

# How to Face the Crisis of Legitimacy: The Transfer and Further Development of Methods of Access from Printed to Digital/Digitised Editions

Dorothee Goetze, Tobias Tenhaef

Rheinische Friedrich-Wilhelms-Universität Bonn / Zentrum für Historische Friedensforschung Bonn

dgoetze@uni-bonn.de, ttenhaef@uni-bonn.de

## Abstract

All media provide media specific methods of access to information and therefore media change affects also these methods of access. But the change of media and hence access methods also raises the question of legitimacy of doing this, in terms of scholarly working as well as in terms of justification in the face of the funding general public financing research either directly by the government or indirectly by research funding organisations which are financed by taxes. The digitalization of a formerly printed critical edition is a case of media change. This paper will shortly describe the full text search as an example of methods of access added by the transformation of classical print editions into a digital Web site, and how far this method accommodates the user habits of the broader, even non-scholarly public. Using the example of APW digital, a full text digitalization of a cutting edge printed edition, online since July 2014, three possibilities to accommodate the users' expectations - and thus to gain legitimacy for the digitalization project - will be presented and discussed: the creation of new introductory material besides the original texts, the using of full text search and the access by metadata.

## 1 Acta Pacis Westphalicae (APW): critical print and online edition

The Congress of Westphalia which ended the Thirty Years War is a key moment in European history. It is considered a model of modern peace negotiations. There had never been such a major and significant secular diplomatic meeting before (Lanzinner 2013, 12). From 1643 to 1649 it provided diplomatic communication in an unprecedented way. It produced a broad variety of textual sources (letters, minutes, diaries and treaties) written in French, Italian, Swedish, Spanish, Latin and German which reflect the congress' activities. These materials have been edited in the series Acta Pacis Westphalicae (APW) since the 1960s (APW digital: [Über die Acta Pacis Westphalicae](#)). To increase the accessibility of the texts it has been decided to establish a digital edition in addition to the print version of the APW. For the transformation of the print into a digital edition, all volumes published in the APW series before 2008 (around 28,000 pages of text written in German, French, Swedish, Latin and Italian) were digitised by OCR. In July 2014 the digital edition of the APW materials went online ([APW digital](#)).

This change of media raises questions about legitimacy, questions which result from the characteristics of editions in general and those of print editions in particular and which have not been commented yet. Up to now discussions focussed mostly on the technical possibilities and limits of digital publications in itself, e.g. their digital

preservation (Ineichen/Flury-Dasen 2001). Below we comment on two challenging main issues that occurred during the transformation of the APW print to the APW online edition and describe the solutions which we found to face this crisis of legitimacy.

## 2 Editions as text-based scholarly communication

Edition principally could be defined as the attempt to reconstruct, establish and publish an authoritative version of a text (Sahle 2007, 64; Apollon et al. 2014, 88-89). To succeed with this task, broad methodological standards are developed and established by the editors who disclose and document the decisions made on the development of this authoritative text (Apollon et al. 2014, 88-89). According to our understanding, the main function of editing is to enable close reading of historical sources without too many logistical expenses.<sup>1</sup> The concept of close reading is first of all used in literary theory as method of access to texts. However, its implications of intensive text analysis describe the practice of historiographical analysis of sources which is the second step of historical criticism (after heuristics, first, and before interpretation, third) as it was formulated by Droysen (Droysen 1977).

By providing a verified text base for further research, editions offer certified supplies which guarantee scientific outcomes. Due to their certifying function, editions are in need of transparency about their critical proceeding and about the establishment of the text. This transparency is typically provided by a meta text, the critical apparatus. It includes critical annotations on the text generation, variations and different states of the text, annotations on the text contents, and indices, registers and regesta as instruments to access the information contained by the edited texts (Apollon et al. 2014, 16-17). Due to this, editing requires complex forms of presentation which combine the text as the base information and the meta information given on it.

Because of its impact on research and research results, editing becomes part of the scientific communication. Scientific communication is not only understood as exchange on science issues between peers of the scientific community. Scientific communication could be differentiated into science communication and scholarly communication. Communication between scientists and non-scholars is called "science communication". Closed communication processes within the scientific community are denominated "scholarly communication" (Hagenhoff et al. 2007, 4-7). Referring to their main functions and recipients, editions are prime examples for text-based scholarly communication as they "have been produced primarily for the peers of the editor, [...] experts in a particular domain" (Apollon et al. 2014, 93), even if they do not aim at producing new research results. In that tradition the APW can be classified as a typically paper-based critical scholarly edition which gains its legitimacy by the high critical standards and the transparency of decision making on the text. These standards are guaranteed by the publication of the rules on which the decisions on the textual reconstruction are based, the comparison of the master version of the text with up to five alternatives, the detection of text variations, down to word level, and the detection of the editor's construction work on the text according to the rules.

## 3 The transfer of publication medium as threat to legitimacy

As the decision was made to transform the APW into a digital edition, the described characteristics of an edition have been questioned by the change of publication medium. This means nothing less than threatening the legitimacy

---

<sup>1</sup> But there is still a debate ongoing about if edition can be estimated as a research result in itself.

of edition. In the current state of digital editing practice, the task is to transfer the typically complex representation of basic and meta information from paper-based to digital publication medium still guaranteeing scientific results. However, as Patrick Sahle pointed out, it is not done by only reproducing the image of the print edition in a digital environment (Sahle 2007, 68). That would restrict the digitalization of editions to being a copy of the text lacking the tools of critics which are the main and legitimating characteristics of a critical edition (Apollon et al. 2014, 64-65). This refers to another risk of changing publication medium for edition which Apollon and his co-authors allude to: digital "text presentations are often made by persons who are very knowledgeable in digital technology but often ignorant of the critical tradition" (Apollon et al. 2014, 17). That means not every opportunity technically possible complies with the conditions of an edition.

## 4 Facing a double crisis of legitimacy

To maintain its legitimacy, a digital or digitalised edition has to manage the balancing act of providing complex contents that still enable close reading and adapting its presentation to user habits without losing sight of its methodological standards and genuine objective. Numerous studies dealt with both Net based scholarly communication and reception habits of Internet users (e.g. Schwabe 2012). For the chosen context, two aspects have to be pointed out which we had to deal with by the transformation of APW to a digital edition: There is a discrepancy between edition inherent scholarly communication and science communication requested by digital publication media (see section 4.1). And to cope with the intended and anticipated expert reception as well as with the change in user expectations and habits appears to be a risk, as Apollon, Bélisle and Régnier state (Apollon et al. 2014, 1) (see section 4.2.). These two aspects show clearly that, in the digital era, the making of edition is facing a double crisis of legitimacy.

### 4.1 New user groups judging the legitimacy of editions

The change of media potentially breaks up the closed scholarly communication. That means that the change of publication media will shift the communicative contents. The exclusive academic communication will disintegrate and editing becomes part of science communication. The communication processes do not longer concentrate on the scientific community, but include a broader and heterogeneous public. As a consequence research results are presented, received and discussed within the scientific community. Non-professionals will also get access to them and both groups will discuss the results. Regarding editorial work, digital presentation and science communication will increase the number of recipients. A broader public will be aware of editions (Hagenhoff et al. 2007; Apollon et al. 2014, 85, 89, 94-95).

These new participants in the communication are not just recipients. They are judges about the legitimacy of Web contents. Thus they influence the digital presentation of contents and their adaptation to user habits. Besides corresponding legal, moral and ethical standards Web contents get their legitimacy due to meeting user interests, expectations and habits. The basis for users' legitimacy as judges is using the Web. Their judgements are reflected in their choice of Web contents. Normally, users pick those contents that match with their need of information and their user habits. Consequently, only the contents that are used own legitimacy.

Another source of the users' legitimacy as judges is financing science by tax money. There are two ways of public research funding, directly, that is state funding of universities and research institutions, or indirectly, that is financing by research funding organisations which are financed by tax money themselves but distribute their financial resources in own responsibility and without governmental interference. Such institutions are for example the Deutsche Forschungsgemeinschaft (DFG) or the Academies of Science and Humanities in Germany. This is the

case of the APW. The work on the APW edition was financed by the North Rhine-Westphalian Academy of Science and Humanities, its digitalization by the DFG. As tax money is public money, its recipients are obliged to spend it for projects and developments of public interest. Thus users influence Web contents in a fundamental way.

The digitalization of the APW aims on increasing accessibility of the texts which also implies a growing number of users – so one could pretend the occurrence of science communication was taken into account.<sup>2</sup> But in addition, the digital edition primarily holds on to the genuine scholarly peer group of the print edition. Nevertheless, the attempt was made to take the new (non-expert) users into account in different ways: General information about the set-up and the intention of the series APW is offered to give users some orientation on the edition's contents. Based on the contents, the digital edition is completed by additional offers, namely biopics (concise biographies) on the most important persons occurring in the sources presented, maps, which show locations, mentioned in the APW and a non-expert as well an expert chronological overview which gives information on the most important occasions of the Congress' proceedings. The expert version provides more detailed information and filter options for the user. These additional offers were composed out of the information given in the annotations in the APW print edition. The names of the persons and locations and the biographical information were detected manually.

The general criteria how to take new users into account were made by editorial board decisions relating to general knowledge about user habits and the set-up of other digital editions.

## 4.2 Desire for fast access to information

The mentioned discrepancy between intended peer and set-up of the edition and the public having access to the digital edition leads to one more challenge, which we had to manage. To understand the set-up of an edition and to be able to use it the reader needs a lot of information on how the edited text has been established. But most of the external recipients, that is non-scholarly public, are interested in a rapid procurement of information. This opposes close reading as the intended function of an edition which main task is the reconstruction of the authoritative form of a text out of its variations. Doing so, close reading becomes indispensable for the observation of the text construction and for deconstructing the composition of its meaning. Additionally, easy and rapid access to edited text opposes the self-conception of most traditional critical editors (Apollon et al. 2014, 95).

## 5 The full text search as method of fast access

To gain legitimacy for a digitalization project it is important to comply with the expected most users' wish for fast access to the information contained by the Web site. This can be achieved by a full text search, which is on the one hand the paradigmatic example of fast access and on the other hand a feature which only a digital edition basing on machine readable text can provide. To elucidate this, in the following sections some general remarks on entering of search requests into the search engine (section 5.1.) and presentation of search results (section 5.2.) are made. Section 5.3. finally shows how far these concepts are implemented in the case of APW digital. The full text search in general is a technique which gives in its most simple form a list of all instances of a given character string in a given textual corpus. Due to performance issues, the full text search is mostly realized as index based full text search. This means that the documents, which are technically spoken mere large character strings, are cut into shorter strings called "words", which are defined as strings of alphabetical signs between two non-alphabetical signs. These words

---

<sup>2</sup> Yet, any quantitative measuring has not been used, but its implementation is planned. An indication for increasing user numbers is the higher ranking of the APWdigital Web site in comparison to the former APW homepage.

are listed together with references to the documents in which they occur (Heyer et al. 2008, 59-62). The amount of words to be indexed can be restrained by the use of stop word lists (Gödert et al. 2012, 259-260). But these aspects and the form and functioning of searching algorithm and engine are issues which feature usually in the backend of a digital edition and are not part of the “human-machine interface”, where the question of usability will be decided (Nielsen, Loranger 2006, xvi). This “human-machine interface” consists in three essential aspects: first the way of giving to the search engine the search request, second the presentation of the list of search results and third the options of further processing the list of search results (Baeza-Yates, Ribeiro-Neto 2011, 5-7).

## 5.1 Entering the search request

In the past several years a *de facto* standard form has been developed for the input of the search request. This is the (in)famous “single search slot”, whose paradigmatic example is the Google-Homepage. The idea of this layout is to accommodate the user as much as possible, so that she or he can put in the requested word or phrase in a most simple way and the “magical” algorithms behind the scene make sure that the documents which the user wants to have are under the first five entries of the list of search results (Röhle 2010, 154-158; Stross 2008, 63). This is a rather comfortable way of using a search engine, especially if a first exploration of the corpus is sought, but it can result in a very vast amount of results, which are mostly ballast, and contradicts in some way the philosophy of the “good” search request which narrows down the amount of text which has to be searched already at the moment of formulating the request (Gaus 2003, 15, 261; Baeza-Yates, Ribeiro-Neto 2011, 4). Such a “good” request becomes more and more similar to a complex database query and the user can be helped in formulating it by the use of a multi slot form with several input slots and drop down menus of searchable categories and Boolean operators. However, a certain disadvantage of this request form consists in the fact that on one side many users perceive this form as somewhat “too technical” and also deterrent, and on the other side a “good”, complex and detailed search request often can be formulated only when the user already has a certain knowledge of structure and content of the corpus of the edited texts, and is insofar already a product of “close reading” of an edition. If both kinds of search request forms are provided by a Web site the single slot form is usually labelled as “Simple Search” and the multi slot form as “Advanced”, “Expert” or “Extended Search”. The latter labels reflect the above mentioned aspects of the complex search. The development of the search form layout has favoured the single slot form, so that this must be provided by every Web site, which shall be searchable (Nielsen, Loranger 2006, 140). Thus, the multi slot form is thus only a bonus, however a useful bonus.

## 5.2 Presentation of search results

The second aspect of searching is the presentation of the search results, which means how the list of search results is sorted by default and by which way the user can get a fast and comprehensive overview of the results. To begin with the latter, here too a *de facto* standard exists (Baeza-Yates, Ribeiro-Neto 2011, 29-32), although it is hard to imagine other possibilities of giving an overview of the search results. This standard is the presentation of the single findings of the search request, ideally highlighted visually, together with their context, which means a certain amount of text before and after the finding. The contextualization of the findings is a feature, which printed indices almost never provide. An even more compelling question is, how the findings can be recovered in the documents which contain them, because every list of search results can only be a representation of a subgroup of the documents of a corpus in kind of a dynamically generated link list, which refers to the finding containing documents only or to the places of findings directly. The solution is the dynamic generation of a second version of the documents which contain findings, in which the findings are marked, either visually or by not visual identification marks in case of direct pointing to the finding’s places or by both. This highlighting of findings is not yet used by all websites, but should become a standard of all websites which content can be controlled entirely by the owner of the website.

The other aspect of presentation of a list of search results is the sorting by default. It is obvious that the results, which are most relevant to the searcher, should be listed first, therefore an algorithmized notion of “relevance” is the favoured sorting criterion. The simplest way of algorithmizing “relevance” is to calculate the quotient of the number of findings in a document and the number of words of the document. Optionally and if it makes sense regarding the search word containing digital documents, the relevance algorithm can be enhanced, for example by counting the links set on or going out from search word containing Web sites in case of Web searching engines (Croft et al. 2010, 25). Though relevance is the most often used sorting criterion, others are imaginable, for example a sorting by date, if the documents are dated in some way.

The question of sorting the list of search results leads then to the third aspect of full text searching, the further processing of the list. Although every list must be sorted in one way, nevertheless it would be useful, if other reasonable options of sorting are given, like sorting by date or alphabet, in ascending and descending order. Another important feature of list processing is the use of filters that is in the end the doing of a “faceted search”. If there are properties that can be filtered, like dates, names or document categories, the amount of search results, which the user has to look at, can be reduced significantly. So the filtering of the search results is the complement on the precise but complex search request. The last option of list processing is the way of exporting the search results to the user, which principally can be done on the client side or can be supported by the server side.

### 5.3 The full text search in APW digital

Where does APW digital stand in this topography of full text search? APW digital fulfils the standard of full text search by providing a single slot search form. Due to the variations of the 17<sup>th</sup> century spelling, the searching algorithm allows a fuzzy search, which can be employed optionally. The list of search results consists of links to the documents, which contain the findings, and a presentation of the findings in their textual context. The single findings are not implemented as links, so that the documents can be reached only as a whole, but in these instances of the documents the findings are marked visually. A server sided exporting function of the list of search results is not featured, neither other options of sorting than sorting by relevance are given, but the user can employ a variety of filters to narrow down the amount of search results. These filters are the volumes of the original printed edition and the dates, categories and places of the documents. An example may illustrate this: If a user searches information about the important conferences held by the envoys of the prince electors in Lengerich, a village halfway between Münster and Osnabrück, she or he can type in the search string “lengerich”. The search engine finds 208 documents which contain the string “Lengerich”. Because we search for events located at Lengerich, the filter option “Place: Lengerich” should be chosen. The amount of findings is reduced to six. Because the conferences of the prince electors’ envoys are documented by minutes, a second filter “Unit: Protokoll” can be chosen, which lets pass to the user only these documents which are categorized as minutes, in German “Protokolle”. After these filtering procedures just four documents remain which are the sole conferences held by the prince electors’ envoys at Lengerich in the current corpus of APW digital. But if, for example, the reports on this conference, dispatched by other envoys at Münster and Osnabrück to their principals, are the documents interested in, the fuzzy search could be used. Now the search results amount to 223 documents. If it is known that the conference of Lengerich was held at 10 and 11 July 1645, the search results can be filtered by using the filters “Year: 1645” (89 documents), “Month: July” (41 documents) and “Unit: Korrespondenz”, which let pass the documents categorized as diplomatic correspondence. The 223 documents from the beginning are now reduced to 12, but also containing Swedish and French spellings of Lengerich like “Lengeriche” or “Lengeriez”.

The filtering function already uses a second method of access that potentially is provided by data storage systems that structure the contained texts by metadata mark-up. This method of access is the grouping together of elements corresponding to their metadata. In the case of APW digital, which employs the XML annotation standard

(Vonhoegen 2009), due to restraints by costs and time a mark-up was chosen which marks mainly the layout of the printed original, so that the information conveyed by the layout remains implicit, but the layout can be reproduced in the presentation. Anyhow, one sort of metadata is annotated explicitly: information about time and place. This information is contained in a TEI compliant mark-up, using the element `<div type="desc" subtype="dating">`, which contains the elements `<placeName>` and `<date>`, the latter providing Gregorian style normalized dates in the attribute `when="yyyy-mm-dd"`. Because almost all documents in the corpus of APW digital are properly dated and most of the documents bear information about the places where they were written, these properties can be accessed directly. The access to the places is carried out by the use of the filtering function which can be employed before or after sending a search request. A second access form provides access to the documents by date. Because dates are structured most effectively in form of a calendar, this layout was chosen as access form to the date containing metadata of the documents in APW digital. This is, besides the full text search, another method of fast and comfortable access, because now the user gets a collection of all the documents which are dated to the same day and were formerly distributed among several volumes of the printed edition. So the user can for example look at the diplomatic activity at 10 July 1645 and gets with one click the access to the minutes of the conference of Lengerich and diary entries reflecting on this diplomatic event. This feature hopefully will also facilitate the research in the corpus of the APW in the future.

## 6 Conclusion

By these means APW digital has tried to solve the questions of legitimacy, which resulted from the change of medium. Because APW digital is a (retro)digitalization of an already existing printed edition, the content of this edition and its “look and feel” as a specific sort of scholarly text were transferred to the new medium virtually unchanged. Nevertheless, to meet the needs of non-scholarly user groups some new offers for looking-up were made in form of biopics, chronograms and maps. But with the full text search and the presentation of informationally marked-up elements in groups defined by the classes of information, e. g. the presentation of dated documents in form of a calendar, methods of access are provided which are specific to the medium “computer” and henceforth “the Internet”. These methods of access enable a fast and comprehensive access to data and are insofar best suited to most users’ wishes for information, which can be looked up in a fast and simple way. Perhaps this sort of “mid-distant reading”, which doesn’t mean the computational analysis of large amounts of text in the sense of Franco Moretti’s “distant reading” (Moretti 2013), of the edited historical sources may lead to and facilitate the detailed and thus “close reading” of historical sources.

## References

APOLLON, Daniel, BÉLISLE, Claire and RÉGNIER, Philippe, 2014, *Digital critical editions*. Urbana, IL : University of Illinois Press. ISBN 9780252038402.

APW digital, [no date]. [Accessed 3 November 2014]. Available from: <http://apw.digitale-sammlungen.de/apw/static.html?lang=en>

BAEZA-YATES, Ricardo and RIBEIRO-NETO, Berthier, 2011, *Modern Information Retrieval. The Concepts and technology behind Search*. Harlow, London, New York : Pearson Education. ISBN 9780321416919.

CROFT, W. Bruce, METZLER, Donald and STROHMAN, Trevor, 2010, *Search Engines. Information Retrieval in Practice*. Boston, Columbus, Indianapolis et al. : Pearson Education. ISBN 9780131364899.

- DROYSEN, Johann Gustav, 1977, *Historik* (historisch-kritische Ausgabe von Peter Leyh und Walter Blanke), (Berlin 1857/1858) Stuttgart. ISBN 9783772806773.
- GAUS, Wilhelm, 2003, *Dokumentations- und Ordnungslehre. Theorie und Praxis des Information Retrieval*. Berlin, Heidelberg, New York : Springer. ISBN 3540669469.
- GÖDERT, Winfried, LEPSKY, Klaus and NAGELSCHMIDT, Matthias, 2012, *Informationerschließung und Automatisches Indexieren. Ein Lehr- und Arbeitsbuch*. Heidelberg, Berlin : Springer. ISBN 9783642235139. DOI 10.1007/978-3-642-23513-9.
- HAGENHOFF, Svenja, SEIDENFADEN, Lutz, ORTELBACH, Björn and SCHUMANN, Matthias, 2007, *Neue Formen der Wissenschaftskommunikation. Eine Fallstudienuntersuchung*. Göttingen : Universitätsverlag Göttingen. ISBN 9783938616758.
- HEYER, Gerhard, QUASTHOFF, Uwe and WITTIG, Thomas, 2008, *Text Mining: Wissensrohstoff Text. Konzepte, Algorithmen, Ergebnisse*. Herdecke, Bochum : W3L-Verlag. ISBN 9783937137308.
- INEICHEN, Andreas, FLURY-DASEN, Erich, 2001, *Geschichtswissenschaftliche Publikationen und Editionen ins Internet*, in: *Geschichte und Informatik = Histoire et Informatique* 12 (2001). Zürich : Chronos Verlag. ISBN 3034005598. ISSN 14205955.
- LANZINNER, Maximilian, 2013, *Einführung. Sicherheit in Vormoderne und Gegenwart. Perspektiven und Leitfragen*, in: LANZINNER, Maximilian (ed.), *Sicherheit in der Vormoderne und Gegenwart. Symposium der Nordrhein-Westfälischen Akademie der Wissenschaften und der Künste*. Paderborn : Schöningh, p. 11-23. ISBN 9783506777645.
- MORETTI, Franco, 2013, *Distant Reading*. London, Brooklyn : Verso. ISBN 9781781681121.
- NIELSEN, Jakob and LORANGER, Hoa, 2006, *Web Usability*. München, Boston, San Francisco et al. : Addison-Wesley. ISBN 9783827324481.
- RÖHLE, Theo, 2010, *Der Google-Komplex. Über Macht im Zeitalter des Internets*. Bielefeld : transcript Verlag. ISBN 9783837614787.
- SAHLE, Patrick, 2007, *Digitales Archiv und Digitale Edition. Anmerkungen zur Begriffsklärung*, in: STOLZ, Michael (ed.), *Literatur und Literaturwissenschaft auf dem Weg zu den Neuen Medien. Eine Standortbestimmung*. Zürich : Germanistik.ch, p. 64-84. ISBN 9783033013520.
- SCHWABE, Astrid, 2012, *Historisches Lernen im World Wide Web: Suchen, flanieren oder forschen? Fachdidaktisch-mediale Konzeption, praktische Umsetzung und empirische Evaluation der regionalhistorischen Website vimu.info*. Göttingen : V&R unipress. ISBN 9783899719963.
- STROSS, Randall, 2008, *Planet Google. How One Company is Transforming our Lives*. London : Atlantic Books. ISBN 9781843549802.
- VONHOEGEN, Helmut, 2009, *Einstieg in XML. Grundlagen, Praxis, Referenz*. Bonn : Galileo Press. ISBN 9783836213677.

## Biography of the authors

Dr **Dorothee Goetze** works as historian at the Department for Early Modern History at Bonn University. She has experience in editing early modern sources as she worked for the project APW from 2008 to 2011 and attended in its digitalization (2010-2012). Currently she deals with developing a concept for a born digital edition of the minutes of 17th century German eternal Imperial diet.

**Tobias Tenhaef** studied from 2006 to 2010 European History and History and Culture of the Native Peoples of the Americas at the University of Bonn finishing with a degree of M. A. From 2011 till 2014 he worked in a leading position in the digitalization project “APW digital” ([apw.digitale-sammlungen.de](http://apw.digitale-sammlungen.de)).