Martin Fechner

A Standardized Interface for Digital Scholarly Editions

[Vortrag, gehalten auf der DHd 2018 – Konferenz der Digital Humanities im deutschsprachigen Raum - unter dem Titel „Eine nachhaltige Präsentationsschicht für digitale Editionen“ ; Köln vom 26.02. – 02.03.2018]

3. April 2018

Persistent Identifier: urn:nbn:de:kobv:b4-opus4-33277

A Standardized Interface for Digital Scholarly Editions[*]

By Martin Fechner[†], 3 April 2018

Introduction

The demands on the creation of scholarly editions have increased over the last decade. Although the printed publication still represents the required result in most edition projects, at the same time digital editions are aimed for, which also require a web publication and the creation of sustainable research data. From the perspective of a digital edition, the core of the edition is therefore in valid data records that can be archived and used for research purposes.[1] The various forms of presentation are then generated from this research data in an optimal workflow, whereat the print is only one form of publication and the web presentation is another. Printed editions are usually based on the traditions of existing editions and thus make it possible to meet important research needs.[2]

The more weight given to the web publications in a digital edition, the greater the demands placed on the form of presentation. It is indisputable that a digital publication allows an approximation to the edited text that goes far beyond static printing.[3] Many digital editions offer additional benefits over their printed counterparts by integrating or linking external resources. At the same time, however, there are still requirements, such as the citation or presentation of apparatuses, which can look back on established traditions in print, but have no uniform equivalent in the web publication yet.

Another problem with web publications is the question of sustainability and long-term archiving.[4] In the area of research data, there are now TEI-XML formats based on open standards that are easy to store and retrieve, but in the area of web publications the production of long-term availability of the publication, including its functionalities, is still a challenge. Therefore, a lot of energy is currently being invested in the establishment of data curation techniques, which should keep the research presentations of even expired projects permanently up to date so that they can be presented in current and future technical environments.[5] This is especially problematic if the form of publication, i. e. presentation and functionalities, is part of the edition itself.[6] After all, there is no differentiation between

---

[*] This paper was originally presented in German at the DHd 2018 conference in cologne under the title “Eine nachhaltige Präsentationsschicht für digitale Editionen”.

[†] Berlin-Brandenburg Academy of Sciences and Humanitites

[1] Archiving sources and high readability are not the only, but the most important goals of edition projects (e.g. Eggert, 2016). Scholarly digital editions are able to support this by their reusability of research data (e.g. Ralle, 2016). Sahle (2016) defines scholarly editions as: “A scholarly edition is the critical representation of historic documents.”

[2] There should be a high readability of the whole edited text, optically separated in apparatuses are editorial comments, text genetic comments, and variant readings. The Citation is established by page and line numberings.

[3] A digital scholarly edition offers full text searches, faceted searches, dynamical indexes, intertextual links, integration of external resources, and links to other editions (e.g. Sahle, 2016).

[4] By now, long-term preservation of research data is possible through the use of standard formats and archival storage in data repositories. But, there’s no great trust in reliability of digital scholarly editions (e.g. Pierazzo, 2015, p. 169).

[5] The considerations of Pempe (2012, p. 141) about the unsolved tasks for achieving sustainability of digital scholarly editions are mainly the same as today.

[6] Ralle (2016) shows that the form of presentation in printed books is part of the edition and can’t be separated from it. More information about the demands and changeability of digital scholarly editions is found in

presentation and research data in the case of web publications as in the book. If only the research data will be archived in the long term, the presentation will eventually become obsolete and the web publication will disappear.

A solution in the field of data curation requires websites must be permanently adapted to the latest technical solutions. However, this effort grows in line with the number of published web publications. If the resources and interfaces associated with the publications have to be maintained as well, the effort increases polynomially. In order to counteract this, the introduction of standards in the area of research data and interfaces is currently being implemented.

Standards and Interfaces

Fortunately, standards are currently being established at various levels that facilitate data exchange between different projects and data archiving.[7] However, such formats only really add value to the use of digital editions if they are supported by the offer of technical interfaces.[8] As an example, only services based on the *Gemeinsame Normdatei* (GND) of the German National Library, which enable the exchange and comparison of personal data in particular, should be mentioned here.[9] For many purposes, however, such interfaces are still deficient, for example, when cross-publication research and intertextual links are to be made possible.[10]

Hopes are high that this gap will be bridged by the *Semantic Web*, which will be used to analyse research data independently of the project in order to enable comprehensive searches and links. A prerequisite for the use of *Semantic Web* technology is the introduction of ontologies, which ultimately only represent another data model into which the data of the projects must be transferred. The focus of the Semantic Web is therefore on information processing and retrieval, while the remaining challenges remain unsolved, such as citation and sustainable availability, which are often linked to the concrete presentation of the web publication. An integral part of scholarly editions is what one could call presentation and functional processing of the research data from a digital research perspective. Editions guidelines and concepts of digital scholarly editions often explain and justify the form of presentation in detail. The TEIPUBLISHER is taking the first steps towards setting up a presentation interface for TEI documents, focusing primarily on the presentation of individual documents.[11]

Pierazzo (2015, p. 127). So, the presentation of editions could be seen as interfaces (e.g. Porter, 2016; Turska, 2016).

[7] For research data and text data in editions the guidelines of the Text Encoding Initiative (TEI, see http://www.tei-c.org/index.xml) is widely accepted as standard. But, these guidelines can only be an first step of an overall standardization process (e.g. Holmes, 2017). Other used standards and systems are for metadata the LMER format (Steinke, 2005), for citations of classical texts the Canonical Text Service (CTS, see http://cite-architecture.github.io/cts/), for persons the Gemeinsame Normdatei (GND) of the German National Library, or for places the common id system of http://www.geonames.org/. For documentation purposes are also standard formats like DITA, or docbook used.

[8] Recent discussion has emphasized that digital scholarly editions can be regarded as interfaces (for examples see the abstracts of the Digital Scholarly Editions as Interfaces conference, University of Graz, https://informationsmodellierung.uni-graz.at/de/veranstaltungen/archiv/digital-scholarly-editions-as-interfaces/; Bleier 2018)

[9] With the GND it is possible to search automatically for data of a person from other resources and to embed the results (for examples see the person index in http://edition-humboldt.de/).

[10] In the project http://correspsearch.net/ is such an interface developed to interconnect the data of different scholarly editions of letters.

[11] See TEIPUBLISHER. THE INSTANT PUBLISHING TOOLBOX at http://teipublisher.com. The data format of the TEIPUBLISHER is based on the ODD specification (One Document Does it all) that allows one to customize the

The introduction of standards and uniform data formats is intended to facilitate the exchange of research data between different projects on the one hand, and to provide the basis for sustainable archiving on the other. It is only by linking research data to an interface that the data is prepared and made available in the desired form, whereat the actual exchange can take place. The digital scholarly edition is increasingly understood as a developed independent publication.[12] As such, it relies on more than just data exchange with other research projects. Digital scholarly editions as publications need sustainable solutions for dealing with a wide variety of digital objects. This includes filters, searches, navigation elements, display of images, text and integration of interfaces. One can say that an digital scholarly edition is *the presentation as functional preparation of research data*. A clearly defined interface that translates the research data into a sustainable functional form of presentation is currently missing.[13]

Proposal

A proposal is now being presented here to address the sustainability of web publications. For this reason, a system is being developed that links the data with the presentation layer, i. e. represents a fully-fledged web publication. Besides the research data the focus here lies on the presentation as functional data presentation (see Fig. 1). However, this general system is to be controlled by means of a project-specific configuration via a standardized interface, so that in order to maintain the web publication of a special edition, it is only necessary to achieve sustainability of the specific configuration in addition to sustainability of the respective research data. If such an interface is developed and accepted as standard, specific web publications including the corresponding functions can be reproduced from the research data using various technical systems. Digital scholarly editions can still technically very different from each other at the moment,[14] so the proposal made here and the following example show where there are already possibilities for standardization and how they could look like.

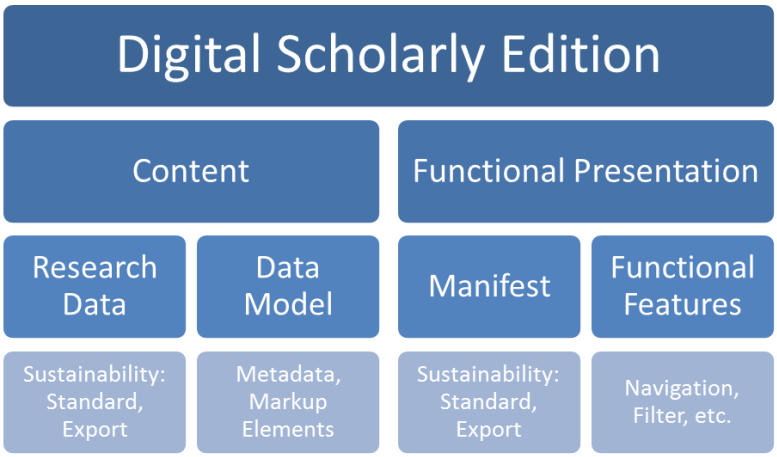

Fig. 1: Parts of a digital scholarly edition

TEI XML, for further information see the documentation of the TEIPUBLISHER, Meier (2016), and Turska (2016).

[12] In the contrary, the printed edition is seen as a reduction of the digital edition (e.g. Sahle, 2016).

[13] Examples of poor usability because of unsustainable interfaces are given by Robinson (2016). Pierazzo (2015, p. 162) sees it as a disadvantage that digital scholarly edition are using different user interfaces. But she considers a future harmonisation to be probable.

[14] The most digital scholarly editions are produced in edition projects with very specific characteristics (e.g. Robinson, 2016).

Goals of a Standardized Interface

The system should be able to encode all the information necessary to display a web publication of a digital scholarly edition.[15] Of course, this includes the dynamic presentation of the texts, as well as the dynamic arrangement of individual text passages. Furthermore, a synchronous representation of texts is useful, as well as the linking of image and text.[16]

The new standard can be designed according to the successful IIIF standard. This was introduced to overcome the different distributed systems for displaying images in scientific web presentations (Cramer, 2011). For digital scholarly editions we are today in the same place as described above.[17]

The IIIF standard provides for a division of the server and client structure. This means there are the IIIF servers, which can be accessed via a well-defined interface and deliver the images in a certain resolution. There are also the IIIF viewers each offering a presentation interface with their own functions, in which the images can be presented in a certain frame. If the accesses are guaranteed, each viewer can display the images of each server. The exchange format is a manifest file containing the image collections, URLs and metadata. Based on this principle, a standardized interface for digital editions can also be defined. In the interests of long-term archiving, it seems sensible to establish compatibility with existing standards.[18]

Proposal for a Manifest

When implementing a system that is able to present a digital edition, numerous features have to be implemented. Thus, the different data types, such as metadata, texts and images, have to be handled. Necessary functionalities of such a web presentation are navigation possibilities through the objects of the edition, in particular through texts and registers, as well as export functions, search and filter options, citation possibilities and the connection of other standard interfaces and external resources. A synchronous representation of related objects, e. g. text and translation, or text and digitized texts appears to be useful. It is also possible to have different views of the contents themselves, as well as individual arrangements of texts or parts of texts. All in all, the interaction of the digital objects of the edition must be defined by the manifest. It is suggested here that the following components should be considered in the manifest.

Metadata

Of course, a manifest should contain the necessary metadata of the digital scholarly edition. It can be discussed which information is necessary for a digital edition and which information is considered desirable.[19] In particular, information such as publisher, title and imprint should be recorded in such a file. In addition to the required entries and desirable standard fields, it is

[15] Shillingsburg (2016, p. 154) argues for not "building theme parks with high entertainment value" but to produce reliable high quality representations of the sources.

[16] This doesn't mean the presentation of images within the text flow. That should be possible anyway. Ralle (2016, p. 154) lists the following features: full text search, faceted search, navigation menus, different text representations, reliable citation, and versioning. She also brings a deposit of a compulsory version to a central library into the discussion to gain long-term preservation. Another detailed list of requirements of digital scholarly editions can be found at Sahle (2014).

[17] Witt (2016) shows the redundancy of the web presentation of digital scientific editions, proposing a similar approach as here to overcome the current situation. He shows how creating a semantic data model can lead to uniform text presentations. However, the flexibility to create such a model is part of the individual editions and differs between them.

[18] There are existing metadata standards for long-term preservation like LMER (see Steinke, 2005).

[19] A compatibility with existing standards like METS or LMER should be established.

conceivable that, depending on the edition, additional information can be added in self-defined fields.

Navigation and Access to Documents and Indexes

In the presentation interface, it must be possible to navigate to the various parts of the edition, the data objects, and to display and sort them in the form of a list.[20] The different types of data objects can be defined for this purpose, such as different text types or registers.

Digital scholarly editions differ in their focal points. This means that navigation and filter options cannot be defined in a sound way across all editions. For this reason, for each edition and its research data it is necessary to explicitly define what kind of filters are useful and possible for which objects. If this information is defined in a manifest, it is then the task of an edition viewer to implement navigation and faceted filtering based on the definitions. Possible are hierarchical filter forms, where the current selection allows further filters, numerical filters or filters with discrete values. It should also be defined whether it makes sense for the respective digital scholarly edition and should therefore be possible to allow the filter values to be combined or mutually exclusive.[21]

Presentation of Documents and Indexes

The manifest file must also define how the research data for a single object is to be presented. This can be done, for example, by referring to interfaces that provide the objects in one or more defined formats.[22]

It is also desirable that additional navigation options are available for the presentation of individual objects, such as documents. So, the internal structure of a document - like individual chapters or sections, but also specific relevant places - can be accessed. To make the internal structure accessible via a definition in the manifest can also open up citation possibilities and make it possible to display passages individually. If you also define how relations to and from other documents are recorded in the data, the viewer can make these relations visible as navigation elements, such as link lists. By defining the internal structure of the object types and the relations to other types, you can also combine objects with distributed information from their parts and display them as an overall document.

If the manifest file also records the relationships between object parts and other object parts, this information can be used for a synoptic representation of different documents, e. g. for the representation of text and translation or text and digitization.

Hierarchy and scientific material

A developed digital scholarly edition contains not only the source material but also accompanying scientific material, such as introductions or scientific essays. The accompanying material can be considered as another object type of the edition. In contrast to documents, access via lists and filters will not be sufficient. There will therefore have to be an additional hierarchy of objects, which is also reflected in the navigation. It should be possible to record in the manifest file what accompanying material is available and in what kind of form this material, as well as the other documents and registers are arranged hierarchically.

---

[20] There are several ways to display an overview of the objects: to show a browsable list, a object cloud, or another graph with navigation links.

[21] Example for a combined filter are searches for objects with references to „Berlin AND Munich", or "Berlin OR Munich", example for a mutually exclusive filter is a search for object with a reference only to "Berlin" or "Munich".

[22] How single XML-TEI documents could be presented in a standardized way shows the TEIPUBLISHER (http://teipublisher.com).

External Resources, Interfaces and Citation

In the manifest file, you can also define which external resources the digital scholarly edition refers to and in what form, i. e. how identifiers of standards are stored. In this way, external resources can also be integrated and linked, which greatly increases the value of a digital edition. In the short term, you have to include resources whose availability is not yet assured.[23] However, if the availability of external resources is not stable, the added value will be lost again. In order to counteract this, it is necessary in the long term that only trustworthy resources are integrated into the edition. The integration is specifically conceivable as a linking or embedding of resources via identifiers.

The production of citation possibilities is one of the current challenges of digital scholarly editions. In order for the presentation system to meet the current requirements of a complete citation system, it must be defined for each object type which parts can be cited and in which form this has to be done. The defined hierarchical structure can also be included. The first established forms for the production of citations in digital space are already available. In addition, systems are currently being developed to systematically cite sections of text in editions.[24]

Elements	Properties to be defined	Has relation to
Metadata	title, imprint, license	
Hierarchy	structure, links to objects and object types routing patterns	Types of objects
Types of objects	type name, type ID context: access to objects by URL or path root: path and conditions (XPath expression) object ID (XPath or XQuery expression) object title (XQuery function) thumbnail	Overall view, filter, relations, interfaces, citation
Overall view	context: related object type ordering functions	
Filter	filter name, filter ID filter type (single, union, intersect, numeric) context: related object type root: path and conditions (XPath) value: labelling function (XPath or XQuery) dependencies to other filters	
Presentation	processing scripts or systems	
Parts of objects, structure of objects	name, ID of structure type root: path and conditions (XPath or URL) ID: (XPath or XQuery expression) title (XQuery function) thumbnail	
Relations	name, ID of relation type subject: ID of object type predicate: URL to ontology object: ID of object type	Object types, parts of objects
Interfaces	name, id of interface access to interchangeable objects (URL)	
Citation	context: related object type citeable objects root (XPath) citation pattern (URL pattern)	

Table 1: Overview of different elements and properties that can be defined by a manifest

[23] Holmes (2017) shows how to integrate external resources in a sustainable way by producing a stand-alone edition.

[24] The Canonical Text Service for example defines a server client based protocol to retrieve text parts cited by a canonical URN (http://cite-architecture.github.io/cts/).

This proposal already includes many requirements of a digital scholarly edition. For an overview about the definitions see in Table 1. With a manifest containing these definitions together and with access to the research data itself, it should be possible for a presentation interface to present a digital edition with the desired functionalities. To what extent special functionalities can also be mapped with this proposal must result in further discussion. What such a manifest file cannot and should not perform is to specify the exact type of presentation. This can only be done through the presentation interface. These could then be further adapted for the presentation of a certain digital edition and provided with a concrete design. Nevertheless, it should be possible to use the manifest file and the research data to restore all desired functionalities of a digital edition regardless of the actual design.

Prototype „ediarum.WEB“

In the EDIARUM research project, we are currently working on the creation of a component that will implement the presented concept and meet the requirements of the presentation.

EDIARUM was developed within the framework of the digital humanities initiative TELOTA at the Berlin-Brandenburg Academy of Sciences and Humanities (BBAW) - in close cooperation and coordination with the scientific research projects of the BBAW. EDIARUM enables manuscripts and printed texts to be transcribed, distinguished, commented on and linked to register entries and other resources in a user-friendly way according to the guidelines of the Text Encoding Initiative (TEI). In addition, EDIARUM supports modern, digital working methods such as decentralized processing of shared data sets or the timely online presentation of selected processing steps. The research software is based on proven technologies and programs and has a modular structure so that it can be easily integrated into existing workflows. In addition, it is very adaptable - an advantage that is very important for scholarly edition projects, despite all the similarities of these projects. In the current further development of EDIARUM, individual modules are developed which can also be used individually, but only develop their full potential when combined. EDIARUM.DB is available for common data management, EDIARUM.EDIT for the creation of research data, EDIARUM.WEB for web presentations, EDIARUM.PDF for printing. The individual components have already been published or the publication is planned.

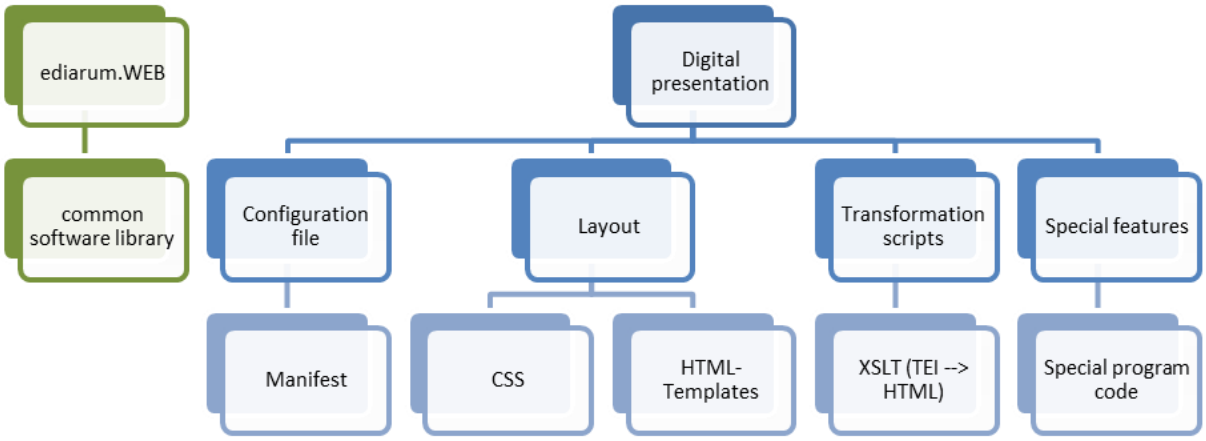

Fig. 2: Schematic figure of the components of the prototype

The concept described here is used in the EDIARUM.WEB module. The EDIARUM.WEB named software solution only contains a program library, which provides the functionalities for display. The core of the specific representation of a digital scholarly edition is formed by a manifest file according to the above-mentioned concept. With this prototype it is already possible to create a website for a digital scholarly edition with the help of a configuration file, which is a precursor to a manifest file, and with a few adjustments, which mainly concern the layout. By using the prototype for several editions, the individual functionalities and configuration options are tested and improved. A schematic function of the prototype is shown in Fig. 2. The aim of the development of the prototype is to provide web pages for existing scholarly edition projects of the BBAW, to continue to implement the functionalities according to the concept described here, as well as to gain experience on how the concept described here has to be further developed and how a manifest file has to look like in order to meet the requirements of a digital scholarly edition. Finally, the prototype should be available as a viewer, which only needs a corresponding manifest file for presentation and access to the data via a corresponding server infrastructure. In order to create presentation options in the edition's own layout, there is a core component that provides all functionalities and a project-specific component that can be adapted in the layout.

In the current state of the development of EDIARUM.WEB it is already possible to create a complete online presence with numerous features.[25] For this purpose, a prototype was developed in the context of different digital editions, which is why the prototype has a high connectivity for different projects. The integration of further projects is currently being tested.

In order to create a website, adjustments for the respective scholarly edition project still have to be made in four different areas (see Fig. 2). First, it requires a central configuration file that is based on the above-mentioned concept, but does not yet contain all components. Especially the connections to external resources and the integration of a citation system are still missing. This configuration file can later serve as a manifest file for the exchange and creation of sustainability. Furthermore, there are project-specific layout adjustments, which mainly concern the CSS and the structure of individual HTML pages themselves. However, these are independent of the research data and rather represent an optical distinction of a scholarly edition project from other digital scholarly editions. In a project-independent viewer, these adjustments would be omitted. Finally, individual transformation scripts have to be included defining how the research data (in this case TEI-XML) is to be integrated into the web page (as HTML) and there is some project-specific program code, which is needed for the correct representation. The project-specific code can be further reduced in future development by introducing conventions, i. e. a further development of the concept presented here. If an individual project has special requirements outside of the defined ones, it is still possible to customize the web presentation with your own program code. When using uniform data formats (here in the project, tests are carried out with DTA-Bf or the corresponding EDIARUM adaptation), a uniform code can also be used for the data-specific transformation.

A high degree of flexibility is achieved by the interaction of the individual components and the ability to adapt them individually. An individual appearance can be created for each project. For projects without individual layout requirements that use only supported standard formats for their research data, access to the research data itself and the corresponding manifest file will be sufficient to provide a digital scholarly edition.

[25] Currently the web presence of the manuscript description catalogue “Commentaria in Aristotelem Graeca et Byzantina” (CAGB, see https://cagb-db.bbaw.de/) will be extended by a scholarly edition part and is rebuild with EDIARUM.WEB.

Conclusion

This article proposes the development of a new standard to improve the sustainability of digital scholarly editions. The proposal made here and the exemplary implementation in the EDIARUM project showed what is already possible with the integration of existing standards. For the discussion that follows, however, it requires the cooperation of the respective projects. Therefore, this is only a starting point for further discussions where the participation of different scholarly edition projects is desired. With the formation of a working group, the various requirements could be discussed in very concrete terms and the necessary standards could be developed and defined. For easy use and acceptance, parallel software solutions should be developed that support the standard and thus enable the presentation of digital editions. To this purpose, the EDIARUM project component listed here will be further developed.

References

Bleier R. Bürgermeister M. Klug H. Neuber F. Schneider G. (eds.) (2018, in print). Digital Scholarly Editions as Interfaces. Schriften des Instituts für Dokumentologie und Editorik 12. Norderstedt: Books on Demand.

Cramer T. (2011). The International Image Interoperability Framework (IIIF): Laying the Foundation for Common Services, Integrated Resources and a Marketplace of Tools for Scholars Worldwide. https://www.cni.org/topics/information-access-retrieval/international-image-interoperability-framework (accessed 7 March 2018).

Eggert P. (2016). The reader-oriented scholarly edition. *Digital Scholarship in the Humanities*, 31(4):797–810.

Holmes M. (2017). Whatever happened to interchange? *Digital Scholarship in the Humanities*, 32(suppl_1):i63–i68.

Meier W. Turska M. (2016). TEI Processing Model Toolbox: Power To The Editor. *Digital Humanities 2016: Conference Abstracts*, 936.

Pempe W. (2012). Geisteswissenschaften. In Neuroth H (ed.), *Langzeitarchivierung von Forschungsdaten: Eine Bestandsaufnahme Version 1.0, Stand 2012*, Boizenburg: Verlag Werner Hülsbusch, pp. 137-159.

Pierazzo E. (2015). Digital Scholarly Editing: Theories, Models and Methods. Abingdon: Routledge.

Porter D. (2016). "What is an edition anyway?" My Keynote for the Digital Scholarly Editions as Interfaces conference, University of Graz. Blogpost. *Dot Porter Digital*. http://www.dotporterdigital.org/what-is-an-edition-anyway-my-keynote-for-the-digital-scholarly-editions-as-interfaces-conference-university-of-graz/ (accessed 7 March 2018).

Ralle I. (2016). Maschinenlesbar – menschenlesbar. Über die grundlegende Ausrichtung der Edition. *Editio*, 30(1):144-156.

Robinson, P. (2016). Project-based digital humanities and social, digital, and scholarly editions. *Digital Scholarship in the Humanities*, 31(4):875–889.

Sahle P. (2014). Criteria for Reviewing Scholarly Digital Editions, version 1.1. https://www.i-d-e.de/publikationen/weitereschriften/kriterien-version-1-1/ (accessed 7 March 2018).

Sahle P. (2016). What is a Scholarly Digital Edition? In Driscoll M. J. Pierazzo E. (eds.), Digital Scholarly Editing: Theories and Practices, Cambridge: Open Book Publishers, pp. 19-40.

Shillingsburg P. (2016). Reliable social scholarly editing, Digital Scholarship in the Humanities, 31(4):890–897.

Steinke T. (2005). LMER: long-term preservation for electronic resources, version 1.2, Leipzig: Dt. Bibliothek. http://www.dnb.de/DE/Standardisierung/LMER/lmer_node.html (accessed 7 March 2018).

Turska M. Cummings J. Rahtz S. (2016). Challenging the Myth of Presentation in Digital Editions. Journal of the Text Encoding Initiative, (9). http://journals.openedition.org/jtei/1453 (accessed 7 March 2018).

Witt J. C. (2016). *Digital Scholarly Editions as API Consuming Applications.* In Book of Abstracts of the Digital Scholarly Editions as Interfaces conference, University of Graz. https://static.uni-graz.at/fileadmin/gewi-zentren/Informationsmodellierung/PDF/dse-interfaces_BoA21092016.pdf (accessed 7 March 2018).