Integrating historical scientific texts into the Bernoulli–Euler Online platform

Tobias Schweizer
t.schweizer@unibas.ch
Digital Humanities Lab, University of Basel, Switzerland

Sepideh Alassi
sepideh.alassi@unibas.ch
Digital Humanities Lab, University of Basel, Switzerland

Martin Mattmüller
martin.mattmueller@unibas.ch
Bernoulli Euler Centre, University of Basel, Switzerland

Lukas Rosenthaler
lukas.rosenthaler@unibas.ch
Digital Humanities Lab, University of Basel, Switzerland

Helmut Harbrecht
helmut.harbrecht@unibas.ch
Bernoulli Euler Centre, University of Basel, Switzerland

Introduction

Bernoulli-Euler Online (BEOL) is an interdisciplinary research project funded by the Swiss National Science Foundation focusing on the mathematics influenced by the Bernoulli dynasty and Leonhard Euler. It is being carried out by the Bernoulli Euler Centre and the Digital Humanities Lab at the University of Basel. Its main goal is the integration of different edition projects relating to the Bernoullis and Leonhard Euler into *one* target platform, offering appropriate functionality for researchers interested in the history of science.

The methodological efforts will also be applicable to other editions since they are developed in a generic way. BEOL is based on Knora, a generic infrastructure for humanities data.

Goal of the BEOL-platform and its technical basis

BEOL aims at integrating three edition projects, that are currently all technically different and thus incompatible with one another:

- *Basler Edition der Bernoulli-Briefwechsel* (BEBB): BEBB is an online edition that is based on the MediaWiki software and hosted by the University Library of Basel. It is connected to the library's metadata catalogue for manuscripts (Basler Inventar der Bernoulli-Briefwechsel). The letters are encoded in Wiki markup and are converted to HTML to represent them on the web. The mathematical formulae are encoded in LaTeX.

- *Leonhardi Euleri Opera Omnia* (LEOO): LEOO is a printed edition of the works of Leonhard Euler that was begun in the early 20th century. In the context of BEOL, the volume containing Euler's correspondence with Christian Goldbach (Euler 2015) will be integrated as a proof of concept. This volume has been prepared using LaTeX (as well as the volume with Euler's correspondence with Daniel Bernoulli that has been published recently). We expect to be able to integrate all the other recent volumes set in LaTeX in a similar manner. For the older volumes, the printed books would have to be scanned (including OCR) and marked up.

- Jacob (I) Bernoulli's scientific notebook *Meditationes*: The manuscript is held in the university library of Basel (shelfmark L Ia 3, 367 pages) and has already been digitized. The manuscript consisting of 287 entries is being transcribed at the Bernoulli Euler Centre using XML (The XML format is specified closely to the TEI specifications P5, so it can be transformed quite easily to TEI/XML) for the text and LaTeX for the mathematical notation that is embedded in the XML.

The three edition projects do not only overlap thematically, but also in terms of the persons involved (authors, mentioned persons) and bibliographical items (literature referred to in the texts, references in-between the editions' texts). Letters exchanged between members of the Bernoulli dynasty, Leonhard Euler and

contemporary mathematicians and scientists are an important part of these edition projects and thus it is desirable to identify and match the persons in all editions in order to display their relations.

The technical basis for BEOL is Knora, an infrastructure for humanities data (Rosenthaler and others 2015) consisting of an RDF-triplestore, an OWL base ontology, and a RESTful API that allows for querying and adding to the data. The base ontology (see prefix 'Knora' in Figure 1) defines common value types (such as a calendar independent format to represent dates using the Julian Day Number) used among humanities projects and can be further extended in project specific ontologies. BEOL will provide such an ontology (see prefix 'BEOL' in Figure 1), defining its own resource classes and properties needed to represent the edition projects' texts and entities. Wherever possible, existing ontologies will be reused by making subclasses and subproperties. BEOL is part of the NIE-INE project, which aims to create a general-purpose infrastructure for digital editions, using Knora as its technical foundation. A focus of this project will be abstracting out concepts shared by different projects and formalising them as ontologies.
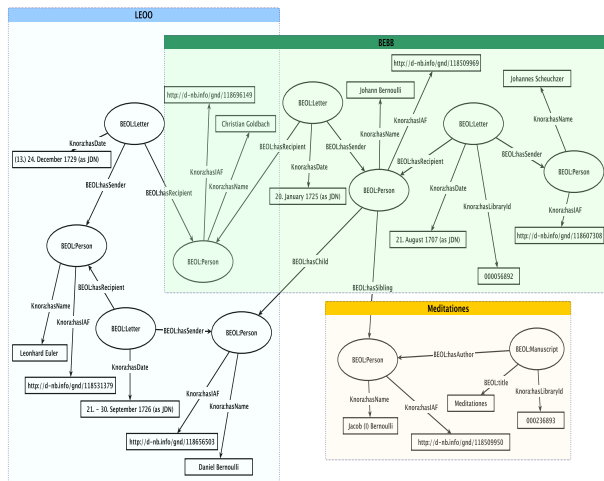

Figure 1: BEOL network and its components

Figure 1 represents all relations between persons (We refer to the Integrated Authority File (GND), and in order to represent locations, we will also refer to GeoNames), letters, and manuscripts (we also link to the catalogue of the Basel university library that keeps many of the original copies of the letters and manuscripts of BEOL), as well as their properties as directed graphs. For reasons of clarity, we use a simplified model here. The coloured rectangles indicate that these have been imported from different edition projects which – considered in isolation – do not allow for

this kind of overview. Moreover, indices and bibliographies have to be unified on the BEOL platform (e.g., Christian Goldbach occurs both in BEBB and LEOO). The BEOL platform will be connected to Early Modern Letters Online, so it will be interoperable with other edition projects.

Importing editions to the same target environment

In order to represent all three editions in the same target environment, they have to be homogenised first. We decided to do so using an XML-based approach. This has the additional advantage that we can make both the texts of BEBB and LEOO available as TEI/XML to the outside world quite easily by applying XSL transformations. We can also use the same routine to import the editions into BEOL. Knora converts XML-encoded texts to RDF in order to store them in the triplestore. From RDF, an XML document can be recreated that is equivalent to the one originally imported. A mapping defines the relations between XML elements and attributes and the entities defined in the ontology.

- BEBB Wiki markup can be transformed to XML using a MediaWiki parser. Wiki tags and structures are mapped to XML tags, and references to other letters, bibliographical items, and images (facsimiles of the letters) can be handled. Once the letters are available on the BEOL platform, the old URLs will have to be forwarded.

- The Goldbach-volume of LEOO is set in LaTeX and can be converted to XML using LaTeXML. Additional mappings to the available standard functionality and customisations can be provided using Perl scripts. LaTeXML provides a MathML conversion for mathematical formulae.

- The Meditationes are transcribed in an XML-based format (see LaTeXML). Derived texts of these files can be generated using XSL-transformations. In this way, several layers (diplomatic, normalized) of the text can be produced. Our approach addresses segments defined on the facsimile (see Figure 2) and turns them into a reading text step by step. The figures (see segment 'M151-03-F' in Figure 2Figure) will be extracted by applying a combination of various image processing techniques and redrawn as vector graphics.
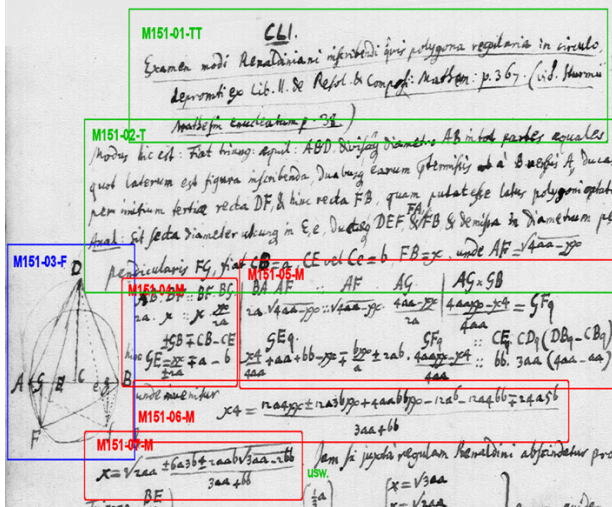
Figure 2. Part of Meditatio 151

One of the main challenges in the BEOL project is the faithful representation of mathematical notation and its relation to the surrounding text (see Figure 2) using web technologies. At the moment, we are using MathJax (which accepts both LaTeX and MathML as input formats) to render the mathematical formulae in the web browser. We also consider MathML as an option, although not all web browsers fully support MathML.

We are aiming at developing a browser based user interface that will be based on the Angular 2 framework (although in the meantime, we are already using Knora's current interface, SALSAH) that not only makes it possible to present the texts on the web and to offer search functionality, but also to add to the data (sufficient permissions provided). The users may create their own annotations on the BEOL platform. Basically, the user interface interacts with the Knora API in order to create new resources, manipulate properties etc. Since BEOL is based on Knora, all of its generic functionality can be used for this purpose.

Conclusion

BEOL integrates three different edition projects into one platform and allows researchers to query previously separated contents and add to them. The specific problems posed by the combination of text and mathematical notation can be addressed in a generic manner. All the functionality to be developed will be part of Knora and can be reused by other projects dealing with scientific texts from mathematics and physics.

Bibliography

Euler, L. (2015): *Leonhardi Euleri Opera Omnia.* Vols. IVA/4: Correspondence of Leonhard Euler with Christian Goldbach, ed. by Martin Mattmüller and Franz Lemmermeyer. Basel 2015.

Rosenthaler, L., et al (2015) *Final Report for the Pilot Project "Data and Service Center for the Humanities".* Swiss Academy of Humanities and Social Sciences. http://www.sagw.ch/dms/sagw/laufende_projekte/DaSCH/FinalReport-DaSCH_print