Digitale Edition in Österreich
Digital Scholarly Edition in Austria

———————————————

herausgegeben von | edited by

Roman Bleier, Helmut W. Klug

2023

Digitale Parallelfassung der gedruckten Publikation zur Archivierung im Kölner Universitäts-Publikations-Server (KUPS). Stand 29. April 2023.

"Don't Worry, We're also Doing a Book!" — A Hybrid Edition of the Correspondence of Bernhard and Hieronymus Pez OSB

Daniel Schopper, Thomas Wallnig, Victor Wang[1]

Abstract

Although digital methods have become an undeniable requisite for scholarly editing, only few projects attempt to standardize data and to envision reusable modules. The learned correspondence of the Austrian Benedictine historians Bernhard and Hieronymus Pez has been object of scholarly efforts for more than fifteen years, transitioning from a traditional print edition to a digital paradigm. The article outlines in detail the features of a hybrid solution for the Pez correspondence edition: a pilot workflow that allows for print and digital publication in the series *Quelleneditionen des Instituts für Österreichische Geschichtsforschung*, while envisioning long-time data storage in the ARCHE repository of the Austrian Academy of Sciences.

Zusammenfassung

Obwohl digitale Methoden heute unerlässlich für wissenschaftliches Edieren sind, bemühen sich doch wenige Projekte um Datenstandardisierung und Wiederverwendbarkeit von Modulen. Die gelehrte Korrespondenz der österreichischen Benediktinerhistoriker Bernhard und Hieronymus Pez ist seit mehr als fünfzehn Jahren Gegenstand editorischer Bemühungen, die eine Transformation von einer klassischen Printedition zu einem digitalen Paradigma mit sich gebracht haben. Der Artikel beschreibt im Detail die Beschaffenheit einer hybriden Lösung für die Pez-Korrespondenz-Edition: einen exemplarischen Arbeitsablauf, der eine gedruckte und digitale Edition in der Editionsreihe *Quelleneditionen des Instituts für Österreichische Geschichtsforschung* ermöglichen soll, während zugleich die Langzeitspeicherung der Daten im Repositorium ARCHE der Österreichischen Akademie der Wissenschaften gewährleistet wird.

---

1 Introduction: on hybrid scholarly editing

> In many cases […] print is still the medium of choice for the publication of the fruits of scholarly endeavour, and digital editions have not yet been accepted among the scholarly community; they well may be used behind the curtains, but when it comes to citations and referencing, it is the printed edition that takes centre stage, thereby depriving the producer of the digital edition of legitimate acknowledgement. (Driscoll and Pierazzo 2016, 15).

Expressed in 2016 in one of the groundbreaking publications on critical scholarly editing, this assertion has lost none of its validity to this day. There may in fact even be an inverse trend, at least concerning the historical disciplines: while digital humanities topics and methods have gradually started entering the scholarly mainstream, the multitude and potentially confusingly broad range of related approaches—from topic modelling to historical GIS and network analysis—have shifted some attention away from digital editing, or at least have not contributed to convincing those hitherto reluctant to engage with it.

Hybrid scholarly editions—that is, editions published digitally *and* in print—hold a particularly difficult position here. To those arguing the cause of the printed book, any effort invested in the digital aspect of an edition may seem a waste of resources, while the opposite may be the case for those advocating purely digital editing. When we argue for the benefits of a hybrid edition in the following, we do so in the full awareness that these benefits depend greatly on the nature of the published source, the target audience and, as we will show, the specific configuration of institutions involved. By putting hybrid editions on the agenda and devoting a separate working group (AG10) to them, the consortium of the Austrian Hochschulraumstrukturmittel project *Kompetenznetzwerk Digitale Edition* (KONDE) has clearly acknowledged the relevance of this field. Moreover, as a chapter of a comprehensive book on critical scholarly editing, this text represents one aspect of a broader context in which digital scholarly editing itself is thoroughly discussed.

We can therefore assume that the reader is familiar with the various types of text encoding as well as with infrastructural matters, with the types of scholarly questions to be answered with the help of digital editions, and with the various issues connected to the creation and maintenance of user interfaces. What we would like to demonstrate on the following pages is how a hybrid edition can be the best answer to specific needs in an appropriate case. Pointing out these needs and sketching out the possible answers is a goal in itself, since other scholars may find different (and perhaps better) solutions to the same problems. We will therefore attempt to frame these problems, explain our choices, and describe the process—but at the same time we would hope not to be understood as unduly advertising a product or preaching a

gospel. This is particularly important when it comes to the prominent participation of a publishing company: we are not seeking publicity, rather trying to work out a solution that assigns each actor a specific role.

One may rightly argue, however, that not all hybrid editions require a publishing company as a key partner in the design of the technical editing process. Some of the respective KONDE partners have found different solutions: the Werner Kofler hybrid edition (Straub and Dürr 2019) is built around an existing copyrighted print edition for the broader public, which it supplements by way of an extensive commentary. Similar approaches in terms of anthologies have been applied to the works of Peter Handke (Kastberger 2015) and Ernst Jandl (Hannesschläger 2015) to respect the copyright protection of initial prints valid for 70 years after an author's death. Conversely, in the case of Robert Musil (Bosse, Boelderl, and Fanta 2016), the entire œuvre along with its commentary can be part of the hybrid design.

Compared to the Pez letter edition (Wallnig, and Stockinger 2010; Wallnig, et al. 2015), which is the main topic of this chapter, the aforementioned projects represent a different way of thinking about hybrid editions. They *supplement* a printed edition mostly designed as a ready-to-use showcase publication for the interested public; they target distinct audiences (scholarly or other) and feature material of varied legal status. Neither is the case with the Pez edition, which, for reasons we will explain presently, deliberately aims to create a digital product that is equivalent to its printed counterpart. Both approaches are consistent with the relevant deliberations in Patrick Sahle's magisterial treatise on the matter, which, after introducing practices of data outsourcing (*Auslagerung*) under the heading of hybrid editing, also acknowledges (though somewhat reluctantly) the legitimacy and validity of calls for printed books (Sahle 2013 2, 61–69).

What we will present on the following pages is a description of precisely the scenario drafted by Sahle: the print publication as a derivative of a digital editorial environment, created primarily as a result of social and academic habits and practices rather than of an intrinsic technical exigency. This context must be taken seriously, however, for what Sahle and others appear to overlook at times is the close connection between the lack of sympathy for the "real-life" analogue aspect among digital editors and the rightly lamented marginalization of digital editions within the scholarly community.

2 The Pez ecosystem

Although digital methods have become a key feature of current humanities projects and an undeniable requisite for editions, only few projects seriously attempt anything but individual solutions when it comes to standardizing data and envisioning reusable

modules. There is no doubt that various attempts and viable practices exist in this regard, but the issue has yet to be addressed from the broader perspective of all the instances constituting the entirety of the workflow of a digital scholarly edition: researcher, hosting institution and series editor, publisher and repository. As will be pointed out on the following pages, it was precisely the relations between these partners, between their perspectives and requirements, that ultimately led to the conclusion that the optimal solution for the edition of the Pez correspondence is that of a hybrid one.

To be sure, there is another fundamental instance in the workflow. For several years now, the Austrian Science Fund (FWF) has been working hard to keep up with the fast-evolving landscapes of digital development that often diverge significantly between specific disciplines. One of the results of this work is the "innovative publication formats" funding scheme; one of several, however, when it comes to planning a hybrid edition: projects also have to take the FWF policies regarding open access and research data management as well as the institution's general research funding schemes into account.

The tangible output of the envisaged project (approved in 2021 as FWF PUD 23) will thus be twofold: its primary goal is the publication of a hybrid scholarly edition of the *full* correspondence of Bernhard and Hieronymus Pez, while the secondary goal is the development of a TEI-based model generic enough to cater to the needs of other editions to be published in the QIÖG (*Quelleneditionen des Instituts für Österreichische Geschichtsforschung*) series. As indicated in the previous paragraphs, the following text will use a pragmatic definition of *hybrid edition* as the publication of the same scholarly content—edited text, commentary and auxiliary material contextualizing and "unlocking" the edited source—in a print and a digital publication.

Although it is generally evident that the modes of access and thus the nature of the two publication formats differ fundamentally,[2] the chosen hybrid approach implies that—on a general level—the versions in both media contain the same scholarly content without favouring one format over the other. While this is a major contrast to a model of hybrid editions where one medium is complementary to the other (e.g. by providing raw transcriptions or facsimile images digitally while the book remains the sole medium for the edited, canonized text),[3] we argue for a clear conceptual and functional distinction between the two publication forms which, in our understanding, qualifies

---

[2] We find this distinction in many different gradations: Sahle (2013, 66) differentiates between "lesen" and "benutzen" (i.e. "reading [a fixated, edited text] vs. using [a digital tool to make both sources and editorial decisions accessible]"); Pierazzo (2014, 151) takes up the difference between "reading" and "seeking information" described by Ciula and Lopez (2009), who see the usage of the digital and analogue publication as a continuum ("a comprehensive process").

[3] Since there is no clear-cut categorization of such intermedial relations in hybrid editions, we take the spectrum of "use cases" collected in Pierazzo 2014, 150–3. Sahle similarly lists examples, but focuses on the variety of publication media rather than on their degree of overlap or complementarity (2013).

our undertaking as a hybrid one.[4] The digital format fosters an "explorative access" to the edition by offering native digital features like a full text search, hypertextualization of its constituents, or orthogonal views on the letters by relating them through common index terms—means of interaction that are inconceivable in printed form. However, these digital enhancements do not represent a disruption but rather a continuum with what the audience of the QIÖG series expects in terms of editorial practice.

Our editorial choice is thus not merely predetermined by the existence of already printed volumes of the Pez correspondence edition or the existence of a print series. It more broadly reflects the needs of a scholarly community accustomed to working with multiple editorial manifestations of the same text.[5] The need to combine a book with an online resource results from everyday practice and can be described as follows:

a) as the need for a synoptic view of resources that can hardly be achieved digit-ally—one can lay out twenty open books in a room, but not reasonably maintain an overview of twenty open browser windows;

b) the need to use a physical book copy as a container for comments, annotations and corrections collected over time—annotating and re-annotating the same *digital* document has simply not proven to be practical for the majority of scholars;

c) the need to collocate a book in a broader physical context representing a specific topic matter—no tagging system can substitute the intuitive topicality of a book shelf; and

d) the wish to safeguard sustainability—while this should not be treated as a religious matter, it is equally evident that there are sufficient examples for digital non-sustainability to justify such concerns.

In conclusion, Sahle (cf. footnote 2) is certainly right to point to a juxtaposition of "reading" (continuous texts conceived as such by their authors, in this case of historical sources) and "using" (bits and pieces of screen-size information); in precisely this

---

[4] "In die Gruppe der Hybrideditionen fallen solche digitalen Editionsformen, die in einer solchen Weise in verschiedenen Medien präsentiert werden, dass es zu einer inhaltlichen und konzeptionellen Rollen-verteilung kommt. […] Entscheidend ist, dass bestimmte Medien die Inhalte und die konzeptionelle Haltung zur Edition beeinflussen." (Sahle 2013, 68).

[5] This is also the case with the Pez letters themselves: digital images (scans) of large parts of the original correspondence kept at the abbey of Melk are published in an adjacent but separate project (Rabl 2013) aiming at a digital representation of the Pez brothers' papers. In this way, it becomes possible to interlink both datasets on an integration layer that draws on common models like RiC (Llanes-Padrón and Pastor-Sánchez 2017).

sense, reading and interpreting texts in their *entirety* has been and still is considered by many as part of the craft and identity of historians. The preferred medium for this is the book.

If one takes all these considerations and practices seriously, the hybrid approach can become a key factor in the current phase of scholarly transition towards a digital paradigm. It can help to reassure the community by way of a "regulatory authority" (Deegan and Sutherland 2009, 72) that counterbalances an often perceived "instability [of digital editions]." Furthermore, it can help to recruit experienced editors reluctant to leave their field of expertise and get them to invest their time in a medium that has yet to prove its sustainability, as well as young editors whose careers will depend on the acceptance of their work by the broader scholarly community.

Besides this user group, however, which we imagine will use the book and the online edition in parallel, we envision that the data itself will be of interest in many other as yet unconceived digital contexts—up to the point where it merges into what Gabler (2017, §9) calls "relational webs of discourse, energized through the dynamics of the digital medium into genuine knowledge sites." To this end, the data, which is also used to produce the digital edition, is being enriched (e.g. by using common identifier systems for named entities) in ways that allow for its easy consumption. Given the epistolary nature of the topic, this can mean services like *correspSearch* but also its potential deeper integration into the LOD paradigm (Ciotti and Tomasi 2016), thereby making it a distinct digital artifact serving a different audience. This separation of user groups with regard to both interest and digital literacy is reflected in the institutional setup of the project: The data consisting of the correspondence of Bernhard and Hieronymus Pez will be archived in ARCHE,[6] the CTS-certified repository for humanities research data run by ACDH-CH (Austrian Centre for Digital Humanities and Cultural Heritage). The user interface will be developed by the Böhlau publishing house and will take into account the needs of further publications in the QIÖG series.

We thus envisage creating a distributed architecture on a sustainable technical foundation that enables interaction with a rich, standards-based dataset in an intuitive manner on the one hand and producing an accompanying print edition that maintains the high standards of the existing series on the other. These goals proactively address the fact that complementary ways of working with editions have emerged as a consequence of the digital turn: besides close reading and the prevalent usage of print publications, the digital humanities have recently brought about programmatic ways of accessing an edition's content. To safeguard the sustainability of the results, we will follow a modular, service-oriented approach with well-defined interfaces between the systems maintained by differentiated institutions, which will allow us to

---

6   https://arche.acdh.oeaw.ac.at/browser.

replace individual modules in the future without having to reimplement the system as a whole.

3 The transforming edition of the Pez correspondence

The following diagram (Figure 1) shows the draft outline of such an environment serving as the blueprint for the development in the project.
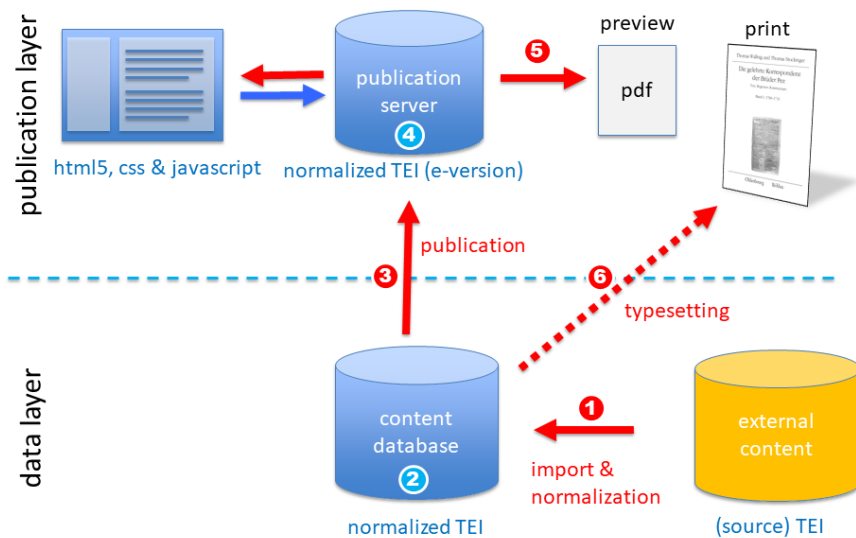
Figure 1: Conceptual diagram.

3.1 Content: the Pez correspondence and the QIÖG series

The Pez correspondence consists of roughly 1150 letters preserved primarily at Melk abbey as well as in various locations scattered all across Europe. It has been the object of editorial efforts since 2004, which to date have led to the *print* publication of roughly half of the letters within the QIÖG series.[7] The ongoing FWF project P28016

---

[7]   Open access: doi:10.7767/9783205794134.

(2016–2020),[8] however, is engaged with preparing the *remaining* letters for *digital publication* while reworking the extant publication into digital formats in parallel.

This process is the background against which the endeavour described in this paper should be viewed, for it reflects the need to create a model fit for digital publishing that simultaneously allows us to preserve previous "non-digital" editorial choices as well as to replicate the print layout adopted for the first two volumes of the Pez correspondence so as to maintain the overall print appearance of the series.

QIÖG is a series edited by the Institute of Austrian Historical Research and published by Böhlau / Vandenhoeck & Ruprecht (since 2021: Brill), with the first volume issued in 2008. The series mostly features stand-alone sources from pre-modern Central Europe whose specific nature is hardly appropriate for the larger series of source editions like the MGH. Each QIÖG volume includes a thorough introduction to the source itself and its genre and individual characteristics, as well as to the specific editorial rules that need to be adopted for it. Though oriented towards best-practice models, these solutions often break new ground and draw on an informed dialogue between the series editors and the individual scholars working with the respective source. This sometimes experimental approach also includes deliberations on digital edition and publication models, and the QIÖG series has therefore added two requirements to the present use case portfolio:

1. it insists, for institutional reasons, on a model allowing for print and online publication; and

2. it wants such a model to be reusable, ideally for all following QIÖG volumes to come.

In order to better understand this second requirement, some of the normalization described below in Chapter 3.2 has also been outlined for other QIÖG publications (namely Volumes 3 and 8).

The Pez correspondence itself is a source relevant to the early modern intellectual history of Central Europe, as it aptly displays the specific Catholic variant of religious antiquarianism and historical criticism. The letters are relevant in regard to the historical—medieval—source material they talk about, the international eighteenth-century networks of scholars traceable through their correspondence, the specific Latinity of Catholic late humanist erudition, and the specific framings of emergent discourses of "Austrian" and "German" history.

These specific features led to certain editorial decisions at the outset of the Pez project that determined the "data model" long before any discussion on digital matters; these decisions are described in detail at the beginning of the two printed volumes.

---

8   www.pezworkshop.org.

For the purposes of this paper, it should be remembered that

a) all preserved and inferred letters are documented by way of basic metadata;

b) each preserved letter is included with its full text, a German summary and an extensive commentary discussing its contents in detail;

c) some bio-bibliographical information is outsourced to the main index referring to the letter number, not the page; and

d) there are additional appendices and elements extracted from the letters, among which is a list of mentioned third-party letters and of material objects sent along with individual letters.

Footnotes are used for the critical apparatus, while the commentary is placed after the respective text passage it refers to. All three parts—letter text, summary and commentary—are divided into associated content segments so as to facilitate orientation within longer letters. This division serves as the edition's document-internal reference system and provides a hook between the three parts relating to each letter.

In the context of Pez volumes 1 and 2, this system was implemented in MS *Word* files by way of standard formats for the individual elements (metadata, summary, text, commentary, literature etc.). These files form the basis for the print layout and PDF files as well as for the digital transformation described in the following chapters.

3.2 Technical execution: implemented and envisaged workflows

Data creation and editorial process

Given the fact that two published volumes already exist, the adoption of a hybrid approach implies establishing two parallel workflows for integration into a single data format:

1. semi-automatic retro-encoding of the existing volumes into TEI-XML, and

2. establishment of an XML-first editing workflow for the letters to be edited in the ongoing project.

Both workflows need to implement a data model expressive enough to be useful as a dataset in its own right within the DH community. The focus in this regard lies on modelling a TEI customization with rich descriptive metadata and sound semantic and structural markup that makes implicit editorial conventions of the printed volumes explicit by means of TEI-encoding. On the other hand, the model needs to be flexible
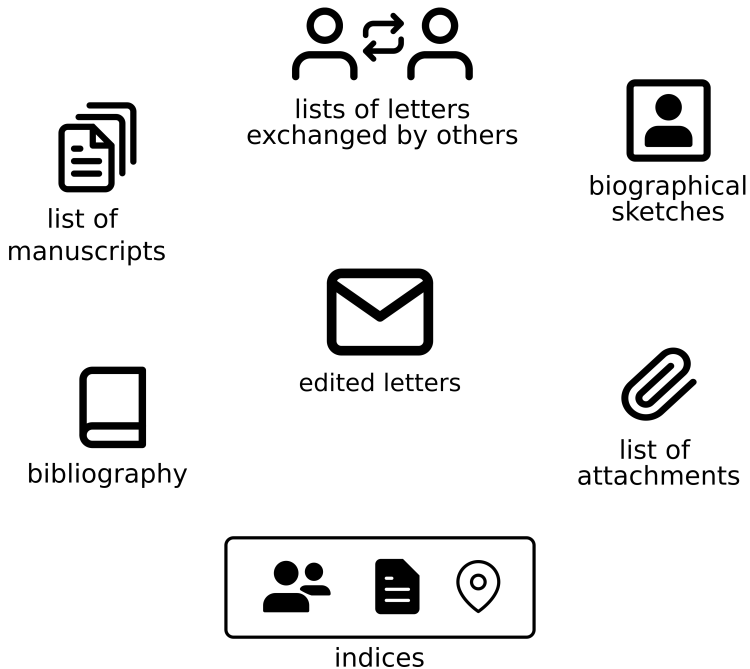
Figure 2: Data architecture of the Pez edition.

enough to support conventions established in the previous published volumes without requiring unreasonable effort. Lastly, the model has to encompass the full resource architecture, which includes material necessary to contextualize and "unlock" the letters as historical sources (see Figure 2).

Naturally, the edited letters are at the centre of the architecture, with each letter represented as a separate TEI document. The document encoding is modelled on the structure of the printed edition and therefore contains a `<front>` matter with the summary, the edited letter in the `<body>`, and a `<back>` matter including the commentary (i.e. a sequence of `<note>` elements attached to segments in the letter's text via a `@target` attribute).

The abovementioned *content segments* are tagged as corresponding `<seg>` elements both in the summary and the letter text, thus making them easily addressable TEI structures. It is important to mention that the editors already refrained from using page numbers as references in the existing print volumes (both in the index and the commentary), instead opting for the letter numbers along with a segment number

```xml
<TEI xmlns="http://www.tei-c.org/ns/1.0">
   <teiHeader>
      <!-- … -->
   </teiHeader>
   <text>
      <front>
         <div type="regest">
            <p>
               <seg type="context" n="1"> … </seg>
            </p>
         </div>
         <div type="editorialNote">
            <!-- … -->
         </div>
      </front>
      <body>
         <div type="editionText">
            <seg type="context" n="1"> …
               <seg type="lemma" xml:id="l_d4618e6102">Adiutores igitur in
                     partem operis vocati</seg>
            </seg>
         </div>
      </body>
      <back>
         <div type="commentary">
            <note target="#l_d4618e6102">
               <ref type="context">1</ref>
               <label>Adiutores <gap/> vocati</label>
               <p>… vgl. <ref type="letter">32 <ref type="context">3</ref></ref>.</p>
            </note>
         </div>
      </back>
   </text>
</TEI>
```

Figure 3: Pez TEI model.

where appropriate. This decision allows existing references to be effortlessly migrated into digital links and ensures the ability to cite across systems and media for both the digital- and analogue-born content.

In our model (Figure 3), most parts of the so-called *editorische Notiz* (roughly "editorial notes") are integrated in the `<teiHeader>`: Literature (including prior editions of the letter in question) is listed as `<relatedItem>` elements of various types, while contextual information is embedded in `<correspDesc>`.

In some cases, several writing stages of a letter are preserved. These variants are recorded by means of double-end-point-attached `<app>` elements placed in a separate division in the `<back>`. The majority of the existing apparatus entries found in the footnotes of the print edition can be mapped to standard TEI markup (mostly adopting the mechanisms provided by the *transcr* module) without loss of information, however in some cases such formalization is unfeasible or unreasonable because flat text in a page-oriented medium requires different modes of expression. The ability to reference null-positions may serve as a simple example: while it is easy from a data-oriented

proxima Deo favente exitura. His^u

---

  ^u   *Davor durchgestrichen* Tum *C.*

```xml
<!-- inside <body> -->
<anchor xml:id="app021"/>His<anchor xml:id="app021e"/> curis …

<!-- inside <back><div type="apparatus"> -->
<app from="#app021" to="#app021e">
     <rdg wit="#C"><del>Tum</del> His</rdg>
     <witDetail wit="#C">Davor durchgestrichen <mentioned>Tum</mentioned></witDetail>
</app>
```

Figure 4: Critical apparatus (Example 1).

```xml
<!-- inside <body> -->
… <anchor type="app-start" xml:id="app019"/>quam suscepi, librorum
sancti Irenaei episcopi Lugdunensis et martyris Adversus
haereses<anchor xml:id="app019e" type="app-end"/>, …

<!-- inside <back><div type="apparatus"> -->
<app from="#app019" to="#app019e">
    <lem wit="#B">quam suscepi <gap/> Adversus haereses</lem>
    <rdg wit="#C">a me concinnata operum sancti Irenaei episcopi
Lugdunensis ac martyris</rdg>
</app>
```

Figure 5: Critical apparatus (Example 2).

view to document that a specific word exists in a witness other than the base text, translating this piece of information into a comprehensible footnote generally requires the addition of a location indicator and thus potential custom wordings (see Figure 4).

Similarly, the lemmata within a printed apparatus (i.e. the words of the base text repeated at the beginning of an apparatus entry to assist the reader in locating the variation) can be easily constructed programmatically from the TEI markup in simple cases, but in cases of ambiguity, editorial intervention is necessary to select a reference that is also meaningful in a printed form. Here too, the model aims to fully consider both the data-centric and the print-centric approach, which comes at the slight cost of the duplication of information—or more precisely, of repeated expression of the same information (see Figure 5).

Besides the letters, the auxiliary material has likewise been converted into TEI-XML from the print edition's source documents (.docx files) by means of custom XSLT scripts. Further curation was necessary to enrich or correct this intermediate step:

a) The (hierarchically structured) indexes from Volumes 1 and 2 were merged into a single index and their entries categorized as referring to persons, organizations /places, manuscripts or works in a simple web interface. Wherever possible, reference resource IDs were manually added to the index entries.

b) The bibliographies of both volumes were extracted and imported into a shared Zotero library for further refinement.

c) The letters were semantically marked up with pointers to the merged index and the bibliography, and first steps towards encoding the critical apparatus were taken.

Since the model has evolved into a largely stable form and has been implemented as a TEI-ODD customization, it is ready to be employed as the foundation of the XML-first workflow that is being set up to encode the remaining letters in the ongoing FWF project. At its end, the completed dataset will be archived and published under a Creative Commons license in the ACDH-CH's repository ARCHE. The digital objects in ARCHE are referenceable via *handle* PIDs and made available through so-called "dissemination services", i.e. specialized web services that render the archived data. Building upon the data provided by such a service, the Böhlau publishing house plans to develop both an intuitive user interface and a workflow to create the print versions of the remaining volumes using the same conventions and typographic features as the existing parts.

Realization of this step requires a substantial investment in technical and social infrastructure, however—an investment that a commercial provider can only make if a critical number of similar publications can be produced with minimal additional overhead, thus leveraging scaling effects. The key thus lies in achieving the level of data homogeneity required for the software components of the publishing ecosystem to be reusable at little cost. Since the data model of any digital edition aims to reproduce its source documents as faithfully as possible (not necessarily in a visual dimension, but at least semantically and/or structurally), the data for volumes in a series like QIÖG will inevitably vary to a great extent. Based on our discussions regarding the Pez correspondence edition, we have drafted a strategy that should help to deal with this intrinsic diversity within the economic boundaries of a publishing house without forcing editors and DH scholars to dispense with the semantic richness required by their particular objects of interest.

Data normalization / mapping of input TEI to publication TEI

In order to make data processing as modular and reusable as possible, we propose a multi-level normalization process that distinguishes between three levels of data preparation:

1. The starting point is **source data** resulting from an edition project. This source or import data must be in TEI markup and represents the digital primary source.

In case of the Pez edition, we assume the primary data to be delivered by the digital long-time preservation repository ARCHE.

2. The first and crucial step of **normalization** prepares the TEI data for the actual publication process. The normalized data remains compatible with TEI, generating document structures or serializing document content in the form required by the publication concept. This corresponds to step (1) in the conceptual sketch in Figure 1.

3. The third step completes the initially media-neutral processed data with regard to an **electronic version**. This corresponds to step (3) in the conceptual sketch in Figure 1.

In the following, the two normalization steps will be presented in more detail (see also Table 1).

Ad 1: Examination of the TEI data is a requirement necessary for all XML processes. In this project, however, it refers to more than mere technical parsing according to the TEI scheme. It is also about examining all detailed structures that are important for publication as well as about detecting semantic inconsistencies. Examples of such checks are reference-relevant structures (links), important metadata, or document hierarchies. All errors must be removed before the next step.

Ad 2: Normalization and enrichment of the data are at the heart of the data preparation process. The goal of normalization is to reduce the variants in TEI tagging. Defined semantic structures (such as names of places or persons) are normalized so that they can later be used uniformly in the publication layer. The concept of data enrichment follows the observation that due to the rigid semanticization, TEI often does not correspond to the desired presentation and sequence in the publication layer. An example of this is metadata that is found in the header in TEI but is usually needed in the text at a defined point (see Figure 6). This necessitates additional effort to prepare the data for presentation. The theory here is that the two media basically have similar requirements and this step can therefore be completed for both of them together. As a result, normalization and enrichment means that significant problems of the publication layer have already been resolved, thereby simplifying the actual print and online publication process.

Ad 3: Accordingly, the last step of data preparation is easily explained. The normalized data can be fed directly into both media forms since all essential intrinsic issues of document delimitation and serialization of the text elements have been addressed. What remains is to implement the technical reference-target-structure and transform it into the desired publication format. For print media, this means an XML typesetting process based on established automatic typesetting techniques (XSL-FO, La-TeX etc.).

Figure 6: Transformation of metadata into text.

Source TEI	Normalized TEI = *step (1)* = *step (3)*	Normalized TEI (e-version)
A. General ·Representation of primary source ·Pure semantic markup	·Flattening of the semantic markup ·Addition of the edition-specific texts	Technical implementation for e-media
Example — Text-critical apparatus	Media-specific implementation	Technical implementation for e-media
B. Media-specific		
Example — Deletion of text	Generation of a (foot)note with deletion remark	Implementation of the (foot)note as an ID-IDREF link with its content
C. Semantic normalization — Semantic ambiguity in TEI markup across projects	Normalized TEI semantics	
Example — Different TEI tagging variants for persons	Unified TEI tagging of persons	
D. Serialization — Semantic tagging scatters information across different places in the markup	Serialization of all text parts for the purpose of the critical edition	
Example — Metadata with state of research, literature and text references is separated from the editorial text	Metadata is inserted as generated text in the desired text flow	
E. Reference and linking mechanism — Original semantics without reference and target structures	Selection and marking of the semantics for which reference and target structures are created	Technical implementation of reference and target structures
Example — Context-based position and counting of lemmata	Concrete counter of lemmata for representation issues	Lemmata with target ID corresponding to concrete counter

Table 1: Overview of normalization tasks.

For online publishing, the normalized TEI can easily be transformed into HTML5 plus CSS, or other electronic formats.

3.3 Publication layer

We focus on the publication layer for the digital medium as the concept of a printed text edition is well-known and established. The publication layer represents the interface between content and user. It therefore has to meet all of the needs of the user, whom we assume to belong to an academic community.

Even for a resilient user group like humanities scholars, a digital publication should be not simply functional but also user-friendly. In the following, we present mockups/sketches of a possible user interface that also include practical experience from the publisher's perspective.

Text representation and readability

A central component of text-based content is the readability of the text. In addition to well-thought-out typographic implementation, it is also important to consider the specific needs of the user.

Fundamentally, the corpus of the edition as well as the printed medium are presented in a similar fashion. Since the on-screen readability of extensive text documents such as the Pez letters suffers quickly, especially because the early modern syntax of various languages is often characterized by long sentences, not all parts of the text are displayed in the initial view. Instead, the document is displayed in a condensed representation providing the user with a quick overview (see Figure 7). Long text passages not needed for this overview are hidden, but the main text structures with their headings are displayed. The hidden text sections can then be displayed with a simple click if desired.

For a more in-depth comparison between e.g. summary and text, a "synoptic representation" and thus parallel access to the different contents is envisioned. Contextual information presented in footnotes in printed editions should ideally be visible in the respective context. Depending on the semantics of such footnotes, they can either be placed directly behind the respective text and displayed, for example, as a popup or in a sidebar next to the edited text. In the case of the Pez edition, footnotes are mostly textual variants that could sensibly be implemented as pop-ups following an appropriately marked passage of text.
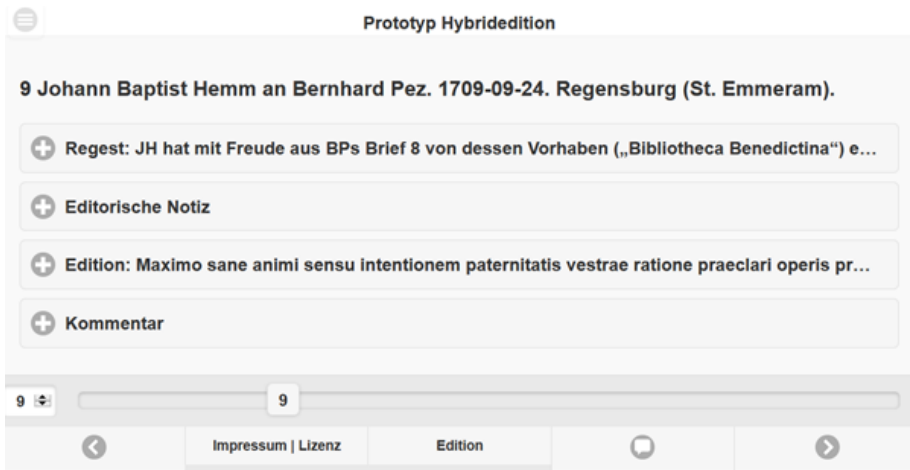
Figure 7: Condensed text representation of a document.

Navigation

In order to guide the user by means of specific information featured in the document, navigation for the usual semantic structures such as persons, chronology, places, works etc. should be provided. These navigation tools fulfil the function of a classic index in the printed edition. A special navigation system makes the individual documents of the edition accessible and shows where the user is located. Hierarchical navigation trees or breadcrumb navigation can be used as display options. In the case of the Pez edition, a simple list of letters is sufficient (see Figure 8).

Other user features and information

Finally, a number of basic features that are nowadays very common in content-based user interfaces should be mentioned—for the very reason that they are practically taken for granted:

- Navigation between documents: it must be possible to navigate between documents directly, once in the form of the usual "back" and "forward" buttons, but also by reference to the preceding or subsequent document of the correspondence.

- Print functionality: although a print function seemingly represents a media rupture, experience has shown that it has its legitimacy in digital use as well. It is crucial for this printing function to be simple and not replace the printed edition.
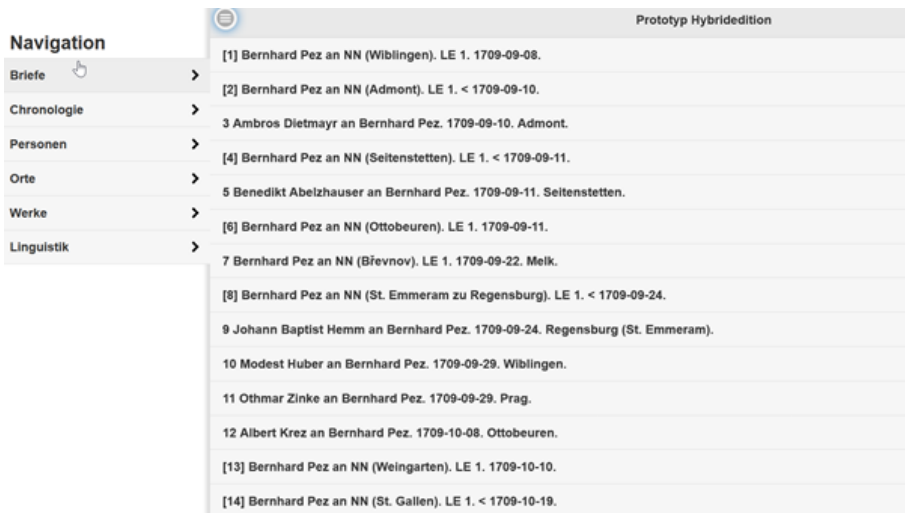
Figure 8: Navigation mockup.

- Simplifying citation: a citation button provides the user with a complete citation proposal for the current document that can be used immediately via the clipboard.

- Full text search: a simple and intelligent full text search should be able to find concatenated search terms as well as recognizing and finding terms with diverging spellings.

- User-specific views: while the order and selection of the letters/documents in the printed edition is fixed and unchangeable, users of the digital edition may use their own document selection or document view. For example, one could restrict the display of letters/documents using specific search criteria and thus create and keep an individual selection of documents. It would also be conceivable to generate an individual view by displaying or hiding certain information that corresponds to the researcher's current question.

Finally, certain information making the context of the document accessible should be displayed. This includes:

- keywords or register information;

- the necessary legal and copyright information of the publication, which includes the imprint and a clear indication of the assigned rights of use;

- ideally, the document should also feature a technically unique and unchangeable identifier. Given its prevalence in the publishing industry and the corresponding familiarity to readers, the use of a DOI seems an obvious choice here, though it is simultaneously debatable for reasons of cost and effort.[9]

Export instead of integrated analysis

The discussion so far has shown that complex search functions are not truly necessary since academic users generally have very specific questions. Instead of such functions, we therefore favor a variety of export functions:

- Export of defined standard lists displaying common search structures (people, places, works, etc.) and their contexts. These indexes should be easily accessible even for non-specialist users and contain links to the respective documents containing occurrences of the index entry.

- In order to allow custom queries as well as for reasons of data transparency, the underlying TEI data should be exportable.[10] Other export options could be provided as well, for example a plain-text version to be used with linguistic tools like *Voyant* (Sinclair et alt. 2016) or *AntConc* (Anthony 2014).[11]

4 Conclusion and strategic outlook: on institutional and technical sustainability

The complexity of the outlined design is a response to some of the concerns frequently associated with digital-first publications: From the point of view of the editors, relying on the expertise and reputation of a well-established publishing house can be an important factor. Even though *digital* (and *hybrid*) has hitherto been limited to page-oriented digital media like e-books or screen-optimized PDFs in many publishing houses, the benefit of optimizing workflows and having a single contact for producing both a book and a web-based edition can outweigh the costs of technical expertise yet to be developed.

---

9   In case of the Pez edition discussed here, the source documents are imported into the publication system from the ARCHE repository, which uses the handle.net system to assign PIDs for the archived resources. Although we can imagine a scenario where PIDs for the frontend view of the documents in the publication system are created during ingest and persisted in the `<teiHeader>` of the archived document, there may be cases where the publisher's system is the sole source of data, so that it must be capable of citable identifiers itself.

10 Although the fully encoded version of the Pez edition's data can be retrieved directly from ARCHE, we cannot assume this for any future publication served by the system.

11 This would be beneficial since workflows like the ones described in Fröstl 2019 are outside of the scope of non-expert users without technical support.

Separating the roles in such a process may also have the advantageous side effect of a higher level of sustainability not only in regard to the source data (which is provided in a well-established standard format and maintained by a certified repository hosted by an academic institution with public support), but also in regard to the much more volatile user-facing front end. Dissevering these concerns requires modularization and a clear-cut definition of procedural and technical interfaces, which can become key factors in the future, should parts of the infrastructure grow obsolete—but only if both the data and the implementation of the publishing platform are published under a license permitting reuse and adaptation by third parties. Bridging the differences between commercial interests, public funding policies, and scholars in both DH and other disciplines will be a matter of discussion for the years to come along with the development of licensing schemes that respect and foster every actor's effort in a complex and dynamic field subject to continuous transformation.

Bibliography

Anthony, Laurence. 2014. *AntConc: Version 3.4.1w* [computer software]. Tokyo: Waseda University. Accessed June 15, 2018. http://www.laurenceanthony.net.

Bosse, Anke, Artur Boelderl, and Walter Fanta, eds. 2016. *Musil Online.* Klagenfurt: Robert Musil Intitut / Kärntner Landesarchiv. Accessed March 7, 2021. http://musilonline.at.

Ciotti, Fabio, and Francesca Tomasi. 2016–2017. "Formal Ontologies, Linked Data, and TEI Semantics." *Journal of the Text Encoding Initiative* 9, online since 24 September 2016. doi:10.4000/jtei.1480.

Driscoll, Matthew James, and Elena Pierazzo, eds. 2016. *Digital Scholarly Editing. Theories and Practices.* Cambridge, UK: Open Book Publishers.

Fröstl, Michael. 2019. "Digital-linguistische Annotation neulateinischer Texte. Theorie und Praxis für Editions- und Geschichtswissenschaft." MA Thesis, University of Vienna (Institute of Austrian Historical Research). doi:10.25365/thesis.58660.

Gabler, Walter. 2018. "Theorizing the Digital Scholarly Edition." In: *Text Genetics in Literary Modernism and Other Essays* [Online]. Cambridge: Open Book Publishers, 2018. Accessed March 7, 2021. http://books.openedition.org/obp/5453. [originally published in 2010 *Literature Compass* 7 (2)].

Hannesschläger, Vanessa, ed. 2015. *Ernst Jandl Online.* Wien: Österreichische Nationalbibliothek / Ludwig Boltzmann Institut für Geschichte und Theorie der Biographie. Accessed March 7, 2021. https://jandl.onb.ac.at.

Kastberger, Klaus (with the collaboration of Katharina Pektor and Christoph Kepplinger-Prinz), eds. 2015. *Handkeonline.* Accessed March 7, 2021. https://handkeonline.onb.ac.at.

Llanes-Padrón, Dunia, and Juan-Antonio Pastor-Sánchez. 2017. "Records in Contexts: the Road of Archives to Semantic Interoperability." *Program* 51 (4): 387–405. doi:10.1108/PROG-03-2017-0021.

Pierazzo, Elena. 2014. *Digital Scholarly Editing: Theories, Models and Methods*. Grenoble. Accessed June 21, 2020. http://hal.univ-grenoble-alpes.fr/hal-01182162.

Rabl, Irene. 2013. "Der digitalisierte Nachlass der Brüder Bernhard und Hieronymus Pez. Ein Projektbericht." *MIÖG* 121: 437–44.

Sahle, Patrick. 2013. *Digitale Editionsformen. Zum Umgang mit der Überlieferung unter den Bedingungen des Medienwandels. Teil 2: Befunde, Theorie und Methodik.* Norderstedt: Books on Demand. Accessed March 7, 2021. https://kups.ub.uni-koeln.de/5352.

Stockinger, Thomas, et al. 2015. *Die gelehrte Korrespondenz der Brüder Pez. Text, Regesten, Kommentare, 2: 1716–1718*. Quelleneditionen des Instituts für Österreichische Geschichtsforschung 2/2. Wien: Böhlau. Accessed March 7, 2021. https://www.oapen.org/search?identifier=576951; http://www.oapen.org/search?identifier=576952.

Straub, Wolfgang, and Claudia Dürr, eds. 2019. *Werner Kofler. Kommentar zur Werkausgabe*. Accessed March 3, 2021. http://wernerkofler.at.

TEI Consortium. 2019. "TEI P5: Guidelines for Electronic Text Encoding and Interchange. Version 3.5.0." Accessed April 23, 2019. https://tei-c.org/Vault/P5/3.5.0/doc/tei-p5-doc/en/html.

Van den Heuvel, Charles, et al. 2019. "Transcribing and Editing text." In *Reassembling the Republic of Letters in the Digital Age. Standards, Systems, Scholarship*, ed. by Howard Hotson, and Thomas Wallnig, 237–64. Göttingen: GUP. doi:10.17875/gup2019-1146.

Wallnig, Thomas, and Thomas Stockinger, eds. 2010. *Die gelehrte Korrespondenz der Brüder Pez. Text, Regesten, Kommentare, 1: 1709–1715*, Quelleneditionen des Instituts für Österreichische Geschichtsforschung 2/1. Wien: Böhlau. Accessed March 7, 2021. http://www.oapen.org/search?identifier=445402.