

Endstation Digital?! Herausforderung Metadaten und Nachhaltigkeit in musikwissenschaftlichen Datenbanken

Blanken, Christine

blanken@bach-leipzig.de
Bach-Archiv Leipzig, Deutschland

Rettinghaus, Klaus

rettinghaus@bach-leipzig.de
Sächsische Akademie der Wissenschaften zu Leipzig,
Projekt Bach-Repertorium

I.

Ausgangspunkt war eine reine Metadaten-Sammlung zu Werken und Quellen Johann Sebastian Bachs (1999-2008: „Göttinger Bach-Katalog“), die zum Abschluss der „Neuen Bach-Ausgabe“ am Johann Sebastian Bach-Institut Göttingen erfolgte. Dieses Metadatensammlung war 2008 die Basis für den Projektstart von „Bach digital“. Seitdem erfolgte mittels kontinuierlicher Förderung durch die DFG ein mehrstufig angelegter Ausbau:

Das erste Digitalisierungsprojekt umfasste die sogenannten Originalquellen zu Johann Sebastian Bachs Musik, also Autographen und originales Aufführungsmaterial Bachs, die sich zu etwa 90 % im Besitz der oben genannten Bibliotheken befinden. 2010 ging diese erste Stufe als www.bach-digital.de online.

Daran schloss sich von 2013 bis 2016 die Digitalisierung von sogenannten Sekundärquellen Bachscher Musik aus der Generation der Bach-Söhne und -Schüler an, ein Bestand, der besonders viel Tastenmusik J. S. Bachs umfasst, die vielfach nicht autograph überliefert ist und damit Forschungen zu individuellen Fassungen ermöglicht sowie Bachs Arbeitsweise in der Klavier- und Orgelmusik zwischen Kunstwerk und Unterrichtspraxis transparent zu machen hilft.

Mittlerweile wurde die dritte Stufe gezündet: die konsequente Ausweitung der Datenbank in Metadaten und Digitalisaten auf die Musik der Bach-Söhne im Projekt „Quellenkorpus Bach-Söhne – Erschließung und Digitalisierung der Primärüberlieferung zu Werken Wilhelm Friedemann, Carl Philipp Emanuel, Johann Christoph Friedrich und Johann Christian Bach sowie deren Einbindung in das zu erweiternde Portal Bach digital“.

Daneben werden Schritt für Schritt auch Werkverzeichnisse der Publikationsreihe „Bach-Repertorium“ sowie das derzeit neu erarbeitete „Bach-

Werke-Verzeichnis III“ integriert. Mittelfristig werden dazu auch musikalische Incipits implementiert bzw. suchbar gemacht.

Diese stufenweise Bearbeitung eines Kernbestandes der Musik des 18. Jahrhunderts strukturiert und systematisiert das gesamte musikalische Quellenmaterial und bildet den soliden Ausgangspunkt für eine quellenbasierte Forschung zur Musik der Bach-Familie: nicht nur als wichtiges Hilfsmittel der Bach-Forschung, sondern auch z. B. als Vergleichsobjekt für Studien zu anderen Repertoires, als unterstützendes Material für Forschungen zu Mitteldeutschland, als Kernbestandteil zur Provenienzforschung wichtiger Sammlungen wie Poelchau, Breitkopf etc. etc. Die Nutzungsmöglichkeiten sind in den vergangenen zehn Jahren stark angewachsen; und damit auf die Verantwortung, ein möglichst den diversen Anforderungen gerecht werdendes Material bestmöglich aufzubereiten.

Das MyCoRe-basierte Projekt wurde dabei von Anfang an durch eine Dokumentation begleitet, die es problemlos nachnutzbar macht: <https://www.bach-digital.de/content/documentation.xml?XSL.lastPage.SESSION=/content/documentation.xml>.

Abseits der Bach-Forschung bzw. Musikwissenschaft werden Daten und Digitalisate von „Bach digital“ auch von einer breitgefächerten Bach-Community gesucht: das sind die weltweit großen Nutzerkreise musikinteressierter Laien sowie auch Musiker, die z. T. direkt nach originalen Quellen-Digitalisaten musizieren. Das Spektrum der Nutzer weitet sich nach unserer Erfahrung mit statistischen Daten zur Datenbank: je divergenter das ins Netz gestellte Material, desto vielfältiger die Nutzung. Inwiefern eine ursprünglich für die Bach-Forschungscommunity entwickelte Datenbank dieser Entwicklung noch stärker Rechnung tragen soll, wäre zu diskutieren.

II.

Die Menge der Datensätze an sich (es sind Stand Januar 2018 immerhin 8230 Musik-Quellen zu 3870 Werken der Bach-Familie) sowie der Umfang der Metadaten innerhalb eines Datensatzes ist nur mit hohem personellem Aufwand auf dem neuesten Stand der Forschung zu halten. Digitalisate und Metadaten werden so gut es geht laufend überprüft, auch mithilfe von Nutzer-Feedbacks – besonders jenen für die Bach-Forschung so wichtigen Power-Usern aus aller Welt. Dies ist eine ständige Anforderung, die die Daten selbst stellen, sobald sie öffentlich sichtbar sind. An der Aktualisierung der Datensätze aufgrund von Neuerkenntnissen der Bach-Forschung sollen deshalb nun auch mehr Mitarbeiter in der Forschungsabteilung des Bach-Archivs beteiligt werden als es Projektmitarbeiter für „Bach digital“ gibt. Ziel ist es, der Veraltung von Forschungsdaten entgegenzuwirken. Das Bach-Archiv sieht sich hier in der Verantwortung, die einmal publizierten Forschungsdaten mit den „Bach digital“-Nutzern möglichst zu teilen. Hierzu

gehört auch die Mehrsprachigkeit, die derzeit nur mit Hilfe von strukturierten Daten umgesetzt werden kann. Fließtexte zu übersetzen ist mangels dafür vorhandener Projektmittel nur sehr begrenzt möglich. Alle anderen Daten sind aber mittlerweile auch in Englisch, Japanisch, Französisch (und Anfang 2018 auch Italienisch und Spanisch) recherchierbar. Hierbei sind wiederum die Nutzer der Datenbank selbst behilflich. Geplant ist als nächstes eine Nutzerbefragung, die über die Interessen und Wünsche sowie Kritik oder weitere Formen der Common Science-Beteiligung Auskunft geben soll. Inwieweit dieses Ergebnis zu einer Umstrukturierung von Daten oder der Präsentation von Modulen führen wird oder muss, ist derzeit noch offen.

Bei dieser prinzipiell optimistischen Sicht auf „Bach digital“ sollen weitere kritische Punkte nicht außer Acht gelassen werden, die aus dieser langjährigen Erfahrung mit den Metadaten resultieren:

Die Datenbank-Struktur suggeriert Eindeutigkeit, suggeriert, dass die Daten dem - in letzter Zeit in den Polit-Medien - so beliebten Faktencheck standhalten. Die Herkunft der Daten, gerade auch bei Neuerkenntnissen, wird dabei oft nicht präzise offengelegt. Die Datenbankstruktur suggeriert indes meist, dass es hier um Fakten geht. Unsicherheiten können nur sehr begrenzt formuliert werden, gerade im Fall von strukturierten Daten. Gerade auch die gegenüber den Printmedien so einfach zu handhabende Datenänderung ist also ein Problem für die Transparenz von Forschungsdaten.

Ein einfacher Daten-Austausch per Schnittstelle, sicher allgemein gewünscht und praktiziert, ist nur insofern dauerhaft praktikabel, so lange ermöglicht wird, dieses Procedere mehrfach zu wiederholen, gerade auch bei Richtigstellungen von Forschungsdaten. Ansonsten finden sich mehrere Versionen von Quellen- oder Werkdaten im Netz, die sicherlich unerwünscht sind, selbst wenn man mit Versionierungsangaben arbeitet.

III.

Immer mehr Datenbanken zu Musikern und musikalischen Quellen tummeln sich im Netz. Doch selbst wenn es inhaltliche Überschneidungen gibt, arbeiten sie zumeist aneinander vorbei. Dabei ist das größte Problem nicht einmal die Vergeudung von Ressourcen, sondern die prinzipielle Unmöglichkeit eines Datenaustauschs bzw. einer einfachen Nachnutzbarkeit der Metadaten – selbst wenn sie in den sogenannten „Quasi-Standards“ TEI oder MEI vorliegen. Bei öffentlich geförderten Projekten ist heutzutage Grundvoraussetzung, dass die „langfristige Sicherung von“ und der „grundsätzlich offene Zugang zu“ Forschungsdaten gewährleistet sein muss, es aber bislang unklar ist, was genau dies heißt. Ist der Zugang schon „offen“ wenn man die Informationen im Internet finden kann, oder erst dann, wenn sie über eine Schnittstelle bereitgestellt werden? Solange Forschungsprojekte nur

„digitale Inseln“ errichten, bringt das „Digitale“ keinen wirklichen Mehrwehrt.

Zwar existieren bereits verschiedene Formate, die speziell für den Datenaustausch gedacht sind, wie z. B. MARC21 und METS/MODS, doch sind diese nur sehr eingeschränkt für (musikwissenschaftliche) Forschungsprojekte und Datenbanken einsetzbar. Auch spezielle Ontologien stehen bereit, die vom W3C zu den „Good Ontologies“ gezählt werden, also Ontologien, die vollständig dokumentiert, dereferenzierbar, von unabhängigen Datenlieferanten verwendet und möglicherweise von bestehenden Tools unterstützt werden („ontologies that are fully documented, dereferenceable, used by independent data providers and possibly supported by existing tools“). Beispiele dafür sind „*Dublin Core*“ und „*The Music Ontology*“. Doch auch hier bleibt das Problem, dass diese Formate zu flexibel, zu schwammig gestaltet sind, um einen sinnvollen, nachvollziehbaren Datenaustausch zu gewährleisten, oder aber spezielle Forschungs-Erkenntnisse nicht hinreichend darin abgebildet werden können – ganz abgesehen davon, dass derlei Lösungen überhaupt erst einmal implementiert werden müssen. Das vielgepriesene RDF, das versucht, einige dieser Probleme zu lösen (oder zu umschiffen), kann dabei kein Selbstläufer sein.

Auch können Projekte a-priori nicht immer vorhersehen, welche Daten genau anfallen werden, bzw. welche von anderen Forschern oder Projekten nachgenutzt werden könnten. Es ist also nicht unbedingt zielführend, Daten in allen möglichen Formaten anbieten zu wollen, selbst wenn die Ressourcen es gestatten verschiedene Daten-Export-Möglichkeiten bereitzustellen (und zu pflegen).

Können RESTful APIs die Lösung aller Probleme sein? Diese ermöglichen es zwar, sehr spezielle Kombinationen aus Metadaten zusammen zu stellen. Dennoch bleibt das Problem der intern verwendeten Formate bestehen; beschreibt ein Feld „date“ ein Aufführungsdatum oder das Datum der Werkgenese?

Um digitale Gräber zu verhindern, sind spezielle, klar definiert und strukturierte Datenformate vonnöten, die für klar definierte Anwendungsfälle einen echten Austausch ermöglichen und somit auch erstmals dezentrale Suchmaschinen ermöglichen. Solche Suchmaschinen können abseits von Google überhaupt erst wirkliche Interdisziplinarität herstellen, denn mit wachsender Zahl an digitalen Projekten – so begrüßenswert dies auch sein mag – steigt die Gefahr, dass man „den Wald vor lauter Bäumen nicht sieht“, also Ergebnisse anderer (vielleicht fachfremder) Projekte nicht wahrnimmt, und dadurch möglicherweise den eigenen Erkenntnisprozess behindert.

Der aktuelle Umgang mit gesammelten Metadaten soll am Beispiel von „Bach digital“ gezeigt sowie mögliche Auswege skizziert und diskutiert werden. Vorgestellt werden dabei standardisierte Formate, die bereits heute den Informationsaustausch und -fluss ermöglichen und aufzeigen, was dadurch zukünftig möglich sein könnte, aber auch, wo die größten Lücken und dringendsten Desiderate bislang bestehen blieben.