# WorldViews: Access to International Textbooks for Digital Humanities Researchers

**Steffen Hennicke**
schwedes@leibniz-gei.de
Georg Eckert Institute for Int'l Textbook Research
Germany

**Lena-Luise Stahn**
stahn@leibniz-gei.de
Georg Eckert Institute for Int'l Textbook Research
Germany

**Ernesteo William De Luca**
deluca@leibniz-gei.de
Georg Eckert Institute for Int'l Textbook Research
Germany

**Kerstin Schwedes**
Georg Eckert Institute for Int'l Textbook Research
Germany

**Andreas Witt**
witt@ids-mannheim.de
Institut für Deutsche Sprache, Germany

## Abstract

This paper introduces the field of international textbook research and discusses how the WorldViews project is working towards enhanced access to textbook resources for digital humanities research.

## Textbook Research

The field of textbook research is one of the more recent and more diverse areas of academic investigation. The condensed and canonical character of the information selected for inclusion in textbooks - here understood as conventional textbooks - gives them central significance in academic, political and educational respects. Textbooks, as carriers of the knowledge and information that one generation wishes to pass on to the next, frequently find themselves at the center of political controversy. As such, their importance as an object of investigation in historical and cultural research has gained significant traction in recent decades. However, textbook research has not yet found dedicated representation as a main subject at universities.

As a non-university institution, the Georg Eckert Institute (GEI) for International Textbook Research conducts and facilitates applied and multidisciplinary research into textbooks and educational media primarily informed by history and cultural studies. For this purpose, the GEI provides digital and social research infrastructure services such as its renowned research library and various dedicated digital information services, such as Edumeres, a virtual network with modules for specific aspects of textbook research. As such, the GEI realizes a unique position in the international field of textbook research.

The study of textbooks has not only been facilitated by growing institutional support and infrastructure but also by the proliferation of new digital methods and resources in humanities research. In the digital humanities, the investigation of research questions is supported by a range of increasingly sophisticated digital methods such as automatic image and text analysis, linguistic text annotation, or data visualization. Digital tools and services combined with the increasing amount of resources available through digital libraries (such as the German Digital Library, the Deutsches Textarchiv, and Europeana) and research infrastructures (such as CLARIN or DARIAH) provide digital support for textbook analysis.

## Digital Information Services

At the GEI, the shift towards more digitally oriented research has resulted in a range of digital information services specifically tailored to textbook research. EurViews, for example, is a multilingual digital platform containing primary sources from twentieth and twenty-first century history textbooks from around the world that manifest particular concepts of Europe and Europeanness (for example, see Gehler and Vietta, 2010; Best et al, 2012, and Chakrabarty, 2000). The service also offers essays, commentaries, and educational histories written by designated experts in the field. EurViews is a useful tool for historians searching for relevant, reliable and hard-to-find research materials on topics related to textbooks. The materials may provide inspiration for research projects or be the starting point for more extensive searches for sources. EurViews also demonstrates that the printed monograph is no longer the dominant

form of publication but that digital representation is gaining in importance.

Digital representations are increasingly becoming objects of investigation themselves. GEI-Digital , for example, a digital library focusing on out-of-copyright works published between the inception of textbooks in the seventeenth century and the demise of imperial Germany in 1918, holds potentially relevant textual resources for EurViews. Other important information services provided by the GEI offer factual and bibliographic data relevant for textbook research: edu.data, for example, provides information about textbook systems in individual countries, and the library catalogue contains bibliographic metadata about textbooks from around the globe.

While these services all contain resources and information that are immediately relevant to textbook research, their content is frequently stored in isolated data silos that lack appropriate interfaces or standardized data models and which prevent convenient use or exchange of data within the GEI or with external services. For example, even though GEI-Digital makes metadata about its resources available as METS/MODS encoded data via an OAI-PMH interface, EurViews lacks the interfaces that would enable it to utilize these full-text resources. Similarly, data on textbook systems from edu.data or bibliographic metadata from the library catalogue cannot be reused by existing or new research projects. External digital humanities infrastructures such as CLARIN-D are also unable to easily access texts stored in GEI-Digital or EurViews since those platforms contain plain text documents only which are not semantically annotated.

## WorldViews

The WorldViews project, which is financed by the Federal Ministry of Education and Research (BMBF), addresses these challenges in order to further elevate access to textbook resources. Its aim being to enhance discoverability, reusability, and sustainability of textbook resources in cultural and historical research and in the digital humanities. The formative use case has been the digital platform for textbook sources, EurViews, a service that has proven to be in high demand– Since 2015 EurViews has had an average of 400 visitors per month, a comparatively high number for this kind of digital academic service– and which constitutes one of the cornerstones of the GEI's research infrastructure.

Two significant milestones have been achieved on the path to enhanced access. The first is the selection of the Text Encoding Initiative (TEI) standard that will be used to encode the semantics of conventional textbook resources and thus facilitate access to full-text sources. In addition, the Component MetaData Infrastructure (CMDI) framework has been chosen for the overall integration of metadata at the GEI; a framework also compatible with the CLARIN-D infrastructure.

An extensive search for publicly available and dedicated encoding schemas for textual characteristics of conventional textbooks yielded no results. Therefore, a profile specifying the most relevant metadata and textual features in textbooks was developed from scratch after consulting historians at the GEI. The only viable option for creating an encoding schema based on the profile turned out to be the TEI encoding standard. The TEI schema created for textbooks focuses on basic elements for the selective and formal description of those structural and semantic features that are immediately relevant for WorldViews. For example, headings of sections or the semantics of particular paragraphs should provide the necessary semantics to enable retrieval scenarios to contextualize search results and to formulate more precise queries targeting particular segments of the text. This ground-breaking schema is designed to provide the nucleus for more comprehensive descriptions of whole textbooks, such as those found in GEI-Digital.

The Component MetaData Infrastructure (CMDI) provides a framework that allows blueprints of distinct metadata components to be defined and reused. CMDI allows standard metadata components to act as profiles for virtually any kind of metadata. CMDI profiles have been created to describe textbooks resources in WorldViews (we have switched to the new version of CMDI, released in mid-2016) and we are currently investigating their application in more fact-based resources such as edu.data. By using CMDI as a general framework for metadata descriptions, full-text resources can immediately be indexed by CLARIN's Virtual Language Observatory and can be analyzed using its various tools and services such as Weblicht. The CMDI description of GEI resources allows for internally standardized search and retrieval operations in federated search scenarios.

The second milestone for better access to textbook resources was the implementation of a logic tier. Central components of this tier are Solr-indices for federated searches, tools for handling digitization workflows, controlled metadata annotation and full-text annotation of textbook sources and, significantly,

a [Fedora](#) repository that will dynamically provide access to standardized representations of textbook resources and other data from the various digital infrastructure services at the GEI for internal as well as external consumers. After extensive evaluation of similar repository software such as DSpace, Fedora was selected due to the greater customizability and flexibility offered by its strong modularity and existing applications with CMDI metadata. Furthermore, use of Fedora is a prerequisite for becoming a CLARIN center, one of the long-term objectives of the GEI. Through the logic tier digital platforms such as EurViews are able to send queries for bibliographic metadata on newly added textbooks directly to the library catalogue or they can directly reuse textbook resources from other systems such as GEI Digital or obtain contextual data from databases such as edu.data.

## Summary

The WorldViews project has four main strategies aimed at improving overall access to digital textbook resources through enhanced reusability, discoverability and sustainability. These are a logic tier based on the Fedora repository, which mediates data between internal and external services; the digitization workflow tool Goobi, which controls metadata descriptions of digitized textbook resources; full-text encoding schemas based on TEI; and the metadata standardization based on the CMDI framework. WorldViews has laid the groundwork for standardized data models at full-text and metadata level in the field of textbook research; thereby providing a firm foundation for textbook resources and related data to be made accessible for digital humanities research.

## Bibliography

**Gehler, M., and Vietta, S.** (Eds). (2010). Europa – Europäisierung – Europäistik, Wien.

**Best, H., Lengyel, G., and Verzichelli, L.** (Eds) (2012). *The Europe of Elites. A Study into the Europeanness of Europe's Political and Economic Elites*, Oxford/ New York 2012

**Chakrabarty,D.** (2000) *Provincializing Europe. Postcolonial Thought and Historial Difference,* Princeton, New Jersey.

**CLARIN.** (n.d.) Component Metadata Infrastructure. https://www.clarin.eu/content/component-metadata [Accessed 24 March 2017]

**CLARIN**. (n.d.) Virtual Language Observatory. https://www.clarin.eu/vlo [Accessed 24 March 2017]

**Gehler, M., and Vietta, S.** (Eds). (2010). Europa – Europäisierung – Europäistik, Wien.

**Georg Eckert Institute for International Textbook Research.** (n.d) EurViews. http://www.eur-views.eu/nc/en/en/home.html [Accessed 24 March 2017]

**Georg Eckert Institute for International Textbook Research.** (n.d) GEI Digital. http://gei-digital.gei.de/viewer/ [Accessed 24 March 2017]

**Georg Eckert Institute for International Textbook Research.** (n.d) WorldViews Project. http://worldviews.gei.de/en/ [Accessed 24 March 2017]