Joseph Haydn Werke Metadata: MEI Way

Joshua Neumann
Akademie der Wissenschaften und der Literatur | Mainz
Germany
joshua.neumann@adwmainz.de

Abstract

The work of the Joseph Haydn Institute has been well-known in eighteenth-century studies and beyond since publishing the first four volumes of the Joseph Haydn Werke in 1958. As with any *Gesamtausgabe* undertaking, sources occupy a central role, and with them come massive amounts of data. Naturally, compilation and organization of the metadata occurred over the life of this project, with digitization only a comparatively recent focus. Multiple factors led to the development of an idiomatic schema. Inasmuch as this system served immediate needs and created a foundation for content findability, it created limitations in accessibility, interoperability, and reusability—all desirable or essential qualities for the online Joseph Haydn Portal. It creates a distinct set of challenges for creating a digital *Werkverzeichnis* within the portal, the most pressing of which is transforming data into a standardized format enabled for the necessary qualities. This poster provides an overview of this process using file samples, concordances for terms and structure, and presents the challenges involved in a project of this size, and the realities of planning the project life cycle.

Introduction

From its inception in 1955, the Joseph Haydn Institute Köln (JHI) has focused on preparing and publishing the first complete academic edition of Haydn's works. As of 2021, 132 volumes have been published, with only the 133rd–a supplemental addition–remaining in progress (Joseph Haydn-Institute, 2021). Digitization of the card-catalogue records for nearly 3,000 works attributed to or associated with Joseph Haydn at some point has occupied approximately the past decade. For development of the *Joseph Haydn Werkverzeichnis* (JHWv), the JHI has partnered with the Centre for Digital Music Documentation (CDMD) at the *Akademie der Wissenschaften und der Literatur Mainz* (ADWMainz). Work on the JHWv began in the autumn of 2021 with the sharing of more than 3,000 XML files via an SVN repository, and the direction to develop it into a widely usable, accessible, online entity to add to the Haydn portal.

1 JHI Records

1.1 Contents

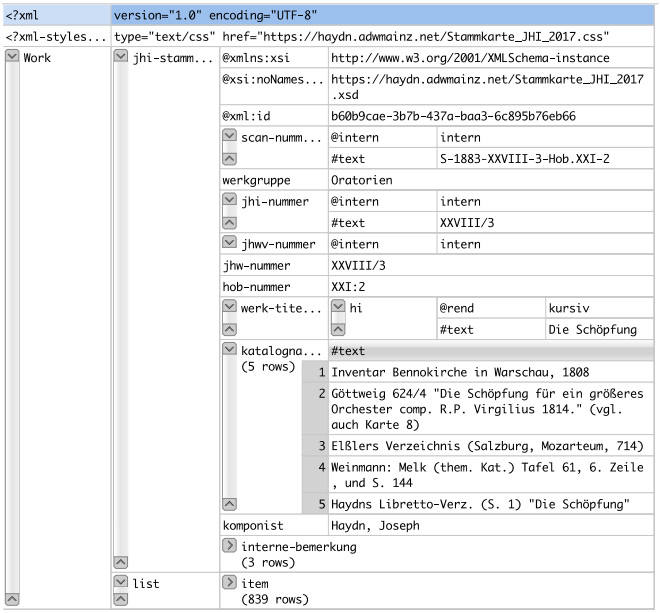
The data corpus for the JHW project consists of work records (XML) and various XSL, CSS, and XSD files. A combination of TUSTEP and XPath analysis revealed the description of this corpus's metadata exhibited in Table 1. Within the files, the best identifier of unique records for individual works is the element `<werk-titel-normiert>`, while `<quelle>` is the best unique identifier for sources. Furthermore, and having identified erroneous attribution as a major flaw in the Hoboken catalogue, JHI researchers across the lifecycle of the JHW identified numerous instances of false or challenged attribution in a variety of ways. Finally, the entire dataset is organized into folders corresponding to the many volumes of the JHW series.

Total files indicated by element `<jhi-stammkarte>`	3,203
Total tag (element) names	168
Files w/ `<werk-titel-normiert>` as element	3,190
Total <quelle> instances	29,214
Files with some degree of false or challenged attribution involved (can be at level of `<werk-titel-normiert>` or `<quelle>`)	1,464

Table 1: Overview of Contents in JHI Data

1.2 Structure

Each file contains information for both works and all sources recognized by the Haydn Institute, with the works section acting as the metadata container for the sources (see Figures 1 and 2). The structure in these files mirrors that of the structure on the hard-copy cards, evidencing that the task of digitization by JHI was for verbatim preservation of the cards and thus dictated the contents and structure of the digital files. Given JHI's source cataloguing efforts, the files exist in a tree structure while also reflecting some minor changes in both structure data labels <tag> terminology. These changes appear to be the result of evolving cataloguing standards and praxis over the nearly 70 years of the Haydn Institute's work.



<?xml	version="1.0" encoding="UTF-8"		
<?xml-styles...	type="text/css" href="https://haydn.adwmainz.net/Stammkarte_JHI_2017.css"		
Work	jhi-stamm...	@xmlns:xsi	http://www.w3.org/2001/XMLSchema-instance
		@xsi:noNames...	https://haydn.adwmainz.net/Stammkarte_JHI_2017.xsd
		@xml:id	b60b9cae-3b7b-437a-baa3-6c895b76eb66
		scan-numm...	@intern: intern; #text: S-1883-XXVIII-3-Hob.XXI-2
		werkgruppe	Oratorien
		jhi-nummer	@intern: intern; #text: XXVIII/3
		jhwv-nummer	@intern: intern
		jhw-nummer	XXVIII/3
		hob-nummer	XXI:2
		werk-tite...	hi: @rend: kursiv; #text: Die Schöpfung
		katalogna... (5 rows)	#text: 1 Inventar Bennokirche in Warschau, 1808; 2 Göttweig 624/4 "Die Schöpfung für ein größeres Orchester comp. R.P. Virgilius 1814." (vgl. auch Karte 8); 3 Elßlers Verzeichnis (Salzburg, Mozarteum, 714); 4 Weinmann: Melk (them. Kat.) Tafel 61, 6. Zeile, und S. 144; 5 Haydns Libretto-Verz. (S. 1) "Die Schöpfung"
		komponist	Haydn, Joseph
		interne-bemerkung (3 rows)	
	list	item (839 rows)	

Figure 1: Sample XML file showing general metadata structure for a musical work, JHW XXVIII/3.

<?xml	version="1.0" encoding="UTF-8"		
<?xml-styles...	type="text/css" href="https://haydn.adwmainz.net/Stammkarte_JHI_2017.css"		
Source	@ref	#a1b5386f-9395-49d6-a02e-cf87891b74e4	
	<!--	Abschriften (Bearbeitungen, Auszüge)	
	q-jhi-sigel	@intern	intern
		#text	CH-Zz, Mus NL 21: Bg 1
	q-scan-nummer-quellenkarte	@andere_quelle	rism-nr
		@intern	intern
		#text	02-12572-ZueZB-Mus.NL21-Bg1
	q-quellentyp	Sammelabschrift/Konvolut	
	q-material-art	Stimme(n)	
	a-ms-rism-bibliothekssigel	CH-Zz	
	a-ms-signatur	Mus NL 21: Bg 1	
	q-sammel-inh	q-sammel-inh-ueber-nach-nr	
		q-sammel-inh-einzel-nr (50 rows)	
	q-sammel-werkspezifika	q-sammel-spez-werk-vollstaendig...	2., 3. Satz
		q-sammel-spez-werk-rism-permalink	https://opac.rism.info/search?id=402000836&View=rism; https://opac.rism.info/search?id=402000851&View=rism
	ms-pr-bearbeitung	ms-pr-bearb-typ	Bearbeitung
		ms-pr-bearb-besetzung	Kl.
	q-in-jhwv-gedruckt	nein	
	q-rism-permalink	https://opac.rism.info/search?id=402000660&View=rism	

Figure 2: Sample XML file showing general metadata structure for a printed source, JHW I/9.

2 End-Use Formatting and Transformation

2.1 Goals and Characteristics

As CDMD operates in conjunction with the *Nationalen Forschungsdaten-infrastruktur* for Culture (NFDI4C), the philosophical alignment also guides praxis and end-usage alignment of characteristics. In this context, the following structural goals are central:

> Standards for handling research data (e.g., linking entities with corresponding authority records, using ontologies and controlled vocabularies) as well as quality features (e.g., versioning and licensing) facilitate the publication of research results in a visible and reusable way and allow for relating them to each other. (NFDI4Culture)

Further guiding the Haydn project and all other CDMD projects are the FAIR Principles of findable, accessible, interoperable, and reusable data management (GO FAIR). To attain FAIR compliance, as with many digital collected works catalogue projects, the JHWv online will be an open access relational database with downloadable XML data for both works and sources.

2.1 Adopting/Imposing MEI Structure

Given its increasingly widespread use, overlap in the aims of both NFDI4Cultre and FAIR Principles, and ability to include multiple encoding formats, MEI is the ideal destination format for this data. With its idiosyncratic structure, a simple transformation into MEI's schema is not possible, even if using the now community-supported Metadata Editor and Repository for MEI Data (Danish Center for Music Editing, 2020) (MerMEId). Even so, the Catalogue of Carl Nielsen's Works (Danish Center for Music Editing, 2019), the Catalogue of the Works of Frederick Delius (University of Oxford), the *Werkverzeichnis* Anton Bruckner (*Österreichische Akademie der Wissenschaften*), and other projects using MerMEId serve as foundational examples.

The workflow for this transformation required first the evaluation of the JHI data to map its structure, then the assessment of MEI standards and experimentation with MerMEId as a potential tool to effect the desired changes. Within the Haydn data, separating sources from works within the files facilitates easier editing of contents, especially with works appearing in hundreds of sources. Given the natural connections between works and sources in Haydn's oeuvre, MEI's use of FRBR relations is invaluable. Collaborations with colleagues fluent in TUSTEP enabled the establishment of content and technical links between sources and works via the element `<scan-nummer-stammkarte>` and the attribute `@xml:id`, respectively.

The second step in the process is the ongoing development of concordances mapping terminology and structure from JHI into MEI (and their appropriate FRBR categories). Most terms in the JHI data have straightforward cognates in MEI; terms that do not require structural analysis, (JHI) usage analysis and translation. These concordances serve both as filters for clarifying overlapping usages between different terms and as templates for final structural transformation via TUSTEP. Structuring a new SVN tree branch with work and source files grouped in master folders permits JHI staff to continue data cleaning while CDMD staff develops

and tests concordances for tag names and structural transformation. Following alpha and beta testing to ensure accuracy, file transformation will occur in batches, first of work/source separation and finally of structural transformation.

Conclusion

While no idiosyncratically constructed XML data set will likely translate into the MEI schema seamlessly, the Joseph Haydn *Werkverzeichnis* project offers valuable perspectives for schema changes and project planning alike. For schema changes, JHWv illustrates the need to consider both terminological and structural translation. For project planning, it serves as a reminder of the value of the MEI schema's practical and theoretical aims and praxis. For project lifecycle planning, the JHWv metadata conversion process encourages anyone undertaking a similar, corpus-based encoding to strongly consider the MEI schema.

Acknowledgements

The author expresses thanks to the following for assistance in various elements of this project: Joseph Haydn Institute Köln, Hans-Werner Bartz and Thomas Kollatz (*Akademie der Wissenschaften und der Literatur Mainz*), Kristina Richts-Matthei (*Universität Paderborn*), Jessica Abbazio (University of Minnesota), and the MEI Digital Pedagogy Interest Group.

References

Austrian Centre for Digital Humanities and Cultural Heritage. (2015). *Werkverzeichnis Anton Bruckner*. http://www.bruckner-online.at/?page_id=1588

Danish Center for Music Editing. (2019). *Catalogue of Carl Nielsen's Works*. Royal Danish Library. http://www5.kb.dk/dcm/cnw/preface.xq

Danish Center for Music Editing. (2020). *MerMEId – Metadata Editor and Repository for MEI Data*. Royal Danish Library. https://mermeid.edirom.de/index.html

GO FAIR. (n.d.) *FAIR Principles*. Retrieved December 10, 2021, from https://www.go-fair.org/fair-principles/

Haydn, J., S. C. Fisher, & S. Gerlach. (2002). *Sinfonien um 1777—1779*. G. Henle.

Haydn, J. & A. Opperman. (2008). *Die Schöpfung : Oratorium, 1798*. G. Henle.

Joseph Haydn-Institut. (2021). *Joseph Haydn Werke*. https://www.haydn-institut.de/index.php/gesamtausgabe/gesamtausgabe-2

NFDI4Culture. (n.d.) *NFDI4Culture Community: Musicology*. Retrieved 07 December, 2021, from https://nfdi4culture.de/about-us/communities/musicology.html University of Oxford. (2016). *Delius Catalogue of Works*. https://delius.music.ox.ac.uk/catalogue/welcome.html

University of Tübingen. *Tübinger System von Textverarbeitungs-Programmen*. Last revised 01 February 2022. http://www.tustep.uni-tuebingen.de/tustep_eng.html