

# Daten auf Wanderschaft

## Migration von Daten als Herausforderung der Digital Humanities am Beispiel von MUSICI und MusMig

Torsten Roeder

### Einleitung

Migration bringt für den Wandernden – oder auch: *das* Wandernde – stets eine Veränderung des Umfelds mit sich. Kontakte und Reibungen mit der neuen Umgebung entstehen und Assimilationsprozesse werden in Gang gesetzt. Das Ergebnis solcher Prozesse ist kaum vorhersehbar: Harmonie und Ausgewogenheit sind dabei ebenso gut möglich wie Konflikte und Differenzen.

Sowohl in den Geisteswissenschaften als auch in der Informatik ist „Migration“ – jeweils mit einer ganz eigenen Ausprägung des Begriffs – zu einem relevanten Thema avanciert. Der Schwerpunkt der folgenden Untersuchung liegt auf einem Überschneidungsbereich der beiden Disziplinen, in dem sich derzeit das Feld der Digital Humanities etabliert.

Diese spezifische Verortung der Digital Humanities, in denen geisteswissenschaftliche und informatische Fragestellungen zusammenfließen, soll in den folgenden Ausführungen veranschaulicht werden. Exemplarisch werden dafür zwei Projekte skizziert, in denen das Thema Migration gleich in zweifacher Hinsicht, auf beiden der oben genannten Ebenen, zum Tragen kommt.

### Begriffliche Grundlagen

#### Der Begriff der Migration in den Geisteswissenschaften

In den Geisteswissenschaften versteht man unter „Migration“ ganz allgemein die Bewegung von Individuen oder ethnologischen Gruppen im geographischen Raum; im weiteren Sinne auch die Wanderung von Gegenständen oder Ideen; und im weitesten Sinne auch „innere“ Bewegungen im menschlichen Denken und Schaffen. Am Ende eines Migrationsweges steht das Subjekt dabei stets in einer neuen Umgebung und erzeugt mit dieser Wechselwirkungen. Die Geisteswissenschaften analysieren Ursachen und Wirkungen dieses Phänomens sowohl in einzelnen Fallstudien als auch zusammenfassend im Hinblick auf größere Kulturströmungen. So hat sich Migration zu einem wichtigen sozialhistorischen Forschungsthema entwickelt, welches

insbesondere am Beispiel der Musik – als stets unmittelbarer und sprachunabhängiger Ausdruck von Identität – seine besondere Ergiebigkeit darin bewies, an ihm Kulturströmungen aufzuzeigen und charakteristisch nachzeichnen zu können.<sup>1</sup>

## Der Begriff der Migration in der Informatik

In den Informationswissenschaften hingegen steht „Migration“ vor allem für die Übertragung von Daten aus einem System in ein anderes System. Auch hier wird der Begriff „Umgebung“ verwendet, der sich in diesem Kontext vor allem auf die informationstechnischen Vorgaben (etwa einer bestimmten Plattform) bezieht. Meist projekt- oder auch geschäftsorientiert ist Migration in der IT ein fast durchwegs praxisbezogenes Betätigungsfeld, in dem Aspekte der Theoriebildung eher im Hintergrund stehen. Erst bedingt durch über mehrere Generationen von Betriebssystemen und Anwendungsprogrammen hinweg erfahrenen Konsequenzen einer unzulänglichen Datenübertragbarkeit konnte die Migration von elektronisch gespeicherten Daten auch die erhöhte Aufmerksamkeit der theoretischen Informatik gewinnen. Ein bereits intensiv bestelltes Feld ist die Migration von Software,<sup>2</sup> welches an dieser Stelle jedoch nicht näher einbezogen werden soll – hier geht es ausschließlich um die Migration von Daten.

Levine 2009 beschrieb das Datenmigrationsproblem wie folgt:

*„Data migration is the process of making a copy of data and moving it from one device or system to another, preferably without disrupting or disabling active business processing.“*

Sieht man von dem Fokus auf Geschäftsvorgänge ab – an dieser Stelle könnte schließlich ebenso gut eine wissenschaftliche Datenverarbeitung stehen – wird hier ein Kernprinzip der Migration erkennbar: Objekte aus einer Ursprungsumgebung müssen zukünftig in einer neuen Umgebung funktionieren. Das Zitat impliziert, dass dieser Prozess nicht zwingend reibungslos verläuft („preferably without disrupting“). Es muss stets mit Inkonsistenzen und Anpassungsbedarf gerechnet werden. Die zu migrierenden Daten müssen daher nicht nur strukturell transformiert, sondern gegebenenfalls auch inhaltlich angepasst werden.

## Datenmigration als Aufgabenfeld der Digital Humanities

Die Verbindung der beiden Migrationsverständnisse hat in zwei Forschungsprojekten konkret Gestalt angenommen, die sich inhaltlich mit der Migration von Musikern im Europa des 17. und 18. Jahrhunderts befassen. Das erste Projekt, MUSICI<sup>3</sup> (2009–2013), widmete sich Musiker-Migrationen nach Italien; das nachfolgende Projekt MusMig<sup>4</sup> (2013–2016) konzentriert sich hin-

<sup>1</sup>Für eine breite musikhistorische Übersicht siehe z.B. Ehrmann-Herfort/Leopold 2013; zu politisch verursachten Migrationen des 20. Jh. siehe z.B. Brinkmann/Wolff 1999 und Davis et al. 2011 (hier Part III, Traveling Sound: Music and Migration); zu zeitgenössischen globalen Strömungen siehe z. B. MAiA 2011/2. Diese Auswahl ist exemplarisch und nicht repräsentativ.

<sup>2</sup>Zur Problematik von *Legacies* (Altsystemen) vgl. z. B. Wagner 2014.

<sup>3</sup>*Musicisti europei a Venezia, Roma e Napoli (1650-1750): musica, identità delle nazioni e scambi culturali*, gefördert durch die Agence Nationale de la Recherche und die Deutsche Forschungsgemeinschaft von 2010–2012, <http://www.musici.eu/> (24.05.2016).

<sup>4</sup>*Music Migrations in the Early Modern Age: the Meeting of the European East, West and South*, gefördert durch Humanities in the European Research Area von 2013–2016, <http://musmig.eu/> (24.05.2016).

gegen auf Migrationsbewegungen zwischen Süd-, Mittel- und Osteuropa. Die MUSICI-Datenbank ging mit dem Projektabschluss im Jahr 2013 in Betrieb.<sup>5</sup> Die Datenbank des Nachfolge-Projektes MusMig basiert technisch auf dem gleichen System wie die MUSICI-Datenbank und soll zukünftig die Daten des MUSICI-Projekts als eine von mehreren Teildatenmengen inkludieren. Dafür müssen die Daten des MUSICI-Projekts in das System der MusMig-Datenbank migriert werden; während die zugrunde liegenden Datenstrukturen auf der technischen Ebene miteinander kompatibel sind, unterscheiden sie sich jedoch inhaltlich in vielerlei Hinsicht. Das gemeinsame Thema der beiden Projekte gab den Anstoß, sich in Analogie zu natürlichen Migrationsstrukturen den inhaltlichen Konsequenzen der Datenübertragung anzunähern. Aus diesen und vergleichbaren Vorhaben ergeben sich die für die Digital Humanities spezifischen Problemstellungen.

Vernachlässigt man den rein praktischen Aspekt der Datenmigration und setzt voraus, dass zwei Umgebungen zumindest von der technischen Seite her miteinander kompatibel sind, öffnet sich die Perspektive auf die zahlreichen inhaltlichen Herausforderungen: Wie verändern sich Informationen<sup>6</sup> in neuen Umgebungen? Wie wirkt sich zum Beispiel abweichendes Vokabular der neuen Umgebungen auf die migrierten Informationen aus? Sind in der neuen Umgebung bereits Daten vorhanden? Werden einheitliche Referenzsysteme genutzt?

Diese Fragen können bei jedem Datenmigrationsprozess erneut auftreten. Durch die aktuelle Forschungslandschaft, die sich zum einen durch eher kurzfristige Projekte und zum anderen durch eine relativ offene Verfügbarkeit von Daten (etwa im Rahmen von Open Access) definiert, bleiben Daten stets der Möglichkeit einer Nachnutzung durch Folgeprojekte oder Dritte ausgesetzt. Dies ist wissenschaftlich gesehen wünschenswert, bringt aber gleichzeitig Aufgaben mit sich.

Eine weitere Form von Datenmigration thematisiert Bernhard Thalheim im Zusammenhang mit der „Evolution von Systemen“, das heißt in diesem Falle: mit dem Anstieg der Funktionalität.<sup>7</sup> Grundannahme ist, dass die Ansprüche an die Funktionalität eines Systems sich mit der Zeit steigern und sich Systeme entsprechend weiterentwickeln müssen. Dies kann etwa durch eine Folge von Ausbaustufen geschehen. Einen deutlich höheren Anstieg der Funktionalität, als es die allmähliche Erweiterung eines bestehenden Systems bewirken könne, bringe der Wechsel eines Systems mit sich. Der Wechsel aber erfordere zwangsläufig die Migration der bereits bestehenden Daten. Als zentrale und zu berücksichtigende Problempunkte werden unter anderem genannt: Die divergente Interpretierbarkeit der Daten in den verschiedenen Systemen, die damit verbundenen unterschiedlichen Methodiken und Mittel, mit denen die Daten erfasst werden und schließlich die Änderung der Spezifikationen von System zu System. All dies ist unabhängig davon, ob die Datenmigration zwischen zwei voneinander gänzlich unabhängigen Systemen erfolgt oder ob sie im Rahmen eines größeren Versionssprungs geschieht.

Das abstrakte Konzept der Evolution von Systemen ist allerdings nicht einseitig linear zu verstehen. Ein Anstieg von Funktionalität kann ebenso auch Verluste in der Leistungsfähigkeit eines Systems mit sich bringen. Ein neues System wird – und soll – nur sehr selten in der Lage sein,

<sup>5</sup>Musicisti europei a Venezia, Roma e Napoli (1650-1750), hrsg. von Michela Berti, Gesa zur Nieden und Torsten Roeder, Berlin/Rom 2013, <http://www.musici.eu/database> (24.05.2016).

<sup>6</sup>Zur begrifflichen Abgrenzung: „Daten“ wird hier unspezifisch aufgefasst und bezeichnet eine undefinierte Menge von Datenobjekten; „Informationen“ hingegen meint hier konkretisierbare, einzelne Datenobjekte mit Aussagegehalt.

<sup>7</sup>Vgl. Thalheim/Wang 2011, 2013; dazu auch Seacord et al. 2003; Klettke 2011.

das bisherige System hundertprozentig zu inkludieren: Systemwechsel sind häufig auch konzeptuell motiviert und dienen nicht nur der Leistungsoptimierung. Analog dazu steht auch in den Geisteswissenschaften einer reinen Anhäufung und Vermehrung von Wissen eine immer wieder neue Kontextualisierung und Reinterpretation der Wissens Elemente entgegen. In den Digital Humanities verbindet sich beides unter anderem auf der Ebene der Datenmodellierung: Hier fließen informationstechnische und geisteswissenschaftliche Konzepte zusammen.

Migrationen von Forschungsdaten sind daher nicht als rein technische Aufgabe zu verstehen, sondern sie müssen von einem wissenschaftlichen Prozess begleitet werden, welcher die Konsequenzen der konzeptionellen Veränderungen in Datenstrukturen transparent werden lässt.

## Das Zusammenspiel von Daten, Kontext und Narration

### Datenbanken und ihre „Weltsichten“ – ein Beispiel

Durch eine Gegenüberstellung mehrerer musikwissenschaftlicher digitaler Ressourcen soll im Folgenden die konzeptuelle Divergenz von Datenbanksystemen exemplifiziert werden. Als Beispiel dienen drei inhaltlich und strukturell sehr unterschiedliche Kurzbiographien des Komponisten Carl Ditters von Dittersdorf. Der Fokus soll, in Anlehnung an das MUSICI-Projekt, auf die Beschreibung seines Verhältnisses zu Italien gerichtet werden. Herangezogen werden das *Österreichische Musik-Lexikon* (ÖML), das *Bayerische Musiker-Lexikon Online* (BMLO) und *Musica Migrants*.

Das ÖML bringt in dem Artikel keinen einzigen Hinweis auf eine Beziehung Dittersdorfs zu Italien.<sup>8</sup> Das BMLO hingegen führt unter dem Schlagwort „Wirkungsorte“ die Städte Bologna, Mantua, Parma, Triest und Venedig.<sup>9</sup> Musica Migrants nennt dagegen eine Konzertreise nach Italien im Jahr 1763 mit Christoph Willibald Gluck.<sup>10</sup> Diese sehr unterschiedliche Auswahl der Informationen hängt sowohl mit der Struktur als auch mit der inhaltlichen Ausrichtung der Datenbanken zusammen.

Das ÖML ist nicht ausschließlich biographisch ausgerichtet und als ursprüngliches Printprodukt vom Umfang her begrenzt, weshalb die Informationen dort auf das für Österreich Relevante konzentriert sind. BMLO und Musica Migrants können hingegen als native Datenbankprojekte mehr Informationen aufnehmen, und so findet sich hier trotz der jeweiligen Interessengebiete (Bayern beziehungsweise Ost- und Mitteleuropa) in beiden Fällen ein Hinweis auf Dittersdorfs Italienaufenthalte. Dabei steht im BMLO eine schlagwortartige Struktur im Vordergrund, welche eine Vielzahl von Wirkungsorten auflistet, jedoch ohne Gewichtung und Kontext. Der Eintrag in

<sup>8</sup> *Österreichisches Musik-Lexikon. Online-Ausgabe*, Wien: Österreichische Akademie der Wissenschaften, 2002–2013, <http://www.musiklexikon.ac.at/> = Österreichisches Musiklexikon, Wien: Verlag der Österreichischen Akademie der Wissenschaften, 2002–2006. Dort: *Ditters von Dittersdorf, Johann Carl*, [http://musiklexikon.ac.at/0xc1aa500d\\_0x0001cbcb](http://musiklexikon.ac.at/0xc1aa500d_0x0001cbcb) (24.05.2016).

<sup>9</sup> *Bayerisches Musiker-Lexikon Online*, hrsg. von Josef Focht, München: Ludwig-Maximilians-Universität, 2004–2014, <http://bmlo.de/>. Dort: *Dittersdorf, Karl Ditters von*, <http://bmlo.de/d0473> (24.05.2016).

<sup>10</sup> *Musica Migrants*, Universität Leipzig, <http://www.musicamigrants.de/>. Dort: *Ditters von Dittersdorf, Carl*, <http://www.musicamigrants.de/pages/individual/person.php?id=topic-233658> (24.05.2016).

Musica Migrans hingegen – eine aus Einzelinformationen zusammengestellte Liste mit semantischem Markup – erklärt die genaueren Umstände eines Italienaufenthalts, allerdings ohne die konkreten Orte zu nennen.

Als Gesamteindruck bleibt, dass in jeder Datenbank eine andere Sichtweise auf Dittersdorfs Italienaufenthalte repräsentiert ist. Aus der Zusammenschau der Datenbanken wird nicht deutlich, ob die in Musica Migrans erwähnte Reise tatsächlich mit den im BMLO genannten Stationen (oder einigen davon) in Zusammenhang steht. Es ist ohne weitere Recherche nicht möglich, die Informationen miteinander in Verbindung zu bringen.<sup>11</sup> Strukturen und Inhalte der Datenbanken sind so divergent, dass sie als miteinander nicht kompatibel erscheinen.

## Die Kontextualisierung von Daten

An dem Beispiel wurde ersichtlich, dass eine Information aus einer Datenbank im Kontext einer anderen Datenbank sowohl die eigene Aussagekraft als auch die der anderen Daten relativiert. Diese Relativität mag problematisch erscheinen, und doch ist sie wichtig: Zwar könnte eine Information prinzipiell auch ohne ihren Kontext erschlossen werden, denn bereits in rein textueller Form liefert sie eine Aussage. Allerdings ließe sich daraus ihre Relevanz nicht bewerten: Diese wird erst durch eine Einordnung in Zusammenhänge erkennbar, wenn also ersichtlich wird, wie sich die Information im Zusammenspiel mit anderen Informationen verhält. Die Möglichkeiten der Kontextualisierung sind fast unbegrenzt variierbar.

So taucht etwa die Information, dass Dittersdorf mit Gluck auf Konzertreise in Bologna war, zwar häufig in den Biographien Dittersdorfs auf, viel seltener aber in denen über Gluck, so dass die Relevanz einer Information offensichtlich mit ihrem jeweiligen Kontext variiert. Angenommen, sie würde in der Biographie Glucks auftauchen, könnte dies sowohl die Wahrnehmung Glucks als auch von Dittersdorf verändern.

Der Kontext entsteht bereits während der Auswahl und Zusammenstellung der Daten, die stets durch eine bestimmte Perspektive und ein vorher festgelegtes Ziel determiniert sind. Wissenschaftlich und methodisch ist dies von großem Interesse, da der Erkenntnisgewinn mehr durch die Modellierung und weniger durch beliebiges Sammeln von Daten befördert wird.<sup>12</sup> In den oben erwähnten Beispielen BMLO und ÖML bestünde dieses in der Fokussierung auf die Bedeutung des Komponisten Dittersdorf für Bayern respektive Österreich.

Umgekehrt gesehen bewirkt schon die Zuordnung einer Datenmenge zu einem bestimmten Vorhaben eine Kontextualisierung. So ist aus der Tatsache, dass Dittersdorf im BMLO und im ÖML vertreten ist, ableitbar, dass der Lebensweg des Komponisten offenbar in einer Verbindung zu Bayern und Österreich stand; ebenso ist damit jede einzelne Information aus der Biographie vor

<sup>11</sup>Die etwas ausführlichere Deutsche Biographie (ADB/NDB) klärt, dass es sich um mehrere Aufenthalte handelte. Vgl. Günter Birkner: *Ditters von Dittersdorf, Carl*. In: Neue Deutsche Biographie 4 (1959), S. 1–2, <http://www.deutsche-biographie.de/ppn118679856.html> (24.05.2016); J. Maehly: *Dittersdorf, Karl von*. In: Allgemeine Deutsche Biographie 5 (1877), S. 262–266, ebd.

<sup>12</sup>In diesem Zusammenhang kann eine Betrachtung von „Big Data“ lohnen, allerdings ist dies für die vorliegende Untersuchung nachrangig. Während „Big Data“ aufgrund der übermäßigen Menge den Inhalt oder die Struktur der Daten nicht im Einzelnen berücksichtigt und auf alternative Methoden der Auswertung zielt, überwiegt in geisteswissenschaftlich ausgerichteten Datenbanken die Bedeutung der konkreten konzeptionellen Anlage. Christof Schöch unterschied in diesem Sinne zwischen „Big Data“ und „Smart Data“ und plädiert für eine Symbiose beider Qualitäten: „only smart big data enables intelligent quantitative methods“ (Schöch 2013).

diesem Hintergrund interpretierbar. Auf diese Weise bleiben Daten dem Kontext ihrer Entstehung verbunden.

Insofern ist das Erfassen von Daten im Idealfall ein von wissenschaftlicher Erkenntnisabsicht geleiteter Prozess, der stets die spezifische Ausrichtung eines Vorhabens transportiert. Die Zielperspektive der Datenerfassung wohnt damit den Daten inne und bestimmt die Wertigkeit der einzelnen Informationen.

Bei einer Migration von Daten entsteht somit zum einen unweigerlich eine Vermischung ursprünglich differierender Kontexte, zum anderen verschiebt sich die Relevanz der Informationen untereinander und die Integrität des Datenkorpus verwässert. Daher ist bei einer Migration der ursprüngliche Kontext der Daten stets zu berücksichtigen.

## Narratives Potenzial und spezifische Terminologien

Der theoretische Ansatz des Personendaten-Repositoriums,<sup>13</sup> einer digitalen Infrastruktur für Personendaten (auf der sowohl MUSICI als auch MusMig basieren), geht davon aus, dass beim biographischen Arbeiten durch Prozesse der Informationsselektion spezifische „Narrationen“ kreiert werden.<sup>14</sup> Durch Veränderung der Sichtweise auf die Daten sind wiederum neue Narrationen erzeugbar: So könnten prosopographische Daten etwa nach chronologischen, geographischen oder sachbezogenen Aspekten geordnet werden. Auch die Auswahl müsste sich nicht zwangsläufig auf die Person als zentrales Narrationsobjekt beschränken, sondern könnte ebenso auch von einer bestimmten Berufsgruppe ausgehen und alle dazu verfügbaren Informationen zu einer neuen Narration formen.

Beispielsweise könnten aus einer Musiker-Datenbank alle Informationen zur Person „Dittersdorf“ extrahiert und zu einer chronologischen Biographie des Komponisten angeordnet werden. Dieses wäre die klassische biographische Narration. Eine tiefe semantische Erschließung der Daten vorausgesetzt, könnten der Datenbank auch alle Informationen zum Ort „Bologna“ entnommen und zu einer lokalen Musikgeschichte zusammengestellt werden (selbstverständlich unter der Berücksichtigung des allgemeinen Erfassungskontextes der Daten). Die Auswahl der Daten kann anhand der Datenbank-Systematik parametrisiert werden, beispielsweise auf Klavierbauer in Wien<sup>15</sup> oder auf Aktivitäten französischer Sänger in Rom zwischen 1650 und 1750.<sup>16</sup> Diesem Prinzip folgend ist anzunehmen, dass jede Datenbank ein fast unerschöpfliches Potenzial von Narrationen birgt, welches lediglich durch die Tiefe der semantischen Erschließung und durch die Zugriffsmöglichkeiten limitiert ist. Hierin liegt das „narrative Potenzial“ einer Datenbank.

Hinter einer semantischen Erschließung verbergen sich in der Regel hochspezialisierte Fachterminologien. So ist die Suche nach „Lautenmachern“ oder „Kastraten“ am ehesten in musikwissenschaftlichen Datenbanken möglich, während sich Termini wie „Zellulärpathologie“ oder

<sup>13</sup>Personendaten-Repositorium, <http://pdr.bbaw.de> (24.05.2016).

<sup>14</sup>„Biographisches Arbeiten ist also ein konstruierender Prozess, der aus der Mannigfaltigkeit an vorhandenen Informationen und Daten schöpft“ (Walkowski 2011).

<sup>15</sup>Beispielrecherche im BMLO: [http://bmlo.de/Q/Musikalische\\_Tätigkeit=Klavierbauer/Wirkungsort=Wien](http://bmlo.de/Q/Musikalische_Tätigkeit=Klavierbauer/Wirkungsort=Wien) (24.05.2016).

<sup>16</sup>Beispielrecherche in der MUSICI-Datenbank: [http://www.musici.eu/?id=90&content=franc\\*&contentTag=provenienza&place=Roma&placeTag=Permanenza](http://www.musici.eu/?id=90&content=franc*&contentTag=provenienza&place=Roma&placeTag=Permanenza) (24.05.2016).



„Verein sozialistischer Ärzte“ vorrangig in medizinhistorischen Datenbanken anwenden lassen. Diese meist streng kontrollierten Vokabulare spiegeln die „Weltordnung“ einer Datenbank und spielen gleichzeitig eine bedeutende Rolle für deren narratives Potenzial. Bei einer Vermischung von Daten infolge einer Migration kommt es zum einen zu einer Veränderung des narrativen Potenzials (es wird größer, aber auch beliebiger), zum anderen kommt es gegebenenfalls zu einer Vermengung der Terminologien, sofern diese nicht miteinander kompatibel sind oder angepasst werden. Diesem Umstand ist bei einer Datenmigration Rechnung zu tragen.

## Konsequenzen für die Datenmigration

In Datenbanken spielen die Parameter Kontext, Datenkorpus und Terminologie bedeutende Rollen und müssen bei einer Datenmigration wie folgt berücksichtigt werden:

- Der abstrakte Kern der Information – also basale Parameter wie Zeit, Raum und Identitäten – dürfen nicht verändert werden.
- Der allgemeine Kontext einer Datenbank bestimmt die Erfassungsvorgänge und damit auch die spätere Einordnung der Daten. Es muss daher transparent bleiben, in welchem Kontext die Daten ursprünglich erfasst wurden.
- Das Datenkorpus als Ganzes bestimmt die Relevanz einzelner Informationen. Bei einer Datenmigration treffen die Parameter der Ausgangs- und Zieldatenbank aufeinander und beeinflussen sich in einem wechselseitigen Prozess: Nicht nur die Daten der Ausgangsdatenbank verändern sich in ihrer neuen Umgebung, sondern auch die Zieldatenbank verändert sich durch eine Integration neuer Daten. Daher ist dafür zu sorgen, dass die Integrität der jeweiligen Korpora erhalten bleibt.
- Die spezifischen Terminologien von Datenbanken repräsentieren deren jeweilige „Weltordnung“ und beeinflusst die Nutzbarkeit auf semantischer Ebene. Sie müssen bei einer Migration den neuen Strukturen so angepasst werden, dass sie in mit den neuen Daten zusammen interagieren können und dass gleichzeitig die ursprünglichen Strukturen nachvollziehbar und nutzbar bleiben.

Der letzte Punkt – die Umstellung von Terminologien – bildet das komplexeste Feld und soll daher eine kurze Ausführung erfahren. Hier ist ein Mapping zu entwickeln, welches die alte Terminologie (T1) auf die neue Terminologie (T2) abbildet. Sofern nicht in beiden Datenbanken ein ontologisches Referenzsystem wie zum Beispiel CIDOC-CRM<sup>17</sup> verwendet wird, welches einen automatischen Abgleich ermöglichen würde, muss ein neues Mapping entworfen werden. Dabei können zwischen den einzelnen Begriffen der Terminologien folgende Differenzen vorliegen:

- Mehrere Termini in T1 entsprechen einem Terminus in T2 (Generalisierung). Die betroffenen Daten müssen zusammengelegt werden.
- Ein Terminus in T1 entspricht mehreren Termini in T2 (Spezialisierung). Die betroffenen Daten müssen inhaltlich analysiert werden, um eine Aufteilung vorzunehmen.

<sup>17</sup>CIDOC Conceptual Reference Model, <http://cidoc-crm.org/> (24.05.2016).

- Ein Terminus in T1 entspricht nur teilweise einem Terminus in T2 (Inkongruenz). Die betroffenen Daten müssen inhaltlich analysiert werden, um Aufteilungen und Zusammenlegungen vorzunehmen.
- Ein Terminus in T1 entspricht keinem Terminus in T2 (Exklusion). Es muss entschieden werden, ob und in welcher Form die betroffenen Daten erhalten bleiben sollen.

Insgesamt können Datenmigrationen eine Vielzahl von inhaltlichen Auswirkungen auf verschiedensten Ebenen bewirken. Um den möglichen Konsequenzen eines Migrationsvorgangs eine höhere Transparenz zu verleihen, empfiehlt sich eine Dokumentation der Datenherkunft in der Zieldatenbank, die auch dem Nutzer kommuniziert wird. Dabei sollten Kontext, Korpus und Terminologie der Ausgangsdatenbank erkennbar bleiben. Eine undokumentierte Integration von Daten würde eine Verzerrung der Informationen bedeuten, die möglicherweise die Datenkorpora wissenschaftlich wertlos und unbrauchbar macht. Daher ist vor einer Datenübernahme zu untersuchen, welche Folgen diese haben könnte und welche Aspekte dokumentiert werden müssen, um die Nutzbarkeit der Daten nicht zu gefährden. Eine Datenmigration wird kaum ohne Veränderungen der Ursprungsdaten möglich sein. Im besten Falle würden die Daten sowohl in dem alten als auch in dem neuen Kontext gleichermaßen gut operieren und innerhalb des neuen Kontextes als Untermenge abgrenzbar bleiben.

Die genannten Punkte berühren mehrere Kernaspekte des Migrationsbegriffes: Die Einpassung in die neue Umgebung und das Ablegen alter Strukturen, das Einbringen eigener Strukturen in die neue Umgebung und die Bewahrung von Integrität der Daten. Migrierende Daten sind dabei nicht selbständig wie Individuen, jedoch begeben auch sie sich gewissermaßen auf Wanderschaft und gelangen in einen im übertragenen Sinne vergleichbaren, wechselseitigen Integrationsprozess.

## Von MUSICI nach MusMig

Anhand des Fallbeispiels der Datenmigration von MUSICI nach MusMig sollen die theoretischen Ansätze einem konkreten Anwendungsfall gegenübergestellt werden. Die beiden Projekte liegen grundsätzlich aufgrund ihrer Thematik, ihrer zeitlichen und räumlichen Abdeckung und ihrer Methodik nah beieinander. Der Fokus liegt beiderseits auf Kulturprozessen, die durch Wanderungsbewegungen von Musikern angetrieben werden. Die Datenbanken enthalten vorrangig biographische Informationen zu Musikern und dokumentieren die beruflichen Aktivitäten und Beziehungen der Musiker an verschiedenen Orten. Kontrollierte Vokabulare und Referenzsysteme helfen bei der systematischen Erfassung von Orten, Institutionen, Berufen, Beziehungen und Sachbegriffen.

### Datenmodell

Beide Datenbanken basieren auf der Architektur des Personendaten-Repositorys der BBAW, deren Datenstruktur den folgenden strukturellen Regeln gehorcht:<sup>18</sup>

<sup>18</sup>Personendaten-Repository: Datenmodellierung, <http://pdr.bbaw.de/projekt/vorstellung/datenmodellierung> (24.05.2016).



- *Aspekte und Personen*: Die basale Dateneinheit besteht in diesem Modell aus einer einzelnen Information zu einer Person, hier Aspekt genannt. Es kann sich dabei sowohl um Lebensereignisse (wie Geburt, Heirat, Ausbildungsabschluss) als auch um Personeneigenschaften handeln (wie Namen, Zugehörigkeiten). Jeder Aspekt ist mindestens einer Person zuzuordnen.
- *Markup*: Die Information jedes einzelnen Aspektes kann mithilfe eines Markup-Vokabulars semantisch angereichert werden. Möglich sind Verweise auf Orte, Personen, Körperschaften, Sachbegriffe und kalendarische Daten. Das Vokabular kann, abgesehen von wenigen Basisbegriffen, frei definiert werden. Es ist hierarchisch in vier Ebenen gegliedert, wobei die unteren, detaillierteren Ebenen die oberen, allgemeineren inkludieren.
- *Kategorien*: Jeder Aspekt muss mindestens einer biographischen Kategorie (zum Beispiel Werdegang, Familienereignisse, Reisen) zugeordnet werden. Die Kategorien können, abgesehen von wenigen Basiskategorien, durch ein Projekt frei definiert werden.
- *Quellen*: Jeder Aspekt muss mit einer bibliographischen Angabe belegt werden.
- *Projekte*: Eine Instanz des Repositoriums kann beliebig viele Projekte enthalten. Aspekte, Personen und Quellen gehören jeweils zu genau einem Projekt.

Aufgrund der identischen Systeme erscheint eine Migration der Daten von MUSICI nach MusMig technisch gesehen relativ leicht.

## Betroffene Bereiche

Schwerwiegender sind die inhaltlichen Konsequenzen der Datenmigration, denn trotz sehr ähnlicher Ausgangsbedingungen bestehen zwischen den Projekten Unterschiede, welche eine Übernahme der Daten nicht ohne Modifikationen ermöglichen. Die Unterschiede schlagen sich in diesen Bereichen nieder:

- Projektkontext
- Erfassungssprache
- Biographische Kategorien
- Kontrollierte Vokabulare
- Referenzsysteme

In den folgenden Abschnitten wird exemplarisch untersucht, welche Maßnahmen konkret durchgeführt werden müssten, um ein harmonisches Zusammenspiel beider Datenmengen zu gewährleisten. Abschließend wird behandelt, welche Konsequenzen aus diesen Erkenntnissen für nachfolgende Datenbankprojekte zu ziehen sind, für die eine Integration der MusMig-Daten infrage käme.

## Projektkontext

Bereits die Unterschiede in der geographischen und chronologischen Ausrichtung der Projekte beeinflussen Wahrnehmung und Bewertung der Informationen. Während MUSICI auf ausländische Musiker in Venedig, Rom und Neapel zwischen 1650 und 1750 einging,<sup>19</sup> erforscht MusMig ein breiteres geographisches Spektrum im Zeitraum des 17. und 18. Jahrhunderts mit Schwerpunkt auf Osteuropa.<sup>20</sup> Könnte eine in MUSICI vertretene Person den Kontext einer auf Italien zielenden Migration hinüber in die MusMig-Datenbank transportieren, wäre dies ein Gewinn an Information; gleichzeitig wäre sie die damit von dem Osteuropaschwerpunkt abgrenzbar.

Die Umsetzung einer solchen „Kontext-Migration“ könnte unterschiedlichste Gestalten annehmen. Ohne weiteres vorstellbar wäre eine zusätzliche Quellenangabe oder auch ein zusätzliches Aspekt-Objekt; zusätzlich würde eine grafische Hervorhebung die Transparenz der Datenherkunft befördern.

Gleichzeitig sollte der Nutzer die Möglichkeit bekommen, die Datenkorpora getrennt voneinander zu nutzen. Dies ist im Personendaten-Repository realisierbar, da Daten nach Projekten trennbar sind, aber dennoch auf einer gemeinsamen Oberfläche erscheinen können. Die Zusammenfassung von Personen, die in mehreren Projekten auftaucht, geschieht mittels Normdateien oder projektinternen Identifikatoren.

## Erfassungssprache

Da bei MUSICI alle Daten aus italienischen Quellen stammen, wurden die Daten direkt in italienischer Sprache erfasst. Zudem war Italienisch die *lingua franca* des Projekts und eignete sich für die Kommunikation besser als Englisch. Bei MusMig hingegen ist die Quellenlage disparater, weshalb dort die Informationen in Kurzform auf Englisch erfasst werden und die Originaltexte zusätzlich als Referenz geführt werden. Es ist in diesem Falle davon auszugehen, dass eine Übersetzung der Inhalte nicht geleistet werden kann. Insofern wird der Sprachunterschied erhalten bleiben und auf mehreren Ebenen erheblichen Einfluss ausüben: Zum einen werden zentrale Funktionen der Datenbank wie zum Beispiel Volltextsuche schwerer handhabbar. Der Nutzer muss darauf vorbereitet werden, dass er mit zwei Sprachen operieren oder ansonsten auf sprachunabhängige Suchfunktionen ausweichen muss, wenn er gleichzeitig in beiden Datenmengen suchen möchte. Ein interessanter Nebeneffekt besteht darin, dass die italienische Sprache der MUSICI-Daten in der MusMig-Datenbank der Hervorhebung dessen dienlich sein wird, dass es sich um Daten aus einem anderen Kontext handelt. Als einfache technische Überbrückung empfiehlt sich eine Kollationsliste für die wichtigsten und häufigsten Termini. Zum anderen müssen die biographischen Kategorien und das Markup-Vokabular sprachlich angepasst werden. Bei den biographischen Kategorien kann dies bewerkstelligt werden, ohne die ursprünglichen Kategorien zu entfernen, da es erlaubt ist, mehrere Kategorien aus unterschiedlichen Projekten parallel zu führen. Bei Markup-Vokabular jedoch funktioniert dies nicht. Sollen die Daten gemeinsam miteinander funktionieren, muss hier Anpassungsarbeit geleistet werden.

<sup>19</sup>Eine Besprechung der MUSICI-Datenbank findet sich bei Berti/Roeder 2015.

<sup>20</sup>Die unterschiedlichen Ausrichtungen von MUSICI und MusMig wurden detaillierter bei Over/Roeder 2015 erörtert.

Informationstechnisch gesehen erscheinen mehrsprachige Inhalte suboptimal. Allerdings repräsentieren die multilingualen Daten die kulturelle Breite des Forschungsgegenstandes viel unmittelbarer als einheitssprachliche Daten. Hier kommt ein interkulturelles Potenzial zum Vorschein, welches bislang wenig untersucht wurde. Im Hinblick auf sich immer enger und globaler vernetzende Datenkorpora erscheint eine Vielfalt von Sprachen auf längere Sicht unumgänglich und bedeutet informationstechnisch eine große Herausforderung. Gleichzeitig steht sie aber in einem globalgesellschaftlich und wissenschaftlich völlig akzeptablen Licht.

### Biographische Kategorien

Die folgende Übersicht zeigt die biographischen Kategorien beider Projekte im direkten Vergleich (die Reihenfolge entspricht der aktuellen systematischen Auflistung der Projekte):

MUSICI	MusMig
01 Nome di norma	16 Normalized name
02 Nomi (altri)	17 Name variants
03 Descrizione principale	18 Biographical data
04 Dati biografici	19 Relatives
05 Provenienza	20 Affiliations
06 Permanenza	21 Profession/function
07 Viaggi	22 Training/schooling
08 Manoscritti	23 Employments
09 Attività musicali	24 Residences/sojourns/travels
10 Altre attività	25 Professional/social contacts
11 Contatti con colleghi	26 Works/products
12 Contatti con mecenati	27 Reception
13 Contatti con maestri di musica	28 Miscellaneous
14 Altri contatti	29 Commentary (public)
15 Informazioni familiari	30 Commentary (private)

In der Gegenüberstellung der Terminologien treten die konzeptionellen Unterschiede und die Interessenschwerpunkte beider Projekte deutlich hervor:

- Die MUSICI-Kategorie „Herkunft“ (05) geht bei MusMig in „Zugehörigkeiten“ (20) auf. MusMig ordnet dieser Kategorie jedoch nicht nur ethnische Herkunft, sondern zum Beispiel auch die Konfession zu.
- Die MUSICI-Kategorien „Aufenthalt“ (06) und „Reisen“ (07) finden am ehesten in der MusMig-Kategorie „Wohnsitz/Aufenthalte/Reisen“ (24) ihre Entsprechungen, wobei das italienische „permanenza“ (dauerhafter Aufenthalt) nicht dem englischen „sojourn“ (kurzer Aufenthalt) entspricht.
- MUSICI führt die Kategorie „Manuskripte“ (08), welche MusMig überhaupt nicht verwendet.

- MUSICI differenziert bei beruflichen Aktivitäten, die in „musikalische Aktivitäten“ (09) und „sonstige Aktivitäten“ (10) unterteilt sind, sowie bei Kontakten, die in vier Untergruppen „Kollegen“ (11), „Mäzene“ (12), „Lehrer“ (13) und „sonstige“ (14) differenziert werden. MusMig generalisiert in beiden Fällen zu „Dienstverhältnissen“ (23) beziehungsweise „beruflichen und sozialen Kontakten“ (25). Die „Lehrerkontakte“ (13) fielen bei MusMig allerdings unter „Ausbildung“ (22).
- MusMig führt eine Kategorie für „Werke“ (26), „Rezeption“ (27), „Verschiedenes“ (28) und zwei Kategorien für „interne“ und „öffentliche Kommentare“ (29, 30).

Offenbar sind die MUSICI-Kategorien tendenziell feiner ausdifferenziert als die MusMig-Kategorien. Dies entspricht auch der jeweiligen inhaltlichen Tendenz der Projekte: MUSICI kann sich aufgrund der Konzentration auf Rom, Venedig und Neapel mit feineren Strukturen wie zum Beispiel dem Mäzenatentum befassen, während MusMig durch den regional weit übergreifenden Kontext allgemeiner arbeiten muss.

Da das Personendaten-Repository im Falle der biographischen Kategorien erlaubt, mehrere Kategoriensysteme parallel zu führen, können die MUSICI-Kategorien beibehalten und weiter genutzt werden, unabhängig davon, welche MusMig-Kategorie ihnen zugewiesen ist. Dies ist für die Nachnutzbarkeit der Daten von sehr großem Vorteil, da auf diese Weise der ursprüngliche Erfassungskontext mittransportiert werden kann und transparent bleibt.

### Kontrollierte Vokabulare

Auch dieser Bereich fällt in die Kategorie der terminologischen Umstellungen. Anders als bei den biographischen Kategorien verhält es sich bei den kontrollierten Vokabularen, die im Datenmodell des Personendaten-Repositories auf der Markup-Ebene eingesetzt werden. Jeder markierten Textstelle kann genau ein Begriff (oder Unterbegriff) eines Vokabulars zugeordnet werden.<sup>21</sup> Damit ist es technisch nicht möglich, die Markup-Vokabulare von MUSICI und MusMig parallel zu führen.<sup>22</sup> Entsprechend muss das MUSICI-Vokabular entweder erhalten bleiben oder an das MusMig-Vokabular angepasst werden. Eine dritte Möglichkeit besteht in der Erstellung eines gemischten Vokabulars, welches auf den oberen hierarchischen Ebenen die Begriffe von MusMig adaptiert, auf den unteren, differenzierteren Ebenen hingegen die Begriffe von MUSICI beibehält.

Zwei weitere Schwachpunkte des PDR-Datenmodells erschweren die reibungslose Migration der Daten. So kann erstens nicht festgehalten werden, nach welchem Vokabular die Auszeichnung erfolgt. Man kann lediglich mit der Annahme operieren, dass sich alle Daten mit einer bestimmten Projektnummer an ein vorher vereinbartes Vokabular halten. Zweitens erfolgt bei Änderungen des Vokabulars weder eine automatische Angleichung der Daten, noch werden

<sup>21</sup>Theoretisch ermöglicht es das Datenschema, die Markup-Attribute zu tokenisieren und außerdem Namespaces zu vergeben, so dass prinzipiell auch mehrere Vokabulare parallel verwendet werden könnten (z. B. @subtype=„musmig:kingdom musici:regno“). Dies ist jedoch im gedanklichen Ansatz des Modells nie formuliert worden, wurde nie getestet und wird von den Software-Komponenten in keiner Weise unterstützt. Im Hinblick auf die praktische Nutzung ist dies also kein gangbarer Weg; für zukünftige Datenmodelle könnte dies jedoch eine Alternative darstellen.

<sup>22</sup>Vgl. *Personendaten-Repository, Schema documentation for aodl.xsd*,  
<https://pdrprod.bbaw.de/wiki/lib/exe/fetch.php?media=de:schema:aodl.pdf> (24.05.2016), S. 18.

die Daten gegen das Vokabular validiert, wodurch im Verlaufe eines Projektes markante Inkonsistenzen in den Daten entstehen können. Infolgedessen gibt es keine Garantie, dass die Daten einem bestimmten Vokabular gehorchen. Ein Test der MUSICI-Daten ergab, dass ein drittes, undokumentiertes, lediglich implizit vorhandenes Vokabular besteht, welches aus den Daten heraus rekonstruierbar ist. Vor dem Transfer nach MusMig wäre dieses mit dem offiziellen Vokabular abzugleichen.

Vom Umfang her sind die Terminologien von MUSICI und MusMig mit jeweils 300-350 Begriffen vergleichbar. Für eine detaillierte Gegenüberstellung ist in dieser Untersuchung nicht der Platz, jedoch seien exemplarisch folgende charakteristische Unterschiede erwähnt:

- MUSICI führt eine fein ausdifferenzierte Terminologie zur Beschreibung von Organisationen, die insgesamt 80 Begriffe und Eigennamen umfasst, während MusMig hier mit 20 Begriffen in einer flachen Hierarchie auskommt.
- MusMig differenziert sehr stark bei der Erfassung der Sprachvarianten von Personennamen, während bei MUSICI die Unterscheidung zwischen Geburtsname und italianisierter Form genügt.
- MusMig stellt ein deutlich komplexeres Vokabular zur Beschreibung von Personenbeziehungen bereit; hier sind zum Beispiel auch Freundschafts- oder Handelsbeziehungen erfassbar, während bei MUSICI Familien- und Förderbeziehungen im Vordergrund standen.
- MUSICI führt Vokabulare sowohl für Instrumente als auch für Instrumentalistenberufe, während MusMig lediglich die Berufe berücksichtigt.

Diese Aufzählung könnte man noch weiter fortsetzen. Unter dem Strich steht, dass MUSICI in vielen Punkten differenzierter arbeitet, während MusMig häufig eine vereinfachte Terminologie bevorzugt, die zudem weniger hierarchisch operiert.

## Referenzsysteme

Eine beschwerliche Angelegenheit ist die Übertragung des geographischen Referenzsystems, welches bei MUSICI völlig anders als bei MusMig gehandhabt wird. Während MUSICI das Markup-Vokabular für eine Normierung der Namensformen nutzt,<sup>23</sup> verwendet MusMig die Identifikatoren von Geonames<sup>24</sup> als Referenzsystem. Das Land „Portugal“ wird bei MUSICI also mit dem italienischen Namen „Portogallo“ in den Markup-Attributen identifiziert, während MusMig in einem Referenzattribut auf die GeoNames-ID 2264397<sup>25</sup> referenziert. Bei der Übertragung nach MusMig müssen die in MUSICI verwendeten Ortsnamen also manuell mit Geoidentifikatoren angereichert werden. Auch Familien- und Organisationsnamen sind in dem

<sup>23</sup>Bei MUSICI liegt eine inkonsistente Verwendung des Datenmodells vor, da innerhalb des Hierarchiebaumes von beschreibenden Begriffen zu Eigennamen gewechselt wird. Dies ist bei einer überschaubaren Menge von Namen noch handhabbar, jedoch hätte dieses Vorgehen bei MusMig den Rahmen des Vokabulars gesprengt. Aus diesem Grund entschied man sich bei MusMig für eine Verwendung eines allgemeinen Georeferenzierungssystems.

<sup>24</sup>GeoNames <http://www.geonames.org/> (24.05.2016).

<sup>25</sup>Siehe <http://www.geonames.org/2264397/portuguese-republic.html> (24.05.2016).

MUSICI-Vokabular enthalten und sollten gegebenenfalls auf eine GND-Referenzierung umgestellt werden.<sup>26</sup>

## Zusammenfassung

Durch die Untersuchung wurde deutlich, dass der Vorgang einer Datenmigration weniger auf der technischen Ebene zu betrachten ist, sondern hierbei vielmehr die inhaltlichen Konsequenzen im Mittelpunkt stehen müssen. Dabei spielen sowohl Kontexte, Sprachen und Terminologien also auch Referenzsysteme eine Rolle. Am problematischsten erscheint die Umstellung von Terminologien, da diese am stärksten die Forschungsinhalte einer Datenbank semantisch strukturieren. Das Beispiel von MUSICI und MusMig zeigt, dass hier zahlreiche Umstellungen notwendig wären, um die MUSICI-Daten innerhalb der MusMig-Datenbank nutzbar zu machen.

Die Vorstellung von einer einseitigen Datenanpassung erscheint angesichts der vielen Faktoren als irrtümlich. Stattdessen sollte die Aufgabe darin gesehen werden, die Erfassungskontexte von Daten deutlich transparenter zu gestalten und die Herkunft von Daten genauer zu dokumentieren. Es gilt, die Übertragungsprozesse sehr bewusst durchzuführen, vor allem in den Digital Humanities, aber ebenso allgemein in den Informationswissenschaften. Bei der Umsetzung erwiesen sich insbesondere parallel führende Terminologien und allgemeine Referenzsysteme als hilfreich und erscheinen auch im Hinblick auf weitere Datenmigrationen als sinnvoll.

Auch wenn die Reise eines Musikers im 17. Jahrhundert nach heutigen Maßstäben beschwerlich gewesen sein mag, so lag auch damals in der Überwindung einer räumlichen Distanz nicht die eigentliche Herausforderung der Migration. Diese ist stattdessen vorrangig in der Auseinandersetzung der hergebrachten Strukturen mit den Strukturen der neuen Umgebung zu sehen. Dabei bleibt einiges erhalten oder verändert sich nur geringfügig, während sich anderes sehr grundsätzlichen Wandlungen unterzogen wird. Dieser Prozess vollzieht sich an jeder Station von neuem; der Wandernde bleibt jedoch in seiner Integrität unverändert.

## Bibliographie

<Berti/Roeder 2015> Michela Berti; Torsten Roeder: *The »Musici« Database. An Interdisciplinary Cooperation*. In: Europäische Musiker in Venedig, Rom und Neapel. Les musiciens européens à Venise, Rome et Naples. Musicisti europei a Venezia, Roma e Napoli. Deutsch-italienische Round-Table-Gespräche (= *Analecta musicologica* 52), hrsg. von Anne-Madeleine Goulet und Gesa zur Nieden, Kassel u. a.: Bärenreiter, 2015, S. 633–645.

<Brinkmann/Wolff 1999> Reinhold Brinkmann; Christoph Wolff: *Driven into Paradise. The musical migration from Nazi Germany to the United States*, Berkeley/London: University of California Press, 1999.

<sup>26</sup>Auch die Verwendung des Getty Thesaurus of Geographic Names <http://www.getty.edu/research/tools/vocabularies/tgn/> (24.05.2016) wurde erwogen, da hier auch eine Referenzierung auf historische Gebiete möglich ist. Allerdings ist die Abdeckung für die untersuchten Regionen nicht ausreichend, weshalb die Entscheidung zugunsten von GeoNames ausfiel.



<Ehrmann-Herfort/Leopold 2010> Sabine Ehrmann-Herfort; Silke Leopold (Hrsg.): *Migration und Identität. Wanderbewegungen und Kulturkontakte in der Musikgeschichte* (= *Analecta musicologica* 49), Kassel u. a.: Bärenreiter, 2013.

<MAiA 2011/2> *Special Issue: Music and Migration* = Music and Arts in Action, Volume 3, Issue 3 (2011/2), <http://musicandartsinaction.net/index.php/maia/issue/view/Vol%203,%20No%203> (24.05.2016).

<Klettke/Thalheim 2011> Meike Klettke; Bernhard Thalheim: *Evolution and Migration of Information Systems*. In: *Handbook of Conceptual Modeling*, hrsg. von David W. Embley und Bernhard Thalheim, Heidelberg: Springer, 2011, S. 381–420.

<Levine 2009> Robert Levine: *Data migration strategies*. In: Wikibon, 2009, [http://wikibon.org/wiki/v/Data\\_migration\\_strategies](http://wikibon.org/wiki/v/Data_migration_strategies) (24.05.2016).

<Over/Roeder 2015> Berthold Over; Torsten Roeder: *MUSICI und MusMig. Kontinuitäten und Diskontinuitäten*. In: *Zeitschrift für Digital Humanities*, Sonderband 1, [http://dx.doi.org/10.17175/sb001\\_017](http://dx.doi.org/10.17175/sb001_017) (24.05.2016).

<Schöch 2013> Christoph Schöch: *Big? Smart? Clean? Messy? Data in the Humanities*. In: *Journal of Digital Humanities*, Vol. 2, No. 3 (Summer 2013), <http://journalofdigitalhumanities.org/2-3/big-smart-clean-messy-data-in-the-humanities/> (24.05.2016).

<Seacord et al. 2003> Robert C. Seacord; Daniel Plakosh; Grace A. Lewis: *Modernizing Legacy Systems: Software Technologies, Engineering. Process and Business Practices*. Boston: Addison-Wesley Longman, 2003.

<Thalheim/Wang 2011> Bernhard Thalheim; Qing Wang: *Towards a Theory of Refinement for Data Migration*. In: *Lecture Notes in Computer Science* 6998 (= *Conceptual Modeling – ER* 2011), Heidelberg: Springer, 2011, S. 318–331.

<Thalheim/Wang 2013> Bernhard Thalheim; Qing Wang: *Data migration: A theoretical perspective*. In: *Data & Knowledge Engineering* 87 (Sept. 2013), S. 260–278.

<Wagner 2014> Christian Wagner (Hrsg.): *Model-Driven Software Migration: A Methodology. Reengineering, Recovery and Modernization of Legacy Systems*, Wiesbaden: Springer, 2014.

<Walkowski 2009> Niels-Oliver Walkowski: *Das Konzept einer polysemischen Datenbank und seine Konkretisierung im Personendaten-Repository der BBAW*. In: *Jahrbuch für Computerphilologie* online, 2011 <http://computerphilologie.digital-humanities.de/jg09/walkowski.html> (24.05.2016).