

Review of

digital Monumenta Germaniae Historica (dMGH)

Monumenta Germaniae Historica and Bavarian State Library. 2005-2010 (current installation). Web <<http://www.dmgh.de>>.

by

Georg Vogeler & Patrick Sahle

to be published in Digital Philology 2 (2013),
Print ISSN: 2162-9544 - Online ISSN: 2162-9552

Author's version with some additional features (screenshots, footnotes, bibliography), April 20, 2013

digital Monumenta Germaniae Historica (dMGH)

Monumenta Germaniae Historica and Bavarian State Library. 2005-2010 (current installation). Web <<http://www.dmgh.de>>.

Patrick Sahle, Universität zu Köln <sahle@uni-koeln.de> - Georg Vogeler, Universität Graz <georg.vogeler@uni-graz.at>

The Monumenta Germaniae Historica (MGH) publishes scholarly editions of source material relevant to the history of the Realm of the Franks and the Holy Roman Empire in the Middle Ages. The enterprise began in the early nineteenth century and became a research institution of its own in 1935. The publication of critical editions remains the central goal of the MGH. Over the last 150 years, the MGH has gained a reputation for the high quality of their editions, although they had to redo some of their early volumes in order to integrate new research results. The series is divided into several sections: historiography in the *scriptores*, law texts in the *leges*, charters in the *diplomata*, letters in the *epistolarum*, and various genres from poetry to necrologies in the *antiquitates*. In the international community, the c. 400 volumes are considered the most prestigious editions of continental European sources of the early and high Middle Ages, both in terms of scholarly standards and completeness. The majority of these sources are in Latin, with a smaller selection in German.

By the early 1990s, the MGH started to digitize its editions in cooperation with Brepols Publishers under the label of *eMGH* (electronic MGH). The *eMGH* as a CD-ROM publication from 1996 onwards was based on the CETEDOC software, and the resulting product is mainly intended as a philological research tool (SAHLE 2002, MÜLLER 2003). From 2004 to 2010 the MGH received three grants from the Deutsche Forschungsgemeinschaft (DFG) to create a new digital version (thus *dMGH*) of the existing printed editions in cooperation with the Digitization Center at the Bavarian State Library. This review draws upon a monographic discussion of the *dMGH* (SAHLE2008), shifts its focus towards philology, and extends it to the project's recent developments.

As of 2012, the *dMGH* present without any access restriction an impressive amount of textual material: some 350 volumes with more than 165.000 pages, presumably all editions that have been published until 2007. 23 of these volumes are parts of two additional series, which are not part of the core sections, i.e. the *Quellen zur Geistesgeschichte des Mittelalters* and *Deutsches Mittelalter, Kritische Studentexte*. New editions are supposed to be added every three years after their publication in print – according to the official statement. However, when checking the website in late November 2012, we found some volumes from 2005 and 2007 that are still missing (*capitula episcoporum* 4; *diplomata* 14,2; *concilia* 6,2). In fact, we have been told by a project staff member that 30 volumes from 2005 to 2009 should be available online in early 2013. All editions are available as digital images as well as electronic full-texts.

A declared aim of the *dMGH* was to create an identical representation of the printed version to ensure that a citation from the *dMGH* could be found as well in the printed version. Accordingly, the website offers a main menu with the series and volumes of the printed version. The interface is primarily a tool to flip through page images in the original order. Once the user has chosen a volume,

further browsing through a table of contents or by page number is possible. The user can zoom in and out of the images, and also switch between scanned images and an HTML version of the full text. The scans can be printed as single images or downloaded as a PDF for larger parts, even entire volumes. However the user should be aware that the PDF contains only images, not the full text. To ensure “citability” in the digital realm, every page has a permanent link that is part of the URN-identification system used by the Bavarian State Library. But the *dMGH* also provides another mechanism that draws on the established canonical references to the volumes and their pages and implements a PURL approach (persistent URL). To give an example, page 40 in the *scriptores* volume 17, can be addressed as <http://www.mgh.de/dmgh/resolving/MGH_SS_17_S_40>.

The section entitled “Datenbanken” of the MGH homepage (but not on the *dMGH* site itself) includes very useful, additional information on MGH editions: additions and new editions of charters, PDF versions of texts to be published in print, and images of manuscripts extracted from photographs in the collection of the institute. Currently the only way to integrate this information into the *dMGH* seems to be the publication in print and re-launch of the print version in the *dMGH* after three years. Most of this data is declared as “preliminary” or “uncritical” documentation, leaving the final scholarly decisions to the print publication, which is still considered to present the canonical text versions.

The screenshot shows the website for Monumenta Germaniae Historica (MGH). At the top, there is a logo for the Bayerische Staatsbibliothek and Deutsche Forschungsgemeinschaft (DFG). The main navigation bar includes links for 'SUCHE', 'BÄNDE', 'HILFE', 'PROJEKT', and 'IMPRESSUM'. Below the navigation, there is a search bar with the text 'Suche im Band' and a search icon. To the right of the search bar, there are links for 'PDF-Download' and 'Scan ausdrucken'. Below the search bar, there is a zoom control with percentages from 25% to 150% and navigation arrows. The main content area displays the title 'MONUMENTA GERMANIAE HISTORICA' and 'DIPLOMATA REGUM FRANCORUM E STIRPE MEROVINGICA'. On the left side, there is a table of contents for the Merovingian section, listing various items such as 'Mer. 2: Merowingier 2 (DD Mer.)', 'Titelblatt (lat.)', 'Titelblatt (dt.) 2. Teil', 'Inhaltsverzeichnis', 'Deperdita', 'Empfängerübersicht (Deperdita)', 'Titelblatt: Appendix', 'Moderne Fälschungen', 'Titelblatt: Register', 'Namenregister', 'Wortregister', 'Empfängerverzeichnis', 'Archivalische Übersicht', 'Abkürzungen', 'Quellen und Literatur', 'Konkordanzen', and 'Tafel: 1. Präzept Chlothars II.' through 'Tafel: 8. Erhaltene Siegel'.

The conversion of the page images in the *dMGH* into full-text allows for the implementation of additional search functionality, which is offered in two places. First, there is the option to restrict the query to particular volumes. Second, a generic interface with the same look is offered through the main navigation panel. Here, the search can be narrowed down according to the major sections and to types of text: edited text, textual apparatus, comments, and other texts (like forewords). The search input field comes with

auto completion suggestions, which increase its heuristic value significantly. The user can further adjust the search results by determining their order (by relevance, title, year, or volume) and appearance. Hits can be shown as image snippets, as bibliographic information, and as text in context.

The search interface offers some basic full-text indexing features such as *and*-concatenation, normalization of German umlaut, a

The screenshot shows a search interface with the following elements:

- search term(s)**: A text input field with a magnifying glass icon.
- search areas**: Four checkboxes for "edited text", "critical apparatus", "historical apparatus", and "other texts".
- section**: A dropdown menu currently set to "all sections".
- search options**: Three dropdown menus for "sort by relevance", "descending", and "no result contexts (faster)", followed by a "find!" button.
- Links for "[help]" and "[new search]" are located at the bottom left of the interface.

not-operator (minus sign), exact phrase search (quotation marks), right truncation (*), single letter wildcard (?) and stop words. The primary search entity is the volume. Queries across the whole corpus deliver a list of volumes containing the search term. Search queries across one single volume separately show a keyword-in-context list. The division of the volumes into single edited texts or other deeper text structures in the volumes is not possible.

The relevance algorithm applied to the search function tends to deliver misleading results. A Latin phrase search proves to be too sensitive to word order, as e.g. is the case of hyperbata: A search for *sacrum imperium*¹ in *Staatsschriften* 1.1 gives only one correct hit within the first ten; in *Const.* 5 *sacrum vacare continget imperium* scores at 75% while *promovendum ad sacrum imperium* gets only 50%.² From the philological point of view they have both the same relevance. All the others are occurrences of the single word *imperium*. Similar problems show up when searching for the phrase in the whole corpus. The above mentioned *Staatsschriften* volume scores at 69.78%,³ while *Const.* 3, with five correct hits in the first ten, scores only at 65%.

One can consider the *dMGH* as a major resource for early and high medieval Latin texts. Together with projects like ALIM (Archivio della Latinità Italiana del Medioevo)⁴ or the CETEDOC data⁵ it constitutes a veritable corpus of the Latin language in this period. The major drawback is that it is still difficult to access the texts with linguistic or philological research interests. Someone looking for ready-made tools (like lemmatized search, explicit proximity search, etc.) may be still better off with the *eMGH*. Someone looking for a preliminary text base to be converted in the format of their own research tools, will be disappointed, as the data source is not available. This person would have to write their own scripts to extract the texts he or she is interested in from the HTML or XML pages

¹<http://www.dmgh.de/de/fs1/search/query.html?fulltext=sacrum+imperium&text=true&subSeriesTitle_str=&sort=score&order=desc&hl=false&action=Finden!>

²<http://www.dmgh.de/de/fs1/object/context/bsb00000666_meta:titlePage.html?context=imperium+sacrum&sortIndex=020%3A080%3A0011%3A010%3A00%3A00&sort=score&order=desc&hl=false&fulltext=sacrum+imperium&text=true&contextSort=sortKey&contextOrder=descending&contextType=scan&action=Finden!>

³<http://www.dmgh.de/de/fs1/object/context/bsb00000646_meta:titlePage.html?text=true&sort=score&order=desc&subSeriesTitle_str=&hl=false&fulltext=sacrum+imperium&sortIndex=010:100:0001:010:01:00&context=sacrum%20imperium>

⁴<<http://www.uan.it/Alim>>

⁵Migne Patrologia Latina at documenta catholica omnia
<http://www.documentacatholicaomnia.eu/25_10_MPL.html> unfortunately provides only images of the printed texts.

respectively.⁶ If the MGH still considers its *raison d'être* to be providing base material for further research, it should consider going further in the open data direction.

The *dMGH* is an extremely rich information resource even apart from the edited texts, as is evident, for example, in the indices on persons and places or the manuscript descriptions in the introductions. This information was incorporated into the *dMGH* as full text, partially with added functionalities, as the *iMGH* (*indices tomorum Monumentorum Germaniae Historicorum*); the digital texts of the indices for the volumes of the *Scriptores in Folio* (excluded vol. 15) were converted into a database format that links the index entries to the texts. In the interface, there is currently only a link via "Zusatzinfo" at the top of the page, including the index entries for the current page, although the complete index can be accessed directly via <http://www.mgh.de/dmgh/indices>. In addition to these indices, the MGH created a database of medieval place names including geo-coordinates. The database contains the information extracted from Graesse-Benedict-Plechls' *Orbis Latinus* (2. ed. 1909) (more than 7000), the *GeoNames-Database* (583 entries) and user contributions. Perhaps due to a lack of publicity, these user contributions amounted to only fifteen entries as of December 10, 2012. In an example, the *dMGH* connected this data with the place names in the edition of the Merovingian charters <http://www.mgh.de/dmgh/geo/demo/merowinger/> and showed the possibilities of a simple GIS-like presentation. A database of the manuscript descriptions in the volumes is under construction and certainly will become a very useful resource.⁷ In the global view, the idea of a collective index of all volumes, including references to manuscript shelfmarks, remains yet to be realized.

The main focus of the MGH has always been a historical one. The texts are sources for the research in the history of Central and Western Europe. This affects mainly the selection of texts but also the editorial method; the edited texts are of a Lachmannian nature: corrected, normalized and cautiously modernized. While the texts are not lemmatized for more powerful searching, they do not follow the exact historical spelling either. Still, the *dMGH* form an impressive and extensive corpus for certain textual genres and can thus be very useful for philological research as well.

The *dMGH* is not primarily a linguistic text corpus or research tool for literary studies. Neither is it a digital scholarly edition that tries to solve editorial problems with digital methods. It is rather a scholarly digital library. Compared to many other digitization projects, which stop indexing at the division of text at chapter level, the *dMGH* implements a deeper data structure, offers reliable full texts, and provides a clear and useful browse and search interface. From this perspective, the *dMGH* has reached its own goals. Admittedly, we would be happy to have *all* formerly printed resources (other than MGH) made available in the same manner, but there are several areas where we could imagine some progress to come. In the long run, a deeper exploitation of the material and a better integration into the evolving digital humanities infrastructure or virtual research environments is desirable. Persistent URLs and web services resolving canonical references point in the right direction. But the accessibility of the raw data only on the page level is a major drawback. The flat search is another one, in particular with regard to the fact that information necessary for the identification of relevant

⁶ The HTML is very much oriented towards the print version; therefore the main text chunk is the line, expressed as an HTML paragraph. As an undocumented feature you can address the TEI source of each page by replacing the html extension of the file with xml (e.g. http://www.dmgh.de/de/fs1/object/display/bsb00000451_00299.xml?sortIndex=030%3A040%3A0006%3A010%3A02%3A00&text=true&subSeriesTitle_str=&html=true&hl=false). This XML provides the display information of the page (e.g. lines) rather than a TEI encoding of the textual structure (e.g. paragraphs).

⁷ Personal note by project staff member Clemens Radl, 12.12.2012.

text entities obviously does exist and indices could be configured to calculate relevance by these text chunks (single charters, edited works) without too much effort. The comparison with a resource like *Perseus*,⁸ which integrates many important texts and tools for the classicist, may demonstrate some aims for which scholarly digital libraries like the *dMGH* could strive. The *dMGH* has already taken some steps in this direction and is working on more; with the *iMGH* semantic information enrichment has been put on the agenda. Work on these aspects is ongoing, and we hope that the *dMGH* will soon deliver more of the rich information from the printed editions in a structured way.

Bibliography

[dMGH 2006] Website of the dMGH: <<http://www.dmgh.de>>

[Müller 2003]Dedo-Alexander Müller: Rezension zu: *MonumentaGermaniaeHistorica (Hrsg.): Elektronische MonumentaGermaniaeHistorica 3 (MGH-3). Turnhout 2002*, in: H-Soz-u-Kult, 16.10.2003, <<http://hsozkult.geschichte.hu-berlin.de/rezensionen/2003-4-030>>.

[Sahle 2002] Patrick Sahle: *Die elektronischen MonumentaGermaniaeHistorica auf CD-ROM: eMGH. Lieferung 2. Turnhout 2000*, in: Zeitschrift für Bibliothekswesen und Bibliographie 49/5-6 (2002), S. 337-340, <http://www.klostermann.de/zeitsch/osw_495.htm> (preprint, available only via web.archive.org).

[Sahle 2008]Bernhard Assmann, Patrick Sahle: *Digital ist besser. Die MonumentaGermaniaeHistorica mit dendMGH auf dem Weg in die Zukunft – eine Momentaufnahme*. Schriften des Instituts für Dokumentologie und Editorik 1. Norderstedt: BoD 2008. ISBN 978-3-8370-2987-1. Online: [urn:nbn:de:hbz:38-23179](http://nbn:de:hbz:38-23179)

⁸ <<http://www.perseus.tufts.edu/>>