



Erhebung, Transformation und Präsentation digitaler Forschungsdaten

Dennis Ried, Karlsruhe

DOI: 10.25366/2020.92

Zitation: Dennis Ried, „Erhebung, Transformation und Präsentation digitaler Forschungsdaten“, in: *Brückenschläge zwischen Musikwissenschaft und Informatik. Theoretische und praktische Aspekte der Kooperation*, in Verbindung mit der Fachgruppe Digitale Musikwissenschaft hrsg. von Stefanie Acquavella-Rauch, Andreas Münzmay und Joachim Veit (= Musikwissenschaft: Aktuelle Perspektiven. Bericht über die Jahrestagung der Gesellschaft für Musikforschung 2019 in Paderborn und Detmold, Bd. 3), Detmold, Musikwissenschaftliches Seminar der Universität Paderborn und der Hochschule für Musik Detmold, 2020, S. 41–46, DOI: 10.25366/2020.92



Erhebung, Transformation und Präsentation digitaler Forschungsdaten

DENNIS RIED, KARLSRUHE

Wird in der digitalen Musikwissenschaft ein neues Projekt ins Leben gerufen, ist es nahezu obligatorisch, Datenbanken aufzubauen. Wie aber gestaltet sich dieser Prozess in kleineren Projekten, bei denen kaum oder gar keine Dateninfrastruktur bzw. technische Unterstützung zur Verfügung steht? Gerade für Dissertationsprojekte kann dies eine große Herausforderung darstellen. Das Problem ist hierbei weniger die Erhebung der digitalen Inhalte als vielmehr deren Verwaltung und (öffentlich zugängliche) Präsentation.

Das zu diesem Text korrespondierende Poster stellt Prozesse der Datenerfassung und Aufbereitungen in zwei vergleichbar kleinen Projekten vor. Bei Baumann Digital handelt es sich um ein Promotionsprojekt zu Leben, Werk und Wirken des Karlsruher Männerchorkomponisten Ludwig Baumann (1866–1944). Der Aufbau eines Online-Portals für das 2018 gegründete Joachim-Raff-Archiv in Lachen/SZ (CH) bildet das zweite Projekt.

Baumann Digital (BauDi)

Am Anfang dieses Dissertationsprojektes¹ stand eine Prämisse: Die Erfassung und Auswertung der Daten sollen genuin digital erfolgen und den Grundstein für eine tiefere Beschäftigung mit dem Thema legen. Eine Gesamtausgabe ist derzeit zwar nicht geplant, doch bietet sich hier die Möglichkeit, die Grenzen des angewandten Systems (auch für eine solche) auszuloten.

Recht schnell wurde deutlich, dass es in der Masse an MEI- und TEI-kodierten Daten nicht leicht ist, den Überblick zu behalten, weshalb eine Form der Organisation gefunden werden musste. Die für BauDi erstellte simple Form der Visualisierung war zunächst nur für den internen Gebrauch vorgesehen – d. h. für eine strukturierte Ansicht, eine Korrektur der digitalen Inhalte und deren erste Auswertung. Mit der Zeit wurde dieses System jedoch immer detaillierter und ergänzt nun seinen ursprünglichen Zweck um die Dimension der öffentlichen Darstellung der Forschungsdaten.

Selbstverständlich gibt es bereits ausgereifte Tools zum Anlegen, Verwalten und Visualisieren von Daten im MEI- und TEI-Format, doch sind diese Umgebungen zwar mächtig im

1 Dennis Ried, *Deutsche bürgerliche Chormusik in der ersten Hälfte des 20. Jahrhunderts. Eine Studie zu Leben, Werk und Wirken Ludwigs Baumanns* (Arbeitstitel), Dissertationsprojekt, Hochschule für Musik Karlsruhe.

Leistungsumfang, dafür jedoch häufig komplex in Installation, Konfiguration und Anpassung.² Aufgrund fehlender Unterstützung im informatisch-technischen Bereich wurde zunächst damit begonnen, eine eigene Umgebung zu schaffen, die den projektspezifischen Ansprüchen genüge. Die schnell wachsenden Datenbestände mussten also organisiert werden, um einen Überblick zu bekommen bzw. diesen nicht zu verlieren. Automatische Register stellten eine schnelle und effiziente Lösung dar. Automatische Register meint hier in erster Linie vorsortierte Listen, die von der Arbeitsumgebung zusammengestellt werden und sich automatisch bei Änderung des Datenbestandes aktualisieren, sodass immer mit der größtmöglichen Datenmenge gearbeitet werden kann. Diese Register erlauben dann auch erste Analysen, beispielsweise über die Menge an Daten in einzelnen Kategorien (Handschriften, Gattungen, Erwähnungen u.v.m.). Ebenso notwendig wurden (rudimentäre) Einzelansichten etwa zum Korrekturlesen der Datensätze.

Joachim-Raff-Archiv (JRA)

Auch das *JRA* hat sich als Ziel gesetzt, die Forschungsdaten in den Standards TEI und MEI zu erfassen, um die X-Technologien zur weiteren Verarbeitung einsetzen zu können. Im Vergleich zu *BauDi* konnte jedoch nicht direkt mit der Erfassung in TEI und MEI begonnen werden. Aufgrund der Vorbedingungen im Archiv gab es vor Projektbeginn keine Zeit zur Erstellung von Kodierungsrichtlinien, für die Schulung von Mitarbeiter*innen oder für die Entwicklung einer digitalen Infrastruktur.³ Die Daten mussten parallel zur Entwicklung eines projektspezifischen TEI- bzw. MEI-Schemas erfasst werden.

Dieser Ausgangslage ist es geschuldet, dass die Daten in der Initiationsphase zunächst in tabellarischer Form (.xlsx) erfasst wurden, um den Mitarbeiter*innen einen einfachen, strukturierten Zugang zum System zu ermöglichen. Zu Projektbeginn konnte lediglich festgelegt werden, dass die Erfassung so kleinteilig wie möglich stattfinden sollte (eine eigene Zelle für jeden Vornamen, jedes Datum, jeden Ort, usw.). Die Herausforderung bestand nun darin, die überaus zahlreichen, zergliederten Metadaten in der Folge in TEI-/MEI-konforme XML-Dateien zu transformieren.

Durch die Exportformate CSV bzw. XML der Eingabe-Datei konnte eine Zwischenstufe hergestellt werden, in der eine erste Bereinigung und Nachjustierung der Struktur vorgenommen werden konnte. Diese Zwischenstufe wurde dann mittels XSLT-Skripten in die gewünschten projektspezifischen XML-Standards, die mittlerweile festgelegt worden waren, überführt. Nach einer weiteren Nachbereitungs- und Bereinigungsphase wurde schließlich ein Zustand erreicht,

2 Beispielhaft zu nennen sind: MerMEId (<<https://music-encoding.org/resources/tools.html#mermeid>>) oder teipublisher (<<https://teipublisher.com/index.html>>).

3 Die Gründungsförderung war auf zwei Jahre begrenzt, weshalb bereits dann Ergebnisse vorliegen mussten, um eine Anschlussförderung erhalten und den Fortbestand des Archivs sicherstellen zu können.

ab dem die Arbeit nun direkt und ausschließlich in XML erfolgen konnte. Den (Erst-)Einstieg über Excel-Dateien zu lösen, hat sich in diesem Fall als überaus hilfreich und problemfrei erwiesen. Zur Erfassung weiterer Daten wird jedoch von dieser Lösung abgesehen, da der Transformationsprozess nach Abschluss der Initiationsphase einen (in diesem Fall unnötigen) Mehraufwand bedeutet.

portal-app

Eine zeitnahe Veröffentlichung der erhobenen Daten war für das *JRA* obligatorisch, während die Daten bei *BauDi* erst nach Verteidigung der Dissertationsschrift vollends publiziert werden können. In Zusammenarbeit mit dem *JRA* wurde die ursprünglich für *BauDi* begonnene Präsentationsplattform⁴ nun detaillierter ausgestaltet.

Um die Datensätze zu visualisieren, wird eine xQuery-basierte Anwendung in Verbindung mit einer eXist-Datenbank verwendet, welche die Daten verwaltet. Der Vorteil der Verwendung von eXist-db liegt einerseits darin, dass es sich hierbei um eine native XML-Datenbank handelt und andererseits, dass diese Open Source Datenbank in DH-Projekten stark verbreitet ist; Entwicklungen und Lösungen anderer Projekte können dadurch direkter nachgenutzt werden. Ein weiterer Vorteil liegt in der Indexierung ausgewählter Daten Bestandteile, die eine eklatante Beschleunigung der Anwendung erlaubt.

Im Wesentlichen besteht das Portal aus mehreren Registerseiten, die den Datenbestand vorstrukturiert und vorsortiert bereitstellen. Diese Register werden dynamisch erzeugt, sodass ein Update im Datenbestand keine manuelle Aktualisierung der Übersichten erfordert. Von den Registerseiten aus gelangen die Nutzer*innen zu den einzelnen Datensätzen, deren Darstellung sich ebenfalls dynamisch aufbaut.

Durch die eindeutige Identifikation von Personen, Institutionen, Werken, usw. ist es möglich, auch eine automatische Referenzierung herzustellen.⁵ Beim Öffnen des Einzeldatenblattes führt das entsprechende xQuery-Skript dann eine Abfrage durch, die über die hinterlegten Daten hinaus derzeit zwei weitere Übersichten zusammenstellt: die bisher erfasste Korrespondenz sowie „sonstige“ Bezüge. Diese „sonstigen“ Bezüge können beispielsweise sein: Widmungsträger, Adressat im Brief, Affiliationen, Nennung in einer Regeste (auch Werke, andere Dokumente, Personen usw.) uvm. Wird eine solche Referenz von der App gefunden, setzt diese automatisch einen Link zum entsprechenden Datensatz. Diese Art der Verknüpfung wird auch auf Textblöcke angewandt, sodass nicht nur innerhalb der Register, sondern auch innerhalb der Volltexte diese Form automatischer Verknüpfung möglich ist. Mit jedem neuen Datensatz und

4 BauDi: <<https://github.com/Baumann-Digital/portal-app>> (Source-Code); JRA: <<https://github.com/Joachim-Raff-Archiv/portal-app>> (online unter: <<http://portal.raff-archiv.ch>>).

5 Bsp.: <persName key="C00695">Raff</persName>; der Wert in @key identifiziert den String „Raff“ eindeutig. Beim Auslesen der Daten, setzt die portal-app einen Link zum referenzierten Datensatz; in diesem Fall zum Personendatensatz von „Joseph Joachim Raff“.

jeder neuen identifizierten Information erweitert sich dieses Netzwerk und stellt immer ausführlichere Zusammenhänge zwischen den Datensätzen her.

Zitation: Dennis Ried, „Erhebung, Transformation und Präsentation digitaler Forschungsdaten“, in: *Brückenschläge zwischen Musikwissenschaft und Informatik. Theoretische und praktische Aspekte der Kooperation*, in Verbindung mit der Fachgruppe Digitale Musikwissenschaft hrsg. von Stefanie Acquavella-Rauch, Andreas Münzmay und Joachim Veit (= Musikwissenschaft: Aktuelle Perspektiven. Bericht über die Jahrestagung der Gesellschaft für Musikforschung 2019 in Paderborn und Detmold, Bd. 3), Detmold, Musikwissenschaftliches Seminar der Universität Paderborn und der Hochschule für Musik Detmold, 2020, S. 41–45, DOI: 10.25366/2020.92

Abstract

The process of data acquisition and transformation of Baumann Digital and Joachim-Raff-Archiv is the subject of this poster description. The data in both projects is (meanwhile) based on TEI and MEI to obtain a machine-readable corpus. The presentation platform created for this project was initially intended as a simple form of visualization for internal use, i. e. for structuring, proofreading of digital contents and first evaluations. The xQuery-based application for eXist-db was developed in detail during the first period of the Baumann Digital project, adding the dimension of public presentation to its original purpose. This poster focuses on the process of collecting, transforming and presenting the data of these two small projects, which have different research interests but deal with similar archival material.

Kurzvita

Dennis Ried (*1989 in Mannheim) studierte von 2010 bis 2017 die Fächer Germanistik (B.A. 2014) am Karlsruher Institut für Technologie sowie Musikwissenschaft (B.A. 2015, M.A. 2017) an der Hochschule für Musik Karlsruhe. Seit Herbst 2017 befindet er sich im Promotionsstudium. Seit 2018 ist Dennis Ried Wissenschaftlicher Mitarbeiter am Max-Reger-Institut und dort in der Reger-Werkausgabe, seit 2019 auch als Editor beschäftigt. Er betreut seit 2018 Forschungsdatenmanagement und Entwicklung des Online-Portals des Joachim-Raff-Archivs.

Creating, transforming and performing data at Baumann-Digital and Joachim-Raff-Archiv

introduction

When a new project in musicology is going to start it's often planned to build up digital data. But what happens in little projects where no deep infrastructure or technical support is given? This could be a very big challenge especially in little projects like a PhD thesis or a newly founded archive with only one or a few employees.

projects



Ludwig Baumann – studies on work and life

BauDi In this dissertation project the collected data are recorded directly as TEI and MEI files. Automated registers and linking help to clean up and explore the data.



Databases for letters, works and persons

JRA The Joachim-Raff-Archiv (Lachen, SZ) had used detailed Excel spreadsheets to create the data sets with the aim of later converting them into TEI and MEI files. Thus, the **JRA** project forms a kind of preliminary stage to the **BauDi** project in dealing with digital registers.

source code on Github

Baumann-Digital/portal-app

The content will also be published there as soon as the dissertation project is completed.

Joachim-Raff-Archiv/portal-app

This repository is a fork from Baumann-Digital

contact

BAUMANN-DIGITAL

Dennis Ried, M.A.
ried-musikforschung@mail.de

JOACHIM-RAFF-ARCHIV

http://raff-archiv.ch
forschung@joachim-raff.ch

creating data

BauDi

- generating XML files by using XQL-scripts
- creating TEI files to store non-musical data
- genuine digital acquisition

JRA

- no time to develop a workflow
- creating hundreds of records (in short time)
- detailed Excel spreadsheets for data recording

transforming data (JRA)

Joachim-Raff-Archiv has started to record metadata with the objective of having TEI and MEI files available for further use. Because of the special conditions within the archive there had been no time for developing coding guidelines, training employees or developing a digital infrastructure prior to the beginning of the project.

- cleaning up the spreadsheets
- set up a transformation workflow
- Excel spreadsheets → raw structured XML
- XSLT-Transformation: raw XML → TEI
- Add to XML database (eXist-DB)

performing data

The app that creates the registers is XQL-based (eXist-DB app). The rendering of the individual data sets is executed with XSLT scripts. XSLT is not only used to display the metadata and the content, but also to link individual items to each other.

Verzeichnis der musikalischen Quellen

DETAILED VIEWS:

- rendering metadata and content
- automatic cross-linking
- searching for and viewing of digital copies
- linking to other contents

Brief von Willy Müller an Ludwig Baumann vom 15. März

OVERVIEWS:

- register (works, sources, letters, documents, persons, institutions, places)
- auto update when new data is entered (live)
- sorting data according to defined categories
- overview of the contained data records

OPPORTUNITIES:

- convenient navigation through the data
- first data analysis
- proofread of individual records
- visualization of relationships between data sets

Visions on the future

- user interface in several languages
- detailed search through the database
- possibility to download individual records (XML, MARC21, print-version)
- ... some other fancy stuff ...

Brückenschläge zwischen Musikwissenschaft und Informatik

**Theoretische und praktische Aspekte
der Kooperation**

**Herausgegeben von Stefanie Acquavella-Rauch,
Andreas Münzmay und Joachim Veit**

**Musikwissenschaft:
Aktuelle Perspektiven 3**

musiconn
für vernetzte Musikwissenschaft

Brückenschläge zwischen Musikwissenschaft und Informatik

Musikwissenschaft: Aktuelle Perspektiven

Bericht über die Jahrestagung der Gesellschaft für Musikforschung 2019 in Paderborn und Detmold

Herausgegeben von Rebecca Grotjahn und Nina Jaeschke

Band 3

Brückenschläge

zwischen Musikwissenschaft und Informatik

Theoretische und praktische Aspekte der Kooperation

Beiträge der Symposien zur Digitalen Musikwissenschaft

Osnabrück 2018 und Paderborn 2019

im Rahmen der Jahrestagungen der Gesellschaft für Musikforschung

In Verbindung mit der Fachgruppe Digitale Musikwissenschaft

herausgegeben von

Stefanie Acquavella-Rauch, Andreas Münzmay und Joachim Veit

Detmold: Musikwissenschaftliches Seminar der Universität Paderborn

und der Hochschule für Musik Detmold

2020



DOI: 10.25366/2020.87

Online-Version verfügbar unter der Lizenz: Urheberrecht 1.0,
<<https://rightsstatements.org/page/InC/1.0/?language=de>>

Bibliografische Information der Deutschen Nationalbibliothek

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über <http://dnb.d-nb.de> abrufbar.

Impressum

Redaktion: Stefanie Acquavella-Rauch, Andreas Münzmay und Joachim Veit

Satz: Nina Jaeschke und Joachim Veit

© Musikwissenschaftliches Seminar der Universität Paderborn und der
Hochschule für Musik Detmold 2020

INHALT

Rebecca Grotjahn, Nina Jaeschke Vorwort zu Band 1–3	IX
Stefanie Acquavella-Rauch, Andreas Münzmay, Joachim Veit Brückenschläge zwischen Musikwissenschaft und Informatik – Vorbemerkung	XI
KOLLABORATIONEN – KO-LABORATORIEN	
Reinhard Keil Der Computer als Denkzeug für hermeneutische Arbeit	3
Ulrich Konrad Philologie und Digitalität. Perspektiven für die Musikwissenschaft im Kontext fächerübergreifender Institutionen	27
Gudrun Oevel Infrastruktureinrichtungen in Forschungsprojekten – Spagat oder Chance?	35
Dennis Ried Erhebung, Transformation und Präsentation digitaler Forschungsdaten	41
Anna Neovesky, Frederic von Vlahovits IncipitSearch – Leitfaden zur Zusammenarbeit	47
Elisabeth Treydte Clara Schumann #digital. 40 Jahre Archiv Frau und Musik und der Start in die Digitalisierung	53
TEXT/DATEN/PROZESSE	
Christine Siegert Komponisten-Gesamtausgaben im digitalen Zeitalter: Perspektiven und Reflexionen am Beispiel Ludwig van Beethovens	61
Markus Neuwirth, Johannes Hentschel, Martin Rohrmeier Perspectives of Musical Corpus Studies: The Annotated Mozart Sonatas	77
Agnes Amminger, Franz Kelnreiter Leopold Mozarts „Gründliche Violinschule“. Zur Textcodierung und -präsentation einer digitalen Edition	83
Oleksii Sapov Algorithmische Automatisierung komplexer Notationsregeln in MEI-XML am Beispiel von Versetzungszeichen	91

Susanne Cox, Richard Sanger	
Digitale Fassungsvergleiche am Beispiel von Beethovens Eigenbearbeitungen	97
Agnes Seipelt	
Digitale Edition und Harmonische Analyse mit MEI von Anton Bruckners Studienbuch	105
Stefanie Acquavella-Rauch	
Musikalische Schaffensprozesse 2.0 – Inkorporation audiovisueller Medien der popularen Musik in Methoden der digitalen Edition	115
DIGITAL(ISIERT)E MATERIALITATEN	
Miriam Akkermann	
(Musik)Instrument (im) Computer	125
Daniel Futterer	
Herausforderungen bei der Kodierung von Paratext am Beispiel Neuer Musik mit Live-Elektronik	141
Matthias Pasdzierny	
How much is the glitch? Das digitale Paradigma als Herausforderung und Chance fur die historische Musikwissenschaft	149
Shintaro Miyazaki	
Musik fur Maschinen?! – Wo sich die Wissenschaft der Medien, des Computers und der Musik treffen und wie sie zusammenarbeiten konnten	173
MUSIKGESCHICHTE(N) IM NETZ	
Matthias Tischer	
Musikgeschichte der DDR: Ein Pilotprojekt zur digitalen Musikvermittlung	181
Annette van Dyck-Hemming, Jan Eberhardt, Melanie Wald-Fuhrmann	
Ansatze zur Analyse historischer Netzwerke mit Neo4j® – Aus der Projekt-Werkstatt der Datenbank zur Fachgeschichte der Musikwissenschaft	191
Axel Beer, Martin Bierwisch, Kristina Kramer	
Das MMM2 – Ein regionalgeschichtliches Onlinelexikon der Arbeitsgemeinschaft fur mittelrheinische Musikgeschichte	199
Matej Santi	
Was erzahlt Fritz Kreislers Geige?	207
Elias Berner	
Alle Menschen werden Bruder?! Ein historisches Dokument aus dem Nationalsozialismus in den sozialen Medien	211

Gabriele Buschmeier in memoriam