Johann Friedrich Blumenbach – Online

Gerhard Lauer and Heiko Weber

Johann Friedrich Blumenbach – Online

Blumenbach’s scientific oeuvre was a crucial part of the rise of modern biology, yet no complete edition of his books exists, and his numerous papers in a range of periodicals are difficult to find. Little work has been done on the history of Blumenbach’s natural history collections, although in his publications he frequently referred to specimens from these collections. A number of nineteenth-century biographical accounts exist (Marx 1840; Bendyshe 1865; Ehlers 1901), but no comprehensive biography of Blumenbach has so far been produced.

The State and University Library of Lower Saxony, colloquially known as Göttingen University Library, owns copies of almost every text published by Blumenbach and of nearly all the books he used for his research. Its archives contain the majority of Blumenbach’s unpublished papers; and the objects in the Academic Museum he curated as well as his private collection of human skulls, with a few exceptions, have been preserved and are present in today’s university collections (cf. Böker, this volume). In 2009, a committee of the Göttingen Academy of Sciences, the State and University Library of Lower Saxony, and the Georg-August-Universität Göttingen, chaired by Nicolaas Rupke, put forward to the Union of German Academies of Sciences and Humanities a proposal for a so-called Langzeitvorhaben (long-term project) to produce a digital edition of Blumenbach’s work and of his collections of natural history. The application was successful, and one year later “Johann Friedrich Blumenbach – Online” took off on what is anticipated to be a fifteen-year journey of application and innovation of the possibilities provided by the Internet in making Blumenbach and his oeuvre available to an international readership (Kerzel, Reich and Weber 2013).

A digital edition differs in two important respects from traditional edition projects. First, it can provide more than just text and include images of, and information about, physical objects which can readily be hyperlinked to a considerable extent of both breadth and depth. Especially in the history of the earth and life sciences it is of importance to have access not only to a text that describes and interprets a specific object but also to images of, and information about, the object. Blumenbach’s writings and his collections serve as an excellent example of this. Second, an online edition makes Blumenbach’s work accessible to digital tools
that are starting to supplement – not wholly supplant – traditional hermeneutic approaches. Needless to say, it is of key importance that digital editions maintain long-established standards of quality. When necessary, new rules have to be defined – for example, regarding citability and long-term preservation of electronic data.

While the basic tasks of producing electronic editions are the same as in pre-digital times, something has changed since the humanities have gone digital (Rosselli Del Turco 2016). A fast and steadily growing number of techniques and tools connect texts and objects in digital space. Key features of scholarly digital editing are the establishment of relations between data, and processing data, on a scale and with a complexity and precision previously unattainable. The same holds true for interoperability, the ability to share data in different computational environments, which enhances interactions between and within different scientific communities (Jannidis 2010).

Instead of presenting texts only as strings of letters and objects as plain images, “Blumenbach – Online” enriches them with metadata and standardized personal names and geographical names. This makes it possible to connect data with other editions, library catalogues, search engines, and even Wikipedia articles. Traditional editions and even present-day search options for digital texts can help us only little to understand the definition of a mammal at a certain point in history, the relevance of comets for natural history around 1800 or how the term “race” was used in French, English, and German sources of that time. By relating terms, names, and objects in semantic ways scholarly digital editions become nodes in machine-based networks of knowledge, the so-called semantic web, which is already part of the current Internet, even if only relatively few people are aware of it.

Digital editions in the history of science

The story of digital editions began in 1946, when Roberto Busa, an Italian Jesuit and an expert on St. Thomas Aquinas, planned the lemmatization of Thomas’s works, intending to produce a concordance of his monumental oeuvre and writings by a few related authors (Winter 1999). Busa’s guiding question was, What is the metaphysics of presence in Thomas Aquinas? Given the voluminosity of the writings, a complete lemmatization was possible only by using machines; and in 1949, Busa convinced Thomas J. Watson, the founder of IBM, to support the development of a computer-based tool for performing text searches. Much later, in 1967, the card-punching of Thomas Aquinas’s texts was completed, and in 1974 the Corpus Thomisticum was printed in fifty-six volumes. More recently, in 2005, a web-based version was made available online, while the following year the “Index Thomisticus Treebank,” a syntactic annotation of the entire edition, was begun.

What Busa did in the 1940s was pioneering work, considering that the first computers came on the market not before 1953, and FORTRAN was presented to the computing world as late as 1957, when programming languages became a tool
Gerhard Lauer and Heiko Weber

for data processing. Today, digital editing is a common practice not only for texts but for cultural heritage in a variety of forms; and digital editions – for example, of the complete musical works of Wolfgang Amadeus Mozart (http://dme.mozarteum.at) and of the paintings, sketches, and writings by Leonardo da Vinci (www.universalleonardo.org) – demonstrate the point.

In the history of science, “The Complete Work of Charles Darwin Online” started in 2002, led by John van Wyhe (http://darwin-online.org.uk). It is the first and probably the most extensive digital edition devoted to a scientist (Wyhe 2002, 2009) and includes Darwin’s complete record of writings – that is, his books, articles, published letters, manuscripts, private papers, and a rich array of supplementary material, such as reviews and critiques. Technical and editorial principles work hand in hand to offer easy access to “the whole Darwin.” The site’s very large number of hits indicates the worldwide interest in Darwin’s work and highlights the potential of digital editions for the history of science.

An edition such as “Darwin Online” addresses both experts and laypeople, and so does another digital edition by van Wyhe, which focuses on Darwin’s contemporary Alfred Russel Wallace (http://wallace-online.org/). Begun in 2012, it includes the published illustrations of Wallace’s animal specimens. Carl Linnaeus has become the subject of two digitization projects: “Linné On Line” by Roland Moberg and his co-editors makes selected research areas of Linnaeus available to a wider audience and informs about his contributions to pharmaceutics, botany and zoology, physics, mathematics, and ecology (http://www2.linnaeus.uu.se/online/index.html). The site “The Linnean Collections,” run by the Linnean Society of London, provides access to Linnaeus’s books, collections, correspondence, manuscripts, and his annotated private library, supplemented by a variety of non-Linnean manuscripts – for example, by Johann Reinhold Forster (http://linnean-online.org/). In a similar manner “The Newton Project,” edited by the historians of science Rob Iliffe and Scott Mandelbrote, publishes an online edition of all of Isaac Newton’s writings, including not only his scientific and mathematical texts but also the alchemical and religious ones, featuring the amendments Newton made to his own writings (www.newtonproject.ox.ac.uk/). The edition uses the platform-independent markup language XML, following the recommendations of the Text Encoding Initiative (TEI), to enable a faceted search for, among other things, references to the Bible, changes of hands, and other editorial details.1

These various editions have a common understanding of what an edition in the history of science should be like. Just as non-digital editions, digital ones are critical editions, but on a scale traditional editions couldn’t possibly attain (Sahle 2013). Digitization can and, in several instances, does present the entire oeuvre of a scientist, which means all published texts in their multiple historical editions. Unpublished manuscripts may be included. In some instances, the editions also provide a wide variety of supplementary material, such as contemporaneous reviews and other responses, diaries, library catalogues, museum records, and descriptions of specimens, not to mention their iconography. Furthermore, digital editions are able to offer side-by-side scans of the original pages and the transcribed and searchable text for critical philological comparisons. An advantage is
that, unlike old-time “critical editions,” they allow the reader to collate the transcribed text with the original. In the digital age, more than ever before, a “critical” perspective on texts is possible in the humanities and in the history of science in particular. Complex and stable bibliographical databases back up the edited texts.

For the purpose of global accessibility, each edition makes use of international encoding standards. The encoded documents must remain valid XML to allow the texts to be indexed automatically, to be displayed in any browser and to guarantee a precise correspondence between the image of original page and digital transcription. All documents are marked up with a tag set to provide an unambiguous matching between the texts and the bibliographical database. Only a few of the editions also provide the markup of semantic entities such as personal names, place names, specific periodicals or document types to make possible, for example, a combined search for illustrations and descriptions.

None of the current editions includes images of material objects from the working contexts of scientists, be it specimens or instruments, let alone experiments. These objects, described or discussed in the historical texts, have not been part of digital history of science editions so far. Thus these digitizations to a large extent follow in the footsteps of philological editions and make only partial use of the potential of digital editing, something that “Blumenbach – Online” is trying to remedy by integrating physical objects and semantics as part of the edition.

Editorial principles and standards of “Blumenbach – Online”

“Blumenbach – Online” provides digital material of high quality: the digital images of book pages and objects will meet or surpass the standards of the German Research Council (Deutsche Forschungsgemeinschaft) for the quality of images (300 dpi color images) and for the accuracy of character recognition of the electronic texts (Liebetruth 2015). Quality is also an issue with regard to the objects of Blumenbach’s collections. Until now, more than 4,000 of these have been identified in various present-day collections at Göttingen University. The project offers high-resolution photos of all objects. If the shape of an object requires this, these are supplemented by rotating and scalable images, as for example in the case of the items in Blumenbach’s collection of skulls (Reich, Böhme, and Numberger-Thuy 2012). For each object there will also be a data set with detailed measurements and metadata about its “biography,” documenting when, where, and by whom a specimen was collected, owned, sent, or brought to Göttingen. The interconnections between texts and collection items are made visible by hyperlinks: whenever Blumenbach in his texts mentions or provides an engraving of an object in his collections, a link is made available to the digital images and the data set of the present-day collection item, and vice versa. It will be possible to see which features of an object Blumenbach highlighted, neglected or – for whatever reason – did not consider worth mentioning.

The Blumenbach project is not limited to the more or less marketable part of his oeuvre – for example, his most famous works in their most popular edition.
The project is to pass on his complete work to the new digital media. In this respect, it will not only guarantee that in the digital cosmos of tomorrow there is as much “Blumenbach” available as in today’s analog world; as a digital repository, it additionally will outperform any analog library by holding copies of all his publications. Also the accessibility of Blumenbach’s scientific collections will increase as the site’s virtual museum of visual Blumenbachiana can be accessed everywhere and by everyone.

Our editing project encodes information in a machine-readable form that is not – or not directly – intelligible to humans. These instructions facilitate the application of advanced digital tools that automatically carry out analytical procedures on, and extract information from, texts and data sets of collection items. Crucial for this are, first, standardized “tags” to encode information and, second, the availability of digital authority databases for referencing information – for example, about people and places and especially bibliographical data. In general, the details of standards are pivotal to the sustainability of an edition such as “Blumenbach – Online.” The more widely adopted the standards are, the better the edition can interact with a digital environment of databases, catalogues, and browsers.

For tagging information, the Blumenbach project follows the guidelines of the Text Encoding Initiative (TEI), here in the version TEI-XML, Best Practice, Level 5. The recommendations are based on the markup language XML and have become a widely used quasi-standard for projects in the digital humanities. Our project uses TEI-tags to encode information about the structure, form, and content of Blumenbach’s texts – for example, headlines, chapters, illustrations, original footnotes, change of language (with identification of the languages), abbreviations (full term added), printing errors (corrected version added), book decoration, uncertain readings, names of people, institutions and places, dates, bibliographical references, quotations, and collection items (see Figure 2.1). More particularly, people, institutions, and places are identified by references to authority databases, and dates are alternatively given in standardized form (ISO 8601: yyyy-mm-dd).

All bibliographical references are identified, and if Blumenbach uses a quotation, the original text, enclosed in a tag, is added to allow human readers or digital devices to compare it to Blumenbach’s version.

Authority databases pose a problem. Machines may, sooner or later, master normal vocabulary – perhaps even in historical orthography; but they might have difficulty identifying and interpreting proper names, especially historical place names, or identical personal names that refer to different individuals. Whereas in most cases human readers are able to make sense of such ambiguous or opaque information, digital devices still need unambiguous identifications in a machine-readable, standardized form. Unfortunately, the requisite universal authority databases do not yet exist. The “Getty Thesaurus of Geographical Names” serves as an informal standard for place names all over the world, but although strong for places in the United States, it is less so for – say – Eastern Europe; and not very useful at all for historical place names, or bygone politico-geographical entities, such as the German principalities of Blumenbach’s day and age. Moreover, the
Gerhard Lauer and Heiko Weber

Getty Thesaurus is a commercial project that until recently charged fees. Presently, it can be used for free, but this may change again in years to come, causing sustainability problems for references to the Getty Thesaurus and similar products. For personal names one can use national databases that commonly exist within a library context, such as the German “Gemeinsame Normdatei” (GND) of the “Deutsche Nationalbibliothek.” They tend to be most useful for identifying people who were involved in writing, producing, and selling books, yet none of them has achieved the status of a worldwide standard. However, Blumenbach also exchanged information and specimens with collectors and amateur scientists, and refers to them in his publications. For the identification of their names, the existing databases are not always adequate.

Currently, this problem remains unsolved. In the case of place names, the Blumenbach project uses the Getty Thesaurus, if only because the datasets contain geo-coordinates, which allow for automated visualizations – for example, of the geographical extent of the provenance of Blumenbach’s collections or of the literature he cites. In order to identify people and publications, the Blumenbach project has created two databases of its own. The one for people, mentioned by Blumenbach or occurring in bibliographical Blumenbachiana (e.g., publishers, translators), offers basic biographical information and if possible the ID number in an authority database. The other, for publications, contains a full bibliographical record of every book or paper referenced by Blumenbach, and offers a link to a digitized version (if available).

Future developments

In a future Blumenbach portal, texts and collection items will be available in various ways and in different digital formats. Users can choose the original printed pages or, if they find the historical black-letter typeface awkward, a transcription (“full text”) in modern typography, and they have the additional option of retrieving extra information. A variety of synoptical arrangements will be possible, of original and transcribed versions, from one and the same or different publications/editions/translations, and of text with relevant collection items and data sets. All material will be searchable in different ways. Because it is difficult if not impossible to anticipate future developments of digital tools, the Blumenbach portal itself will not provide such devices. Instead, users will be able to download texts, images, and datasets in different digital standard formats and analyze the data with devices of their own choosing.

Digital editions such as “Blumenbach – Online” are also capable of responding to another innovative development: a change in the understanding of editions. Currently, editions are commonly understood as discrete entities, centered on single authors. However, in the history of science as well as in other fields of scholarly inquiry, practitioners are increasingly interested in specific layers of history, extended across many historical places – in the geography of scientific knowledge. Not just single editions are needed, but also editions as part of a wider corpus, customized to meet specific and thematic research interests. No edition
can fully anticipate future research questions. What can be done, however, is to support the integration of data into a corpus individually compiled by a researcher for analysis. Therefore it is increasingly imperative for projects like “Blumenbach – Online” to develop interoperable editions that can be integrated into such research corpora easily. The Blumenbach project has already made its texts available in larger corpora like the German Text Archive (Geyken and Gloning 2015). The Blumenbach portal will support corpus-based research in the history of sciences by offering a complete as well as a selective download of all data. There is no such thing as “the definitive edition” and this holds true also for “Johann Friedrich Blumenbach – Online.”

Notes
1 Other digitization projects of the works of remarkable figures in the history of science include “The Humboldt Digital Library” (www.avhumboldt.net/), the “Portal Alexander von Humboldt” (http://humboldt.hs-offenburg.de) and “Journals of the Lewis & Clark Expedition” (https://lewisandclarkjournals.unl.edu/).
2 This section is based on work by Claudia Kroke of the Blumenbach – Online project, whom we would like to thank for her contribution.
3 The Blumenbach project is making some parts of its editorial work – as soon as they become available – accessible on its homepage (www.blumenbach-online.de). An online version of the Blumenbach bibliography by Claudia Kroke (2010) includes links to the digitized versions (PDF) of all texts, and for most of them also a preliminary full text (HTML), searchable in browsers. Like the Darwin portal, “Blumenbach – Online” augments the edition of Blumenbach’s publications and specimens with supplementary material.

Bibliography


