Presentation Abstract
Digital Humanities in the Nordic Countries 2018

Katarina Pihlflyckt
Svenska litteratursällskapet i Finland
Zacharias Topelius Skrifter
katarina.pihlflyckt@sls.fi

# ZACHARIAS TOPELIUS SKRIFTER – APPROACHING A DIGITAL SCHOLARLY EDITION THROUGH METADATA

This poster presents an overview of the database structure in the digital critical edition of Zacharias Topelius Skrifter (ZTS). The poster specifies what kind of information is available for the user and shows how the entity relations open a possibility for the user to approach the edition from other angles than the texts, by taking advantage of the information and descriptive metadata in indexing systems. Through this data, a historian can capture for example events, meetings between people or editions of books, as they are presented in Zacharias Topelius' (1818–1898) texts. Presented here are both already available features and features in progress.

ZTS comprises eight digital volumes hitherto, the first published in 2010. This includes the equivalent of about 8 500 pages of text by Topelius, 600 pages of introduction by editors and 13 000 annotations. The published volumes cover poetry, short stories, correspondences, children's textbooks, historical-geographical works and university lectures on history and geography. Genres still to be published include children's books, novels, journalism, diaries, religious and academic texts. It is freely accessible at topelius.sls.fi.

## DATABASE STRUCTURE

The ZTS database structure consists of six connected databases: people, places, bibliography, manuscripts, letters and a chronology. So far, the people database consists of about 10 000 unique persons, and a possibility to link them to a family or group level (250 records). It has separate chapters for mythological persons (500 records) and fictive characters (250 records). The geographic database has 6 000 registered places. The bibliographic database has 6 000 editions divided on 3 500 different works, and the manuscript database has 1 400 texts by Topelius on 350 physical manuscripts. The letter database has 4 000 registered letters to and from Topelius, divided on 2 000 correspondences. The chronology of Topelius life has 7 000 marked events. The indexing of objects started in 2005, using the FileMaker system. The preliminary idea at that time was to give the users similar information as in indexes and annotations in printed scholarly editions of the same type. The work with finding more possibilities on how to use, link and present the data is in constant progress. Restrictions are mostly due to time-limiting factors within the editorial work. Challenges lay in the requirement of having shared data for different genres and in working coherently with a large amount of material for many years in the ever-developing digital world.

The bibliographic database is the most complex database, and follows the *Functional Requirements for Bibliographic Records* (FRBR) model. This means we are making a difference between the abstract work and the published manifestations (editions) of that work. The FRBR focuses on the content relationship and continuum between the levels; anything regarded a separate work starts as a new abstract record, from where its own editions are created. Within ZTS, the abstract level has a practical significance, in cases when it is impossible to determine to which exact edition Topelius is referring. Also taken in consideration is that for example articles and short stories can have their own independent editions as well as being included in editions (e.g. magazines, anthologies). This requires two different manifestation levels subordinated the abstract level; the regular editions and the texts included in other editions, the records of the latter type must always link to records of the former.

The manuscript database has a content relationship to the bibliographic database through the abstract entity of a work. A manuscript text can be regarded as an independent edition of a work in this context (a manuscript that was never published can easily have a future edition added in the bibliographic database). The manuscript text itself might share physical paper with another manuscript text. Therefore, the description of the physical manuscript is created on a separate level in the manuscript database, to which the manuscript text is connected. The display of the manuscript database is still under planning.

The letter database follows the FRBR example; an upper level presents the complete abstract correspondence between Topelius and another person, and a subordinated level describes each physical letter within the correspondence. It is also possible to attach additional corresponding persons to occasional letters. The letter database is currently only available through the manuscript descriptions for the separate published letters.

The people database connects to the letter database and the bibliographic database, creating a one-to-many relationship. Any writer or author has to be in the people database in order to have their information inserted into the other two databases. Within the people database there is also a family or group level, where the records of individual family members can be grouped, but in contrary to the letter database, this is not a superordinate level. The people database is available through the indexes and the annotated people in the XML files. The group level is still under construction and currently not included in the display view.

The geographic database follows a one-level structure. Places in the letter and manuscript databases are linked from the geographic database, in a similar way as people. The geographic database is available through the indexes and the annotated places in the text.

The chronology database contains manually added key events from Topelius' life, as well as short diary entries from various calendars during his life. The main content, however, is automatically gathered records from other databases, linked through the marked dates when Topelius works were published or when he wrote a letter or a manuscript. Dates of birth and death of family members and close friends can be linked from the people database. The chronology creates a timeline that would not only give the user key events from Topelius' life, but also links to the other database records. Encoded dates in the XML files (letters, diaries, lectures, manuscripts etc.) could lead the user directly to the relevant text passages. The work with the chronology is for the moment not prioritized and a timeline display currently not available.

## POSSIBILITIES FOR THE USER

Approaching a digital scholarly edition with over 8 500 pages can be a heavy task, and many will likely use the edition more as an object to study, rather than texts to read. For a user not familiar with the content of the different volumes, but still looking for specific information, advanced searches and indexing systems offer an alternative path into the relevant text passages. The information in the ZTS database records helps provide a picture of Finland in the 19th century as it appears in Topelius' works and life. A future feature for users is freely access to the data in the published records through an API (Application Programming Interface). This will create opportunities for the user to take advantage of the data in almost any wanted way: to create a 19th century bookshelf, an app for the most popular 19th century names or a map of popular student hangouts in 1830's Helsinki.

Through the indexes formed by the linked data from the texts, the user can find all the occurrences of a person, a place or a book in the whole edition. One record can build a set of ontological relations, and the user can follow a theme, while moving between texts. A search for a person will provide the user with information not only about the person, his/hers close family and the possible personal relation to Topelius, but also where Topelius mentions this person, what he has to say about the person, or if they possibly meet or interact. Furthermore, the user will be able to see if this person was the author of a

book mentioned by Topelius in his texts, or if the editors at ZTS have used the book as a source for editorial comments. The user will also be able to find the person's possible correspondence with Topelius. The geographic index can help the user create a geographic ontology with an overview of Topelius' whereabouts through the annotated mentions of places in Topelius' diaries, letters and manuscripts. Map coordinates are not included in the database, partly because of time-limiting reasons and partly because of the difficulty in finding correct coordinates for places that have disappeared or changed since the 1800's. Available are instead descriptions of the places and the bigger entity where it is situated (e.g. city, region or country). A division is made between places in Finland and outside Finland. Uncertainty is encoded in the XML file in cases where the editor cannot be sure, whether a person, a place or a literary work that Topelius mentions is the same as linked to in the database, and the connection shown as uncertain in the view.

The entity relation between the bibliographic database and the manuscript database creates a complete bibliography over everything Topelius wrote, all known manuscripts and printed editions that relate to a specific work. So far, there are 900 registered independent works by Topelius in the bibliographic database; these works are implemented in 300 published editions (manifestations) and 2 900 text versions included in those manifestations or in other independent manifestations. The manuscript database consists of 1 400 manuscript texts, of which 600 are linked from 350 abstract works in the bibliography. The FRBR model offers different ways of structuring the layout of a bibliography according to the user's needs, either through the titles of the abstract works with subordinate manifestations, or directly through the separate manifestations. Topelius' texts on the abstract level are also marked according to genre (e.g. journalism, children's books etc.) and the bibliography can be arranged based on this. Furthermore, the bibliography points the user to the published texts and manuscripts of a specific work in the ZTS edition and to text passages where the author himself discusses the work in question. The bibliography will be available in 2019, together with a register over all works by other authors that Topelius is mentioning in his text, no matter whether we know the exact edition or only the abstract work.

The level of detail is high in the records. For example, we register different name forms and spellings (*Warschau* vs *Warszawa*). Such information is included in the index search function and thereby eliminates problems for the end user trying to find information. Topelius often uses many different forms and abbreviations, and performing even an advanced search in the texts would seldom give a comprehensive result in these cases. The letter database includes reference words describing the contents of the correspondences. Thus, the possibilities for searching in the material are expanded beyond the wordings of the original texts.