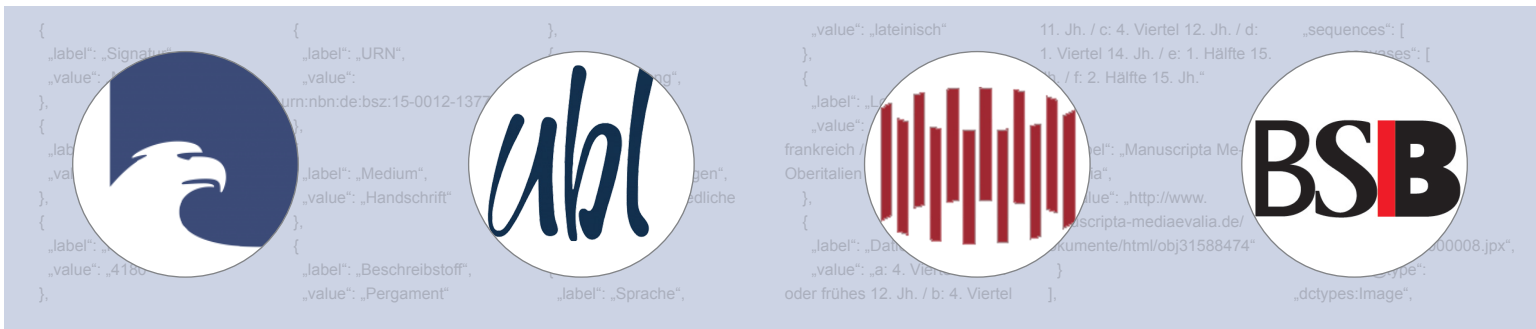


Handschriftenportal (HSP)

Antrag zur Entwicklung eines zentralen Onlineportals
für Erschließungs- und Bilddaten zu Buchhandschriften

Staatsbibliothek zu Berlin - Preußischer Kulturbesitz
Universitätsbibliothek Leipzig
Herzog August Bibliothek Wolfenbüttel
Bayerische Staatsbibliothek München



Handschriftenportal (HSP)

Antrag zur Entwicklung eines zentralen Onlineportals für
Erschließungs- und Bilddaten zu Buchhandschriften

Teil B: Beschreibung des Vorhabens

Inhalt:

1. Ausgangslage und eigene Vorarbeiten	3
1.1 Ausgangslage	3
1.2 Vorarbeiten	5
2. Ziele und Arbeitsprogramm	11
2.1 Voraussichtliche Gesamtdauer des Projekts	11
2.2 Ziele	12
2.3 Arbeitsprogramm und Umsetzung	17
2.3.1 Projektorganisation	17
AP 1: Projektmanagement	17
AP 2: Öffentlichkeitsarbeit und Dokumentation	18
2.3.2 Qualitätsoffensive und Datenmigration	18
AP 3: Erstellung der Kulturobjekt-Dokumente und Bearbeitung der Beschreibungsdokumente	18
AP 3.1: Technischer Support für die Datenanalyse und -bereinigung	20
AP 4: Technische Migration der MM-Daten	21
2.3.3 IT-Entwicklung	21
AP 5: Definition der systeminternen Schnittstellen und Review der Gesamtarchitektur	21
2.3.3.1 Nachweis- und verbundene Module	21
AP 6: Nachweismodul	21
AP 6.1: Erfassung	22
AP 6.2: Versionierung	23
AP 7: Normdaten	23
AP 8: Import	23
AP 9: Export	23
2.3.3.2 Präsentation und verbundene Module	24
AP 10: Suchmaschinensystem für Weboberfläche	24
AP 10.1: Indexschema und Hostingkonzept	24
AP 10.2: Indexierungskomponente	25
AP 11: Weboberfläche	25
AP 11.1: Web Content Management System	25
AP 11.2: Integriertes Discovery System	26
AP 11.3: Einbettung von Mirador in TYPO3	27
AP 11.4: Anzeige von Textdokumenten	27

AP 11.5: Interaktion zwischen Dokumentfenstern	28
AP 11.6: JSONBLOB Service	28
AP 11.7: Interoperabilität der Präsentationskomponenten	28
AP 12: Integration von Digitalisaten	29
2.3.3.3 Funktionsübergreifende Module	30
AP 13: Identity-Management	30
AP 14: Schema- und Skriptpflege	30
2.3.3.4 Aufbau der technischen Infrastruktur	30
AP 15: Aufbau und Betrieb der Infrastruktur und Administration	30
AP 16: IIIF-Hosting und Geschäftsmodell	31
AP 17: Entwicklungsunterstützende Systeme	32
AP 17.1: Versionskontrolle und Continuous Integration (CI) (Aufbau und Betreuung)	32
AP 17.2: Weitere Werkzeuge zur Unterstützung der Entwicklung (Aufbau und Betreuung)	32
2.3.4 Meilensteinplanung	33

1. Ausgangslage und eigene Vorarbeiten

1.1 Ausgangslage

Die Erschließung des Handschriftenerbes in Deutschland gilt seit Jahrzehnten international als vorbildlich, sowohl was die an wissenschaftlichen Bedürfnissen ausgerichteten Erschließungsstandards angeht als auch hinsichtlich der erreichten Breite des Erschließungsstandes. Ein zentraler Grund für diese internationale Vorreiterrolle ist in der gezielten Förderung der Handschriftenkatalogisierung durch die Deutsche Forschungsgemeinschaft seit den 1960er Jahren zu sehen und in der damit verknüpften Infrastrukturmaßnahme, die Erschließungsaktivitäten an Handschriftenzentren als Kompetenz- und Serviceinstitutionen zu bündeln.¹ Diese Bündelung, die im internationalen Vergleich ein Alleinstellungsmerkmal ist, führte auch dazu, dass der Wert elektronisch aufbereiteter Information für die Handschriftenforschung zu einem sehr frühen Zeitpunkt erkannt und mit Manuscripta Mediaevalia (MM; www.manuscripta-mediaevalia.de) durch die Staatsbibliotheken in Berlin (SBB-PK) und München (BSB) zusammen mit dem Bildarchiv Foto Marburg schon 1999 ein zentraler deutscher Onlinekatalog für das Handschriftenerbe geschaffen wurde. Dies war international ein Novum und stellt bis heute eines der wenigen Angebote dieser Art weltweit dar. Mit aktuell über 120.000 Datensätzen zu mittelalterlichen (und zunehmend auch neuzeitlichen) Handschriften ist MM zur Zeit das umfangreichste Informationsangebot für die internationale Handschriftenforschung.

¹ Vgl. www.handschriftenzentren.de sowie Christoph Mackert: Die Arbeitsgruppe der deutschen Handschriftenzentren – Servicezentren für Handschriftenererschließung und -digitalisierung, in: o-bib 2 (2015), Nr. 1, S. 1–14, <https://www.o-bib.de/article/view/2015H1S1-14>.

Umfang und Bedeutung der Daten in MM standen allerdings von Anfang an in deutlichem Missverhältnis zur Akzeptanz des Portals durch die Nutzer*innen. Diese kritisierten vor allem seine mangelhafte Usability und schleppende Innovation, was zur Folge hatte, dass MM von Teilen der Community ungenutzt oder gar nicht genutzt wurde, wie interne Stellungnahmen des Wissenschaftlichen Beirats der Handschriftenzentren mehrfach monierten. Systematische Schwachpunkte wurden insbesondere seit Aufkommen der Digitalisierung virulent. Insbesondere zu nennen sind: 1) die fehlende Nachhaltigkeit und nur zögerliche Weiterentwicklung aufgrund von Programmierung durch kommerzielle Dienstleister unter Verzicht auf Open Source-Softwares, 2) die technische Ansiedlung im Kontext eines kunsthistorischen Datenangebots mit entsprechenden musealen Formaten statt im bibliothekarischen Bereich, 3) die für Nutzer undurchschaubare Heterogenität des Datenangebots in Verbindung mit dem hierarchischen Datenbankmodell fein differenzierter Feldkategorien sowie 4) die Ausrichtung an den einzelnen Erschließungsdokumenten als zentraler Bezugsgröße des Systems statt an den (Handschriften-)Objekten. Gleichzeitig präsentierten Angebote wie die digitalen Sammlungen der UB Heidelberg für einen einzelnen Bestand (www.ub.uni-heidelberg.de/helios/digi/handschriften.html) oder e-codices als ein nationales Schweizer Portal (www.e-codices.unifr.ch) neue, optisch ansprechende und funktional leistungsfähige Lösungen mit engen Innovationszyklen, die die Bereitstellung wissenschaftlicher Erschließungsdaten konsequent von den digitalisierten Objekten und von den sich schnell wandelnden Usability-Aspekten her dachten. Parallel entstanden in anderen europäischen Ländern nationale oder fachlich orientierte Portale, die entweder der zentralen Präsentation von Digitalisaten und/oder von Erschließungsdaten dienten.²

In diesem veränderten Umfeld war es ein Ergebnis der „Pilotphase Handschriftendigitalisierung“ (2013-2015), dass Manuscripta Mediaevalia den Anforderungen der fortschreitenden Digitalisierung und des digital sozialisierten Nutzerverhaltens nicht mehr gewachsen war.³ Als Ergebnis ging im Juni 2015 aus dem Kreis der Handschriftenzentren und ihres Wissenschaftlichen Beirats eine Initiative hervor, deren Ziel es war, MM durch ein neues Handschriftenportal abzulösen. Dieses neue Portal soll nach aktuellen Standards sowohl die Funktion des Zentralkatalogs für das Handschriftenerbe Deutschlands erfüllen als auch dem zentralen Nachweis digitalisierter Handschriften dienen und dabei streng an den Bedürfnissen der wissenschaftlichen und bibliothekarischen Communities ausgerichtet sein. Die im März 2016 abgeschlossene DFG-Begutachtung der Pilotphase Handschriftendigitalisierung forderte folglich „ein zukunftsfähiges Konzept für ein neues Handschriftenportal“⁴ als Voraussetzung für eine DFG-Förderlinie zur Digitalisierung mittelalterlicher Handschriften.

Mit einem solchen neuen Portal können nicht nur die Ergebnisse der international vorbildlichen Handschriftenerschließung Deutschlands adäquat sichtbar und nutzbar gemacht, sondern darüber hinaus auch mit den digitalisierten Handschriftenbeständen zusammengeführt werden.

² Digital Scriptorium (www.digital-scriptorium.org), Manuscriptorium (www.manuscriptorium.com), manuscripta.at (manuscripta.at), Manus (manus.iccu.sbn.it), IRHT (www.irht.cnrs.fr/fr/ressources/les-ressources-electroniques), Gallica (gallica.bnf.fr), BVMM (bvmm.irht.cnrs.fr).

³ Zu dieser Pilotphase und ihren positiv begutachteten Ergebnissen s. <https://www.bsb-muenchen.de/kompetenzzentren-und-landesweite-dienste/kompetenzzentren/handschriftenerschliessung/services-und-projekte/> und <http://www.handschriftenzentren.de/materialien/>.

⁴ Die Gutachterstellungnahme ist auszugsweise publiziert unter <https://www.bsb-muenchen.de/kompetenzzentren-und-landesweite-dienste/kompetenzzentren/handschriftenerschliessung/services-und-projekte>.

1.2 Vorarbeiten

Anfang Juni 2016 wurde ein erster Antrag für die Entwicklung eines Handschriftenportals bei der DFG eingereicht, an dem neben der SBB-PK und der BSB auch die Universitätsbibliothek Leipzig (UBL) und die Herzog August Bibliothek Wolfenbüttel (HAB) beteiligt waren. Das Bildarchiv Foto Marburg, das an einem neuen Handschriftenportal nicht beteiligt sein wird, unterstützt das Vorhaben. Dieser Antrag wurde zwar im September 2016 aufgrund eines zu geringen Konkretisierungsgrads abgelehnt, doch betonte das Gutachtergremium zugleich, dass eine „Überarbeitung unter Berücksichtigung der Hinweise [...] sehr begrüßt“ würde.

Dieser grundlegend überarbeitete Neuantrag wird hiermit vorgelegt. Er wird für die technische Entwicklung von der SBB-PK, der UBL und der HAB sowie im Bereich der Daten-Qualitätsoffensive und der redaktionellen Aufgaben zusätzlich von der BSB als Antragstellerinnen getragen.

Planungsphase

Als Ausgangsbasis für den vorliegenden Antrag sind die intensiven Vorarbeiten für den Erstantrag nachgenutzt worden, insbesondere die detaillierte Analyse der Use Cases, welche die verschiedenen Funktionalitäten des Portals sowohl für die internen Prozesse als auch nach außen hin analysieren. In der Folge wurde ein neues Konzept entwickelt, das zwei klar voneinander geschiedene Phasen vorsieht: eine erste dreijährige Aufbauphase mit Freischaltung einer Beta-Version nach zwei Jahren und allen nutzerrelevanten Grundfunktionalitäten und eine zweite Dreijahresphase mit erweiterten Funktionalitäten und zusätzlicher Datenqualifizierung (s. u. 2.1).

Auf mehreren Arbeitstreffen der Antragspartner zwischen Herbst 2016 und Frühjahr 2017 wurden außerdem von den IT- und Fachabteilungen Fragen des logischen Datenmodells (s. u. „Datenmodell“), der Gesamtarchitektur und des Datenflusses (s. u. „Gesamtarchitektur“), der Datenqualität und -aufbereitung nach hierarchisch gestuften Datensets (s. AP 3), der Mengengerüste sowie des Projektmanagements (s. AP 1) und der Communitybeteiligung (s. AP 2) vertieft beraten und konkretisiert. Auf dieser Grundlage wurden durch die Entwicklerteams möglichst konkrete Aufwandsschätzungen erarbeitet, wobei Erfahrungen aus früheren oder laufenden Projekten herangezogen werden konnten.⁵

Kooperationen

Der Antrag wurde in engster Abstimmung mit den anderen Handschriftenzentren, dem international besetzten Wissenschaftlichen Beirat der Handschriftenzentren (www.handschriftenzentren.de/?id=13#beirat) und einem für die Portalentwicklung erweiterten Beirat ausgearbeitet, in dem vor allem Vertreter*innen von Infrastruktureinrichtungen Mitglieder sind (ausländische Handschriftenportale,

⁵ Für die einzelnen Antragsteller sind insbesondere die Erfahrungswerte aus folgenden Projekten zu nennen: 1) SBB-PK: Aufbau einer Plattform-Service- und Microservice-Infrastruktur für [KALLIOPE](#), den [GW](#) und [Cross Asia](#); 2) UBL: a) Suchmaschinenprojekte: [finc](#), [LerXe](#), b) Präsentationsprojekte zu Handschriften: [PapyrusEbers](#), [Codex Sinaiticus](#), c) Linked Data-Projekt: [AMSL](#), d) Handschriftenportalprojekte: [Orientalia](#), [Papyrusportal](#), e) Webdesign: Website der [Hs.zentren](#), der [UBL](#), von [finc.info](#); 3) HAB für Schema- und Skriptpflege: [Europeana regia](#). Zahlreiche Editionsprojekte, in denen die HAB den technischen Support und die Skriptpflege übernommen hat, u.a. [Tagebücher Fürst Christian II. von Anhalt-Bernburg](#), [Schriften und Briefe Andreas Bodensteins von Karlstadt](#).

Handschriftencensus, WZIS [www.wasserzeichen-online.de], Altbestandsbibliotheken, Archive, Mittelalterblog [mittelalter.hypotheses.org]).⁶ Über den Wissenschaftlichen Beirat und über verschiedene Kooperationen der Handschriftenzentren bestehen intensive Vernetzungen zu wichtigen handschriftenbezogenen Informationsanbietern in Frankreich und Italien (IRHT [www.irht.cnrs.fr], SISMEL [www.sismelfirenze.it], ICCU als Betreiber von Manus online [www.iccu.sbn.it, <http://manus.iccu.sbn.it>]), zum Mediävistenverband e.V. (www.mediaevistenverband.de) sowie zu prominenten Vertretern der Handschriftenforschung, -erschließung und -digitalisierung im englischsprachigen Raum, in dem bis heute keine zentralen Angebote entwickelt wurden, sondern nur Einzelinitiativen existieren. Der intensive Austausch mit der Autographendatenbank Kalliope (<http://kalliope.staatsbibliothek-berlin.de>) und der Einbanddatenbank (<http://www.hist-einband.de>) ist durch die gemeinsame institutionelle Verankerung an der SBB-PK sichergestellt. Bei Treffen mit Vertretern der Deutschen Digitalen Bibliothek (DDB - <https://www.deutsche-digitale-bibliothek.de>) und des Archivportals-D (<https://www.archivportal-d.de>) wurde ein grundsätzlicher Konsens über das künftige funktionale Zusammenwirken der Portale erzielt. Ebenso besteht die Bereitschaft des neu bewilligten Akademievorhabens Handschriftencensus⁷, des Schweizer Handschriftenportals e-codices sowie des internationalen Fragmentarium-Projekts (fragmentarium.ms), Daten auszutauschen und hierfür Schnittstellen und Workflows zu erarbeiten. Geplant ist außerdem die gegenseitige Nachnutzung bzw. gemeinsame Entwicklung von Normvokabular im Bereich von Schriftarten/Paläographie, Buchschmuck, Musiknotation und differenzierter Schreibsprachenbestimmung.

Für die im System befindlichen Daten der Österreichischen Nationalbibliothek, der Österreichischen Akademie der Wissenschaften (als Betreiber des Handschriftenportals manuscripta.at) und der UB Basel (für den HAN-Verbund) wurden Vorabsprachen getroffen, wonach die veralteten Daten, welche derzeit bereits in MM vorgehalten werden, grundsätzlich über regelmäßige automatisierte Einspielungsprozesse durch die jeweils aktuellen Daten aus den einzelnen Nachweissystemen ersetzt werden (s. AP 3). Die Projektpläne wurden außerdem von Vertretern der Antragsteller auf verschiedenen Tagungen zur Diskussion gestellt.⁸

Um die sehr gut genutzten Images der digitalisierten Handschriftenkataloge auch im neuen Handschriftenportal zeigen und sie darüber hinaus für die Volltexterstellung und -nutzung verwenden zu können, werden die hierfür notwendigen Rechte mit den Verlagen geklärt. Das Einverständnis des Verlags Harrassowitz, aus dem die überwiegende Mehrzahl der betreffenden Kataloge stammt, liegt den Antragstellern bereits vor.⁹

⁶ Der Erweiterte Beirat umfasst Vertreter*innen von ausländischen Handschriftenportalen (e-codices, Fragmentarium [fragmentarium.ms], manuscripta.at), vom Handschriftencensus (s. Fußnote 7), vom Archivbereich, vom Mittelalterblog sowie von Altbestandsbibliotheken.

⁷ Das Vorhaben wird den bisherigen Handschriftencensus (www.handschriftencensus.de) weiterentwickeln und ersetzen, vgl. http://www.adwmainz.de/fileadmin/adwmainz/Pressestimmen/2016_11_02PMNeues_Akademieprojekt2017.pdf.

⁸ Tagung "Illuminierte Handschriften und Inkunabeln im Zeitalter Gutenbergs" (Wien Januar 2016), Maschinen und Manuskripte III (Darmstadt Februar 2016), Forschungsworkshop der SBB-PK (Berlin Juni 2016), Sitzung der AG Regionalbibliotheken (Jena Oktober 2016), IIF/Mirador Research Workspace Meeting (Amsterdam Oktober 2016), Symposium "Handschriften und Alte Drucke" (Blaubeuren November 2016), Fragmentarium Case Study Workshop (Wolfenbüttel Mai/Juni 2017), Arbeitstreffen zu einem Portal oriental. Handschriften (Berlin Juni 2017).

⁹ Mit den Verlagen Reichert, Klostermann, Hauswedell, Akademie-Verlag und weiteren ist eine möglichst zügige Einigung bis spätestens Projektbeginn projektiert.

Datenmodell

Als wichtige Vorarbeit und konzeptionelle Weiterentwicklung gegenüber dem ersten Antrag wurde ein neues logisches Datenmodell mit klar definierten Entitäten erstellt:

Während in MM bislang einzelne Erschließungsdokumente nebeneinander stehen, soll im Handschriftenportal das **“Kulturobjekt”** die zentrale Entität im System sein, mit der alle weiteren Informationen wie identifizierende Kerndaten, Beschreibungsdokumente, Annotationen, Digitalisate und Normdaten verknüpft sind. Der Begriff **“Kulturobjekt”** statt **“Handschrift”** berücksichtigt, dass neben Handschriften (inkl. Fragmenten, Einzelblättern und Rollen) auch Überlieferungssymbiosen mit Drucken und buchbezogene Artefakte wie etwa separierte Einbände im System zu verwalten sind. Jedes real existierende sowie jedes virtuell (re-)konstruierte Kulturobjekt wird durch genau ein **Kulturobjektdokument (KOD)** repräsentiert.

Das KOD umfasst eine definierte Menge an Informationen zum Kulturobjekt, die aus Beschreibungsdokumenten und Annotationen gewonnen werden, wobei im System beliebig viele Beschreibungsdokumente und Annotationen zu jedem KOD existieren können. Während die Beschreibungsdokumente und Annotationen mit einem bestimmten Urheber verbunden sind und letztlich immer eine historische Station im Prozess der Wissensgenerierung repräsentieren, dient das KOD als gemeinsamer Anker für alle Daten zu einem bestimmten Kulturobjekt und bildet mit den in ihm selbst hinterlegten Informationen den jeweils möglichst aktuellen Kenntnisstand über die zentralen Eigenschaften eines Kulturobjekts ab. Im Gegensatz zu Beschreibungen, die mit dem Ende eines Katalogisierungs- oder Forschungsvorhabens als abgeschlossen gelten müssen, bleibt das KOD offen und dynamisch. Ergänzungen oder Änderungen des Kenntnisstandes können auf der Basis neuer Beschreibungsdokumente oder nutzerinduzierter Annotationen in die KODs eingebracht werden. Die fachliche Herkunft der neuen Informationen wird durch Referenzen auf diese Quellen feldbezogen nachvollziehbar gemacht.

Der Attributumfang¹⁰ des KODs ist zweistufig definiert. Für jede Handschrift ist ein **Minimaldatenset** verpflichtend, mit dessen Hilfe das Kulturobjekt überhaupt erst individualisierbar wird. Es umfasst Ort, Namen und GND-ID der besitzenden Entität (Institution oder Person), eine Identifikation des Kulturobjekts (in der Regel die Signatur) sowie eine im Portal zu vergebende interne KOD-ID und schließlich die ID des korrespondierenden Normdatensatzes für **“Schriftdenkmäler”** der Gemeinsamen Normdatei (GND) der Deutschen Nationalbibliothek (DNB), dessen genaue Ausgestaltung derzeit in Arbeit ist. Vertreter aller Antragsteller sind Mitglieder in der neu gegründeten AG **“Handschriften”** der DNB zum internationalen Regelwerk **Resource Description and Access (RDA)** (<https://wiki.dnb.de/display/RDAINFO/RDA+und+Sondermaterialien>). Da die Definition der Schriftdenkmäler auch in diesem Zusammenhang von zentraler Bedeutung ist, sind Austausch mit der GND und Mitsprache der Partner bei der einschlägigen Entwicklung gewährleistet. Die KODs des Handschriftenportals sollen nach dem derzeitigen Diskussionsstand in der GND in einem festzulegenden Format als Schriftdenkmal abgelegt und mit einer ein-eindeutigen Kennung versehen werden. Idealerweise werden die KODs des Handschriftenportals also die Quelle für die GND-Normdokumente bilden.

Durch die geplante flächendeckende Generierung von GND-Normdatensätzen der Kategorie **“Schriftdenkmäler”** aus den KODs wird darüber hinaus die Verknüpfung der Portaldaten mit der GND und

¹⁰ Ein Attribut in dem Datenmodell entspricht einem Feld bzw. einer Informationskategorie in der Beschreibungspraxis. Beispiele für Attribute sind etwa die Entstehungszeit, die Schreibsprache oder das Format des Kulturobjekts. Die jeweiligen Inhalte eines Attributs werden im Folgenden als **“Werte”** bezeichnet.

damit in das semantische Netz realisiert, wobei die derzeit angestoßene Entwicklung eines internationalen Handschriften-Identifiers einbezogen wird.¹¹ Durch die Synchronisation von GND-Schriftdenkmälern und KODs des Portals erhalten letztere gleichzeitig Normdatenqualität. Für die Norm-Entitäten tritt das Portal der GND gegenüber grundsätzlich als Konsument auf. Hinsichtlich der in Rücksprache mit den besitzenden Einrichtungen von ihm erstellten und verwalteten KODs hat es aber die Rolle des primären nationalen Lieferanten. Sie werden der GND initial zur Verfügung gestellt und im Dauerbetrieb mit dieser synchronisiert. Die Schreibrechte in der GND zur Veränderung des korrespondierenden Normsatzes zum Schriftdenkmal hinsichtlich der Angaben des Minimaldatensets des Kulturobjekts müssen daher klar definiert sein.

Die Angaben des Minimaldatensets werden durch diejenigen des **Kerndatensets** ergänzt, dessen Attribute die zentralen Merkmale des Objekts bezeichnen und dessen Belegung deshalb zwar nicht verpflichtend, aber für alle Handschriften anzustreben ist. Zu diesem Kerndatenset, das enge Übereinstimmungen mit dem Datenset des DFG-Viewers hat und perspektivisch auch im IIF-Kontext Verwendung finden soll, gehören:

- der (fingierte) Titel des Kulturobjekts
- alternative Signaturschreibweisen der besitzenden Institution
- Entstehungsort/e und -datum/en
- Beschreibstoff/e
- Format/Abmessungen
- Blattzahl/Umfang
- Schreibsprache/n (nur Grundkategorien: dt., lat., griech., etc.)
- Buchschmuck (ja/nein, ggfs. Grundkategorien)
- Musiknotation (ja/nein, ggfs. Grundkategorien)
- Formtyp (Codex, Rotulus, Fragment etc.)
- Status des Objekts (vorhanden, zerstört, verschollen etc.)
- bisherige Signaturen bei Besitzwechsel

Nutzer-Recherchen zu weiteren Inhalten wie Autoren, Werken, Initien u. ä. werden in der ersten Phase über die Volltexte und Registerangaben der Beschreibungen möglich sein (s. auch AP 3). Eine weitergehende Normierung dieser Angaben für eine verbesserte Suche ist für die zweite Phase geplant.

Die Attribute des Kerndatensets werden im Gegensatz zu denen des Minimaldatensets aus den - potentiell unterschiedlichen - Beschreibungsdokumenten bezogen. Jedes dieser Attribute des KODs wird daher von einem Link auf das informationsgebende Beschreibungsdokument im Sinn einer Quellenangabe begleitet. Digitalisate sowie Beschreibungen werden im Modell als eigene Entität gestaltet, die über die interne KOD-ID mit dem Kulturobjekt verknüpft sind. Damit stellt das KOD den Zusammenhang aller für eine Handschrift im Portal vorhandenen Datenangebote sicher.

Ort der Aufnahme von Erschließungsdaten - von basalen Kurzaufnahmen bis hin zur wissenschaftlichen Tiefenerschließung - ist die Entität der **Beschreibung**. Diese Entität ist so modelliert, dass die durch die Richtlinien Handschriftenkatalogisierung der DFG (www.manuscripta-mediaevalia.de/hs/kataloge/HSKRICH.htm) sachlich definierten Beschreibungsabschnitte in jeweils eigenen Modulen abgelegt

¹¹ Hier wird an die Ergebnisse des Pariser Workshops vom 27.04.2017 zu einer International Standard Manuscript Shelfmark Number angeknüpft werden können, an dem Vertreter*innen der Antragsteller teilgenommen haben: www.irht.cnrs.fr/fr/agenda/manuscript-ids-identifiants-des-manuscrits.

werden können (z. B. Äußeres, Einband, Inhalt). Den Ausgangspunkt bildet die auf das Kulturobjekt als Ganzes bezogene Beschreibung, welche die Informationen aus Schlagzeile, Äußerer Beschreibung und Geschichte der Handschrift aufnimmt, auf welche die Referenzen des Kerndatensets der KOD verweisen. Über diese Kerndaten hinaus werden im Beschreibungsdokument Urheber und Datum der Beschreibung selbst festgehalten. Die Entität Beschreibung ist prinzipiell seriell rekursiv verwendbar, um den spezifischen Anforderungen an die Erschließung von Einbänden, Texten, Fragmenten in situ, Faszikeln oder Buchschmuckelementen zu genügen. Für diese wird je ein eigener Typ der Entität Beschreibung modelliert, der jeweils spezifische Attribute für die Erschließung aufweisen kann. Dabei werden alle Einzeltypen der Beschreibung des Kulturobjekts als Ganzes zugeordnet. Allen Beschreibungstypen gemeinsam ist, dass sie neben der Möglichkeit der kontrollierten Erfassung dieser spezifischen Attribute auch die Formulierung von Beschreibungsfließtext erlauben. Damit haben der Beschreibungstext und sein Index ihren logischen Ort in einer gemeinsamen Entität.

Im Datenmodell wird die Entität **Annotation** im engeren Sinne als Repräsentation nutzergenerierter Daten verstanden und ist funktional eine Teil- oder Kurzbeschreibung. Annotationen können von den am System registrierten Nutzern in der Web-Präsentation erstellt werden und unterschiedliche, vom Nutzer zu definierende Öffentlichkeitszustände (privat, gruppenintern, veröffentlicht) haben. Wird eine Annotation veröffentlicht, wird sie zugleich als Mikropublikation mit Urheberschaft belegt und als Entität persistiert. In ihrer persistierten Form werden diese Annotationen neben den Beschreibungsdokumenten Teil des Nachweismoduls und damit eine Quelle für die Aktualisierung der Datensets des KOD.

Die Entitäten **Personen**, **Körperschaften**, **Literarische Werke** und **Orte** werden - soweit dort vorhanden - alle mit einer GND-Referenz versehen und über das Normdatenmodul (AP 3) mit den entsprechenden Normdaten der GND synchronisiert. Noch nicht in der GND vertretene Angaben werden über das Normdatenmodul einem Redaktionsgeschäftsgang zugeführt. In den Beschreibungsdokumenten werden Referenzen auf diese Entitäten für die Erschließung genutzt. Über das Kerndatenset der betreffenden Beschreibungsdokumente werden die Daten zu Orten und Körperschaften mit ihren GND-Referenzen auch in das KOD vermittelt.

Gesamtarchitektur

Als weitere zentrale konzeptionelle Vorarbeit konnten sich die Partner bei der technischen Entwicklung auf die Architektur des geplanten Gesamtsystems verständigen. Diese basiert auf dem in den beteiligten Einrichtungen bereits angewandten Muster der Microservices. Dieser moderne Software-Enterprise-Entwicklungsansatz hat einen modularen Aufbau des Gesamtsystems zum Ziel und bietet im Unterschied zu einem monolithischen System folgende Vorteile:

- einfache Erweiterbarkeit durch flexible Anbindung weiterer Komponenten
- Skalierbarkeit und einfacher Austausch einzelner Komponenten
- effektive Nachnutzung einzelner Module und damit die Möglichkeit schneller Entwicklungszyklen
- Robustheit gegen Softwarefehler und Ausfälle

Der Gesamtsystemverbund des künftigen Handschriftenportals unterteilt sich in folgende Bereiche von IT-Komponenten:

- Clientsysteme
- Plattform-Services
- Backend-Services
- Externe Dienste
- Message Queue

Der Bereich der **Clientsysteme** umfasst alle Komponenten, die über ein graphisches Benutzerinterface (GUI) verfügen und im Internet erreichbar sind, also sowohl die offen zugängliche Weboberfläche wie auch die zugangsgeregelten Bereiche der Erfassung und Workspaces. Für das Handschriftenportal wird ein Content-Management-System (CMS) mit entsprechenden Erweiterungen aufgebaut. Als Produkt kommt das Open-Source-CMS TYPO3 zum Einsatz, geplant sind Erweiterungen für die Integration des CMS in die Microservice-Architektur sowie für die Integration des Mirador-IIIF-Viewers und der Suche.

Die **Plattform-Services** gewährleisten die funktionale Einheit des modularen Gesamtsystems. Nur mit ihrer Hilfe kann der konsistente Betrieb zahlreicher miteinander kommunizierender Microservices gewährleistet werden. In diesem Bereich stellen alle IT-Systeme zentrale Dienste für die Backend-Services zur Verfügung. Im Einzelnen werden folgende Systeme benötigt:

- Zentrales Logging
- Zentrales Monitoring
- Service-Discovery und Routing
- Zentrales Konfigurationsmanagement

Diese Services wurden teilweise bereits innerhalb der SBB-PK aufgebaut (s. AP 15).

Durch die **Backend-Services** werden die Daten des Systems verwaltet und Prozesse gesteuert. Folgende Systeme werden als Microservices umgesetzt: Nachweismodul mit Erfassung, Versionierung, Im- und Export, Normdaten- und Identity-Management, CMS mit Anzeige und Suche, in der zweiten Phase auch das Katalogmodul sowie die Workspaces mit der Möglichkeit, Annotationen anzulegen und zu veröffentlichen.

Der Teil der **externen Dienste** besteht aus Systemen, die durch das Handschriftenportal genutzt werden (GND, DFG-Viewer, lokale Bildangebote). Die logische Integration dieser Systeme in die Gesamtdarstellung dient dazu, die Abhängigkeiten einzelner Architekturkomponenten zu erkennen und zu berücksichtigen.

Die **Message-Queue** steuert den Datenaustausch einiger wichtiger Komponenten und ermöglicht so eine permanent aktuelle Darstellung aller Daten in den Clientsystemen des Handschriftenportals. Eine Messaging-Lösung bietet den Vorteil asynchroner Kommunikation, welche die Unabhängigkeit der einzelnen Systeme erhöht. Darüber hinaus toleriert die Messaging-Lösung Ausfälle einzelner Systeme ohne Datenverluste, was die Widerstandsfähigkeit des Gesamtsystems erhöht. Die Message-Queue in Verbindung mit dem SOLR-Cloud-Modul stellt damit die Schnittstelle zwischen den Backend-Systemen und der CMS-Weboberfläche dar.

Prüfung bestehender Anwendungen

In der Phase der Antragsvorbereitung wurden alternative Infrastrukturen, insbesondere D-Work der UB Heidelberg und e-codices, für eine mögliche Nachnutzung in Betracht gezogen. Es stellte sich heraus, dass die jeweils spezifisch implementierten Infrastrukturen nicht einfach kopiert werden können und dass die entwickelten Komponenten auf ihre jeweiligen Umgebungen individuell ausgerichtet sind, weshalb eine unmittelbare Nachnutzung nicht möglich ist. Bei e-codices, dessen Programmierung durch einen privaten Dienstleister stattfindet, sind Backend und Präsentation der Daten so eng miteinander verwoben, dass einzelne Elemente nicht herausgelöst und nachgenutzt werden können. Dies betrifft insbesondere die Einbindung von Digitalisaten aus lokalen Serversystemen, die Vorverarbeitung der TEI-XML-Beschreibungen für die Suche, eine MySQL-Datenbank für die gesamte Workflow-Organisation und zur Erfassung von Normdaten, sowie die Einbindung spezieller Daten pro Handschrift (die Übersichtsseite) und die Mehrsprachigkeit der Oberfläche. Gleiches gilt für D-Work: D-Work wurde als Tool zur Organisation des Digitalisierungsworkflows konzipiert und initial implementiert und erst später sukzessive um Bestandteile zur Präsentation von Daten und andere Funktionalitäten erweitert. Auch hier sind zahlreiche lokale Besonderheiten in der Programmierung berücksichtigt, die nicht ohne Weiteres verallgemeinert und für ein zentrales Handschriftenportal nachgenutzt werden können. Gleichwohl ist im Rahmen der Projektrealisierung der fachliche Austausch mit den jeweiligen Betreibern gesichert (s. o. "Kooperationen").

Übergang von MM zum neuen Handschriftenportal

Die Web-Präsentation von MM wird bis zur Freischaltung des neuen Handschriftenportals und der Migration aller Daten weiter betrieben und wie bisher bei der Firma Stegmann gehostet. Für diesen Übergangsbetrieb wird ab Herbst 2017 die verbesserte Version MM3 bereitgestellt. Die SBB-PK hat unter finanzieller Beteiligung der BSB einen entsprechenden Wartungsvertrag mit Stegmann abgeschlossen.

Die Neueingabe von Erschließungsdaten/Beschreibungen und der Nachweis neuer Digitalisate können während der Entwicklung des neuen Portals weiterhin über die Erfassungssoftware MXML erfolgen. Das Hosting der Erfassungsdatenbank wird mit Projektbeginn in der SBB-PK stattfinden; die Wartung des Systems durch den Dienstleister wird während des Projekts wie bislang durch einen mit SBB-PK und BSB geschlossenen Vertrag garantiert. Hierdurch werden auch die redaktionelle Nutzung und Bearbeitung der Daten durch die Partner im Projekt und im späteren Betrieb sichergestellt. Alle in dieser Zeit generierten Daten werden zusammen mit dem bestehenden Datenpool mit der Produktivschaltung im neuen Portal verfügbar gemacht.

2. Ziele und Arbeitsprogramm

2.1 Voraussichtliche Gesamtdauer des Projekts

Das Gesamtprojekt ist auf insgesamt sechs Jahre angelegt und gliedert sich in **zwei klar voneinander abgegrenzte Phasen von je drei Jahren**.

In der **ersten Projektphase** wird die Basis der Infrastruktur inklusive des Nachweises aller Daten in der zentralen Datenbank, der Erfassungskomponente sowie der Präsentation geschaffen und nach zwei Jahren produktiv freigeschaltet. Zu diesem Zeitpunkt werden alle Daten aus MM in das neue System überführt und auf die neue Einheit des Kulturobjekts bezogen sein. Dafür werden in der ersten Phase

für alle Kulturobjekte KODs mit möglichst vollständigen und normierten Kerndatensets generiert. Außerdem werden alle bislang fehlenden Volltexte der jetzt schon in MM vorhandenen Katalogdaten per OCR erzeugt und in das System integriert. Auf diesem Weg werden die Recherchemöglichkeiten speziell zu den bislang elektronisch nur durch Indexdaten erschlossenen deutschen Handschriftenkatalogen wesentlich verbessert. Parallel wird ein Webportal entwickelt, welches nutzerseitig die Suche in den Metadaten mit ihren normierten Werten sowie in den Beschreibungsvolltexten ermöglicht. Bild- und Textdaten können in einem einheitlichen Präsentationsmodus auf der Grundlage des International Image Interoperability Framework (IIIF - iiif.io) angezeigt werden, der es auch erlaubt, mehrere Dokumente - auch aus verschiedenen Institutionen bzw. Repositorien - in unterschiedlichen Anordnungen parallel zu benutzen. Zur Recherche im gesamten Datenbestand wird aus dem Nachweismodul ein Suchmaschinenindex befüllt und mit der Weboberfläche verbunden. Für Einrichtungen, die selbst keine Infrastruktur für die Bereitstellung von Digitalisaten nach dem IIIF-Standard betreiben können oder wollen, wird die UBL im Projektverlauf ein Dienstleistungsangebot und Geschäftsmodell zum Hosting von IIIF-Images und -Metadaten erarbeiten.

In der **zweiten Projektphase** soll dieses Basisangebot durch folgende Maßnahmen ausgebaut und zusätzlich qualifiziert werden: 1. Workspaces, 2. vertiefte Datenbereinigung und -normierung, 3. Datenaustauschroutinen, 4. Katalogmodul zur Generierung von Publikationen, 5. Aktualisierung der DFG-Richtlinien (im Einzelnen s. u. 2.2).

2.2 Ziele

Das Handschriftenportal (HSP) soll als das zentrale deutsche Repository für Buchhandschriften des Mittelalters und der Neuzeit fungieren, unter Einbeziehung verwandter Objekte wie Fragmente, Rollen, Wachstafeln, Einzelblattüberlieferungen, Handschrift-Druck-Symbiosen oder separierte Einbände.¹² In seiner neuen Qualität eignet sich das HSP auch für die Erschließung und Präsentation neuzeitlicher Buchhandschriften und überwindet so eine bibliothekarisch wie wissenschaftsdisziplinär künstliche Grenzziehung. Infrastrukturell wird damit zugleich eine Nachweislücke geschlossen, indem das Handschriftenportal komplementär neben Kalliope tritt, das in seiner Materialsicht und Systemarchitektur auf Autographe und Nachlässe ausgerichtet ist. Der Fokus des Handschriftenportals liegt zunächst primär auf europäischen Handschriften auf Latein, Griechisch und in den Volkssprachen, es kann aber von Beginn an auch als Basisnachweis für außereuropäische Handschriften genutzt werden.¹³

Das Portal wird standardisierte Erschließungsdaten unterschiedlicher Tiefe gebündelt anbieten und mit den Digitalisaten zusammenführen. Neue Erschließungsdaten werden über ein Erfassungstool und über Schnittstellen in das System gelangen. Der Nachweis der Digitalisate wird zumindest über Links in die lokalen Präsentationen erfolgen, angestrebt wird aber eine möglichst weitgehende Präsentation

¹² Standards für die Beschreibung von Einbänden als eigenständige Kunstwerke werden derzeit an der BSB im Rahmen eines DFG-Projekts entwickelt, vgl. <https://www.bsb-muenchen.de/ueber-uns/projekte/erschliessung-und-digitalisierung-von-einbaenden-als-eigenstaendige-kunstobjekte/>.

¹³ Für deren vertiefte Erschließung ist auf Spezialdatenbanken zu verweisen. Zu nennen sind hier etwa die bereits existierenden Datenbanken zu orientalischen Handschriften, wie sie an der UBL und an der SBB-PK betreut werden: www.islamic-manuscripts.net, www.refaiya.uni-leipzig.de und orient-digital.staatsbibliothek-berlin.de. Mit dem geplanten Zentralportal orientalischer Handschriften ist eine konzeptionelle Abstimmung weiterer Planungen ins Auge gefasst.

innerhalb des Systems auf der Grundlage des IIF-Standards, wofür eigene Services angeboten werden. Innerhalb des Systems vorgehaltene Images werden auch über den DFG-Viewer zur Verfügung gestellt.

Bestimmend für die Ziele des neuen Handschriftenportals sind die Bedürfnisse seiner potentiellen Nutzer*innen. Dabei stehen nach Analyse der Antragsteller **vier Akteurgruppen bzw. Anforderungsszenarien** im Mittelpunkt:

- Wissenschaftler*innen möchten für ihre handschriftenbezogenen Forschungen im Portal Informationen zu einzelnen Handschriften/Kulturobjekten und deren Digitalisaten erhalten und diese integriert nutzen, sie möchten thematische Recherchen unterschiedlicher Komplexität durchführen und benötigen interaktive Features zur privaten und öffentlichen Annotation sowie zur Kollaboration.
- Bibliotheken und weitere handschriftenbesitzende Einrichtungen möchten die vorhandenen Daten über ihre eigenen Bestände für lokale Anwendungen (z. B. Digitalisierungsworkflows) nachnutzen, pflegen und ergänzen sowie neue Daten über Schnittstellen automatisiert einspielen.
- Wissenschaftliche Bearbeiter*innen an Handschriftenzentren sowie an weiteren Bibliotheken und Forschungseinrichtungen generieren neue Daten mit einer Differenzierung bis hin zur Tiefenerschließung nach den Richtlinien der DFG: Diese Daten sollen entweder mittels eines geeigneten Erfassungstools in den Datenbestand des Portals eingebracht und mit Normdaten verknüpft oder über schnittstellenbasierte Austauschprozesse für den zentralen Nachweis bereitgestellt werden. Die in einem Projektkontext entstandenen Daten sollen außerdem im Zusammenhang des Portals gebündelt als Publikation (Katalog) dokumentiert werden können.
- Betreiber verwandter fachspezifischer und regionaler Nachweissysteme wünschen eine möglichst große Sichtbarkeit bzw. Referenzierung ihrer Daten auf der Grundlage von automatisierten Austauschroutinen. Nationale und internationale Portale (DDB [www.deutsche-digitale-bibliothek.de], europeana [www.europeana.eu]) sind an einem Nachweis von Informationen aus dem Portal interessiert. Dabei soll das Portal aus ihrer Sicht perspektivisch eine material-spezifische Aggregatorenfunktion erfüllen.

Um diesen verschiedenen Nutzungsszenarien gerecht zu werden, wird das Portal in der **ersten Projektphase** als eine Vielzahl miteinander verbundener funktionaler Einheiten entwickelt werden, die jeweils eines oder mehrere dieser Szenarien bedienen:

- **Nutzeranbindung:** Die möglichst enge An- und Einbindung der wissenschaftlichen Communities und ein Hochmaß an Transparenz nach außen ist Voraussetzung für den Projekterfolg. Die hierfür notwendige Kommunikation ist im Projekt als eigene Arbeitsaufgabe definiert (AP 2). Zur Information über Projektetappen dienen ein redaktionell betreutes Blog mit Social Media-Kanälen und RSS-Feed¹⁴ sowie entsprechende Meldungen über die Website der Handschriftenzentren und die einschlägigen Mailinglisten. Nach 18 bzw. 30 Monaten werden Workshops zur Evaluierung durch Nutzergruppen durchgeführt; hinzu kommen zwei Best-Practice-Workshops für die erforderliche Weiterentwicklung von IIF (s. AP 11.7). Über diese Kanäle hinaus ermöglicht die Anwendung des Paradigmas der Agilen Softwareentwicklung bereits ab einem frühen Zeitpunkt die kontinuierliche Einbeziehung der Communities in die Entwicklung des

¹⁴ Zumindest für die Frühphase des Projekts ist hier eine Zusammenarbeit mit dem von der Community gut rezipierten Mittelalterblog vereinbart.

Portals (s. AP 1). Nach den regelmäßigen Programmiersprints werden deren Ergebnisse in dedizierten Teststellungen mit Feedbackfunktionen allen Interessierten zugänglich gemacht. Forscher*innen und bestandshaltende Einrichtungen können das Projekt so permanent eng begleiten. Als institutionalisiertes Kooperations- und Beratungsgremium stehen die anderen Handschriftenzentren und der Wissenschaftliche Beirat sowie der für die Projektzeit erweiterte Beirat zur Verfügung. Auch innerhalb der Projektgruppe werden durch das agile Vorgehen kurze Kommunikationswege und -zyklen sichergestellt. Auf Leitungsebene garantiert die Anlage des Projektmanagements mit seiner Hierarchie von (Teil-) Projektleitung(en) und Steuerungsgruppe klare Entscheidungskompetenzen und Eskalationswege (AP 1).

- **Datengrundlage:** Alle bisher in MM verfügbaren Daten werden auch im neuen Portal verfügbar gemacht. Auf der Grundlage des neuen fachlichen Datenmodells werden darüber hinaus Datenergänzungs- und -qualifizierungsmaßnahmen durchgeführt (AP 3-4). Durch die Generierung von KODs mit einem Set möglichst normierter Kerndaten wird nicht nur der gezielte Zugriff auf alle Informationen zu einer bestimmten Handschrift garantiert, sondern werden auch facetthaltige Suchen möglich, die auf diesen qualifizierten Kerndaten aufsetzen. Weiterhin werden über OCR-Verfahren systematisch Beschreibungsvolltexte derjenigen Handschriftenkataloge erzeugt und ins System eingebracht, zu denen MM bereits Indexdaten aus den Katalogregistern vorhält. Dies ermöglicht einerseits die möglichst automatisierte Datenextraktion zur Befüllung wichtiger Kerndatenfelder im KOD, vor allem aber verbessern die nachgeführten Volltexte und ein speziell auf diese differenzierte Datenlage ausgerichtetes Discoverysystem deutlich die Recherchequalität (APs 10-11). Über ein Importmodul (AP 8) können Metadaten zu Handschriften aus Bibliotheksverbänden in den Portalbestand integriert und damit z. B. Ergebnisse verteilter Digitalisierungsvorhaben zentral fachspezifisch verfügbar gemacht werden.
- **Nachweissystem:** Alle im Portal vorgehaltenen Informationen werden im Nachweismodul, dem Kernstück der Systemarchitektur, dauerhaft gespeichert und verwaltet (AP 6). Über ein Erfassungstool (AP 6.1), standardisierte Import- sowie Normdatenschnittstellen (APs 7-8) gelangen neue Informationen manuell und maschinell in dieses Nachweismodul. Über eine standardisierte Exportschnittstelle (AP 9) werden Daten nach einem differenzierten Creative Commons-Lizenzierungsmodell (de.creativecommons.org) zur Nachnutzung in lokalen Angeboten und anderen Online-Portalen bereitgestellt. Angestrebt ist grundsätzlich eine möglichst liberale Lizenzregelung. Eine versionierte Datenhaltung sorgt für die persistente Zitierbarkeit aller Dokumente (AP 6.2). Um die Rechte am Datenhandling im Gesamtsystem nach Nutzergruppen steuern zu können, werden Zugangsberechtigungen (Rechte und Rollen) durch ein Identity-Management geregelt (AP 13).
- **Suchmaschine:** Um eine komfortable, integrierte Weboberfläche bereitstellen zu können, ist es erforderlich, alle relevanten Daten in einer integrierten Suchmaschine für den Echtzeitbetrieb aufzubereiten. Hierzu ist es erforderlich, die Daten unter verschiedenen Gesichtspunkten vollautomatisch und ereignisgesteuert umzuformen und in die Suchmaschine zu indexieren. Dabei sind die Abhängigkeiten zwischen Datensätzen zu berücksichtigen sowie Anreicherungsprozesse zu implementieren (AP 10).
- **Web-Präsentation:** Für die Präsentation der Daten einschließlich der an Usability-Kriterien ausgerichteten Recherche und Gestaltung des Webauftritts werden ein CMS sowie eine geeignete Suchmaschine eingesetzt (APs 11.1 und 11.2). Zur Anzeige von Digitalisaten (Handschriften- und Katalogimages) werden der DFG-Viewer sowie Mirador für IIIF-fähige Daten eingebunden (AP 11.3). Der Aufbau eines IIIF-Servers ermöglicht die Integration IIIF-fähiger Bildangebote auch aus verteilten Umgebungen. Dieser Server steht auch für Institutionen ohne IIIF-

fähige Präsentation oder ganz ohne eigenes Präsentationssystem zur Bildanzeige innerhalb des Portals zur Verfügung (AP 16). Im Projekt soll IIF um die Anzeige von Textdaten erweitert werden. Alle Entwicklungen folgen dem Paradigma des Responsive Design. Durch den konsequenten Einsatz von Identifikatoren und Normdaten werden zentrale Informationseinheiten im Portal als Microdata eingebettet und sind so im Sinne von Linked Open Data maschinell nachnutzbar. Die Internationalität der Communities bedingt die Mehrsprachigkeit des Angebots, sobald das Portal produktiv arbeitet. Die für die Nutzerführung wesentlichen Bereiche der Web-Präsentation werden daher auch auf Englisch angeboten werden.

- **Interoperabilität:** Für die formale und inhaltliche Kontrolle der Datenstruktur der KODs und Beschreibungsdokumente sind Schemata notwendig. Für den Im- und Export von Beschreibungsdokumenten in das bzw. aus dem Nachweismodul sowie für die Generierung von Katalogpublikationen müssen Skripte erstellt und gepflegt werden (AP 14).

Auf diese Weise können bereits innerhalb der ersten Projektphase entscheidende Verbesserungen gegenüber dem bisherigen Angebot von MM erreicht werden. Das HSP bietet 1) eine zukunftsfähige Weboberfläche, die 2) einem intuitiven Rechercheverhalten entspricht und zugleich 3) alle Informationen zu einer Handschrift jeweils unter einem Anker bündelt. Es verfügt 4) über eine deutlich verbesserte Datengrundlage aufgrund der Integration aller Volltexte der Handschriftenbeschreibungen. Dank seiner Interoperabilität ist es 5) eng mit den bibliothekarischen und wissenschaftlichen Communities sowie relevanten Datenaggregatoren vernetzt.

Zusätzliche Verbesserungen für die Nutzer*innen sind für die **zweite Projektphase** geplant:

- Entsprechend dem Bedürfnis der Forschung nach interaktiven Funktionalitäten und elektronischen Arbeitsumgebungen soll das Portal für angemeldete Nutzer*innen **Workspaces** anbieten, in denen mit einem differenzierten Rechtemanagement Text- und Bilddaten erhoben und gesammelt, mit definierten Anderen geteilt oder im System unter einem definierten Lizenzmodell als Annotation/Mikropublikation freigeschaltet werden und damit in das zentrale Nachweismodul übergehen. Auf diese Art wird ein virtueller Handschriftenlesesaal mit den kommunikativen Aspekten des Forschungsaustauschs, ergänzt um die neuen Möglichkeiten des Digital Age, entstehen. Die Entwicklung dieser Workspaces erst in der zweiten Projektphase hat den Vorteil, dass die gegenwärtig sehr dynamische internationale Entwicklung auf diesem Gebiet - vornehmlich in der IIF-Community - mitgestaltet und so Synergien genutzt werden können.
- **Datenbearbeitung und -ergänzung:** Aufgrund der Historizität handschriftenkundlichen Wissens und des ebenfalls historisch bedingten unterschiedlichen Grads an Normdatenanreicherung wird das neue Handschriftenportal jenseits der KODs auf der Ebene der einzelnen Beschreibungsdokumente Daten unterschiedlicher Tiefe, Aktualität und Normierung anbieten. Eine weitere Normierung von Daten zu Personen, Institutionen und Werken in diesen Dokumenten mit geringerem Normierungsgrad ist für die zweite Projektphase vorgesehen. Das gleiche gilt für die Integration von fachlichen Thesauri in die Datenqualifizierung und Neu-erfassung, bei der bereits vorhandene Lösungen der internationalen Community nachgenutzt werden.
- Ebenfalls für die zweite Phase ist die Entwicklung von **Datenaustauschroutinen** mit weiteren Anbietern (e-codices, HAN, manuscripta.at, ÖNB, fragmentarium) vorgesehen. Im Dialog mit dem HSC ist darüber hinaus ein Datenabgleich vorgesehen, wobei in erster Linie die Kerndaten

zur Schreibsprache verbessert werden sollen. Hierfür ist auch die Erweiterung der Austauschformate auf diese externen Quellen bzw. Forschungsdaten (z. B. Geschichtsquellen des Mittelalters [www.geschichtsquellen.de/index.html], KdiH [www.dlma.badw.de/eine-seite/kdih]) notwendig.

- Die Entwicklung einer **Katalogfunktion** soll es ermöglichen, aus dem Nachweismodul Gruppen von Erschließungsdokumenten auszugeben und mit ebenfalls im System gespeicherten Paratexten (Einleitungen, Register etc.) zu versehen. Auf diesem Weg können digitale und analoge Publikationen aus dem System heraus erzeugt werden.

Im Anschluss an die von der Community begleitete Entwicklung des Erfassungstools und an die Verfeinerung des Datenmodells in der ersten Projektphase sollen die **DFG-Richtlinien Handschriftenkatalogisierung** in der zweiten Projektphase an den aktuellen Stand der Praxis angepasst werden. Die Notwendigkeit einer solchen Aktualisierung und Überarbeitung unter Berücksichtigung der veränderten Bedingungen des Digital Age wurde in den letzten Jahren sowohl vom zuständigen DFG-Gutachtergremium als auch von den Handschriftenzentren und ihrem Wiss. Beirat mehrfach betont. Mit den neuen Richtlinien wird eine Kodifizierung und Vereinheitlichung des Vorgehens erreicht, wobei auch Erfahrungen aus den ersten Handschriftendigitalisierungsprojekten nach Aufhebung des DFG-Moratoriums einbezogen werden können. Gegebenenfalls wird das Erfassungstool entsprechend aktualisiert.

Schließlich werden alle Entwicklungen der zweiten Projektphase in die bis dahin aufgebaute Systemarchitektur integriert und die von den Erweiterungen betroffenen bereits bestehenden Module angepasst.

Die **Nachhaltigkeit** des geplanten Angebots ist über die geplante Gesamtprojektlaufzeit hinaus strukturell und institutionell sichergestellt:

Erstens wird durch die konsequente Modularisierung aller Softwarebestandteile deren flexible Nutzbarkeit erreicht. Die Verwendung von Open Source-Software einerseits und die Publikation der eigenen Codes als Open Source andererseits garantieren dem System maximale technische Transparenz.

Zweitens gewährleistet die SBB-PK als technischer Betreiber des Gesamtsystems den dauerhaften Betrieb des Handschriftenportals und die zentrale Datenspeicherung im Rahmen ihrer eigenen umfassenden Infrastruktur. Für die technische Administration, die kontinuierliche Betreuung und Entwicklungsunterstützung sagt die SBB-PK eine ausreichende personelle Unterstützung zu.

Im Bereich des Datenmanagements stellen SBB-PK und BSB drittens die redaktionelle Pflege des Handschriftenportals auch nach Projektende nachhaltig sicher. Diese umfasst die Betreuung und Optimierung der Prozesse in den Bereichen der Datenerfassung und des maschinellen Daten-Ingests sowie die inhaltliche Koordinierung des Datenaustauschs mit weiteren Anbietern, außerdem die Normdatenredaktion und die Aktualisierung der eingesetzten Thesauri im Dialog mit der Wissenschaft. Im konzeptionellen Bereich werden Datenmodell und Erfassungsformate in Abstimmung mit Normdatenanbietern sowie nationalen und internationalen Gremien und Standards kontinuierlich weiterentwickelt. Im Austausch mit den Communities wird entstehender Verbesserungsbedarf an den Funktionalitäten des Systems evaluiert, daraus Entwicklungsempfehlungen abgeleitet und diese an die SBB als technischen Betreiber kommuniziert. Kontinuierliche Öffentlichkeitsarbeit über die jeweils einschlägigen Kanäle (Präsentationen, Newsletter, Blog und Social Media), die Dokumentation aller Prozesse sowie die Anwenderbetreuung in Form von Schulungen runden das künftige Aufgabenspektrum ab.

Hinsichtlich der IIIF-Services verpflichtet sich zum vierten die UBL, dauerhaft einen entsprechenden Server zu betreiben und hierzu gehörige Dienstleistungen für Institutionen ohne IIIF-fähige Präsentation und insbesondere auch für Kleinsammlungen ohne eigene Bildpräsentation anzubieten.

Die Pflege von Schemata und Skripten wird fünftens auch über den Projektzeitraum hinaus durch die HAB sichergestellt.

2.3 Arbeitsprogramm und Umsetzung

Die Arbeiten im Projekt verteilen sich auf folgende 17 Arbeitspakete (AP). Dabei ordnen sich die einzelnen APs den drei Themenbereichen Projektorganisation (2.3.1), Qualitätsoffensive und Datenmigration (2.3.2) sowie IT-Entwicklung (2.3.3) zu.

Die Umsetzung der technischen Entwicklungsaufgaben geschieht durch die drei Projektpartner SBB-PK, UBL und HAB mit folgenden verteilten Rollen: Die SBB-PK verantwortet schwerpunktmäßig die Umsetzung des Nachweismoduls und seiner Teile, während die UBL in erster Linie die Präsentation realisiert und die HAB vorrangig die fachlichen Schnittstellen ausarbeitet. Alle vier Antragsteller-Institutionen (SBB-PK, UBL, BSB, HAB) sind gemeinsam an der Daten-Qualitätsoffensive beteiligt.

2.3.1 Projektorganisation

Die Organisation des Projekts basiert nach innen auf Festlegungen zum Projektmanagement und nach außen auf der systematischen Einbeziehung der Communities.

AP 1: Projektmanagement

Das Projekt Handschriftenportal wird in einem klassischen Projektmanagement-Rahmen durchgeführt. Dabei werden die Rollen des Projektleiters, des Lenkungsausschusses und der Teilprojektleiter besetzt. Die Realisierungsphase, also der gesamte eigentliche Entwicklungsprozess, wird agil gestaltet. Kennzeichnend hierfür sind die Arbeit im Team, die enge Anbindung an die fachliche Seite – vertreten durch einen Product Owner im Team selbst – sowie das iterative Vorgehen im Scrum-Prozess.

Die klassische Meilensteinsetzung in der Planungsphase nach 6-Monats-Zyklen (s. u. 2.3.4) wird in der Realisierungsphase durch die iterative Planung von Sprints (4-Wochen-Zyklen) ergänzt. Die Sprintziele werden auf das Erreichen der Meilensteine ausgerichtet und teamübergreifend kontinuierlich abgestimmt. Um eine effektive Zusammenarbeit der Teams zu gewährleisten, wird es neben Telefon- und Videokonferenzen regelmäßige Projekttreffen geben (in der Regel acht pro Jahr).

Die Rolle des Auftraggebers im Projekt hat der Lenkungsausschuss. Er setzt sich aus Vertretern der Hausleitungen der Partner zusammen und übernimmt das Controlling der Phasen sowie die Abnahme der Meilensteine. Außerdem steht er dem Projektleiter als übergeordnetes Entscheidungs- und Eskalationsgremium zur Verfügung. Die operative Leitung des Gesamtprojekts hat der Projektleiter zusammen mit seinem Vertreter.¹⁵ Sie koordinieren teamübergreifend die Arbeiten, die für das Erreichen der Meilensteine notwendig sind. Die Teilprojektleiter planen und überwachen die Realisierung der Arbeitspakete. Sie sind für das Aufgaben- und Ressourcenmanagement in Bezug auf das Erreichen der Meilensteine verantwortlich und geben regelmäßige Statusmeldungen an den Projektleiter. Im Projekt

¹⁵ Die verschiedenen Rollen können - wo sinnvoll - durchaus in einer Person miteinander kombiniert werden. So kann etwa der Projektleiter gleichzeitig Product Owner eines Teams sein.

wird jedes Team einen Teilprojektleiter haben, außerdem wird es einen Teilprojektleiter für das häuserübergreifende Arbeitspaket zur Datenveredelung geben. Innerhalb der Teams ist der Product Owner der fachliche Ansprechpartner für die Entwicklung. Er vertritt die fachliche Auftraggeberseite und bestimmt und priorisiert die fachlichen Anforderungen (Product Backlog). Gemeinsam mit dem Scrum-Master und dem Scrum-Team legt er die Ziele und Prioritäten für die einzelnen Sprints fest, nimmt im Sprint-Reviewmeeting die Teilergebnisse des Scrum-Teams ab und entscheidet über den Einsatz und/oder die Veröffentlichung des (Teil-)Produkts. Der Scrum-Master organisiert und optimiert die Projektarbeit im Team. Er aktualisiert die Sprintplanung im Sprint-Backlog und visualisiert den aktuellen Status in Bezug auf das Erreichen der Sprintziele im Sprint Burndown Chart. Das Entwickler- bzw. Scrum-Team setzt die Arbeitspakete bzw. die Anforderungen in Produktfunktionalitäten um.

AP 2: Öffentlichkeitsarbeit und Dokumentation

Um das Projekt von Anfang an gegenüber den verschiedenen Communities transparent zu gestalten, eine gute Außenkommunikation zu gewährleisten und die verschiedenen Aktivitäten zur Nutzer-Einbeziehung zu organisieren, ist eine Stelle für Öffentlichkeitsarbeit vorgesehen, bei der die Projektkommunikation zentral zusammenläuft. Sie betreut das Projekt-Blog und sichert die regelmäßige Informationstätigkeit über Blogbeiträge, Newsletter und Meldungen auf der Website der Handschriftenzentren sowie über die einschlägigen Mailinglisten ab. Über das Blog informiert sie die Community über aktualisierte Teststellungen (s. AP 1). Sie ist gleichzeitig dafür verantwortlich, Texte und Materialien zu erarbeiten und zu verbreiten, die über das Projekt informieren. Außerdem übernimmt sie die Vorbereitung und Organisation der beiden Nutzer-Workshops.

Mit dieser Stelle soll ebenfalls die Dokumentation der Projektergebnisse verbunden werden, insbesondere die Erstellung des Handbuchs und die Zusammenstellung von Best-Practice-Informationen.

2.3.2 Qualitätsoffensive und Datenmigration

AP 3: Erstellung der Kulturobjekt-Dokumente und Bearbeitung der Beschreibungsdokumente

Die in Katalogisierungs- und Konversionsprojekten über Jahrzehnte erstellten Beschreibungsdaten zu aktuell über 120.000 Buchhandschriften in MM bilden auch künftig die Basis des Informationsangebots für die Wissenschaft. Dies hat schon die Prüfgruppe zur Pilotphase Handschriftendigitalisierung betont und die Übernahme aller Daten aus der aktuellen Umgebung in das neue Handschriftenportal gefordert. Außerdem würde eine parallele oder verteilte Datenhaltung zwischen MM und HSP Recherchen erschweren und wäre Nutzern nicht zuzumuten.

Das HSP wird daher bereits bei seiner Produktivschaltung den gesamten zu diesem Zeitpunkt in MM vorhandenen Datenbestand anbieten. Dieser umfasst qualitativ hochwertige Volltextbeschreibungen mit normierten Indexdaten aus der MXML-Erfassung, ältere in HiDA3 primär erfasste Beschreibungsdaten und retrokonvertierte Indexdaten gedruckter Handschriftenkataloge sowie die Volltext- und Indexdaten aus dem 2004-2009 durchgeführten Retrokonversionsprojekt der BSB München zu Altkatalogen aus der Zeit vor der DFG-Förderung von Handschriftenerschließungsprojekten.

Für das Gesamtziel einer verbesserten Datengrundlage und Recherchequalität sind vorrangig zwei Maßnahmen erforderlich: die umfassende Bereitstellung möglichst aller Beschreibungsvolltexte sowie die flächendeckende Generierung und Befüllung von KODs als zentralen Informationsankern.

Volltextgenerierung

Ein zentraler Mangel des derzeitigen MM ist, dass gerade die inhaltlich hochwertige DFG-geförderte Katalogisierung aus dem Zeitraum von etwa 1970 bis 2010 im Datenbestand größtenteils nur durch die retrokonvertierten Indexdaten (sog. Registerdokumente) vertreten ist, die eigentlichen Beschreibungstexte aber bestenfalls als Images, nicht jedoch als digitale Volltexte vorliegen.

Um die Recherchequalität bei diesen Handschriftenbeschreibungen deutlich zu verbessern und um eine möglichst einheitliche Nutzererfahrung zu ermöglichen, werden in der ersten Projektphase die Volltexte der Beschreibungen in den entsprechenden Dokumenten ergänzt und diese für Recherche und Dokumentanzeige verfügbar gemacht. Da die Qualität der bereits vorhandenen älteren Images für die Generierung der Volltexte nicht ausreicht, sind zuerst ca. 66.000 hochauflösende Images aus etwa 220 Katalogen neu zu erstellen und katalogweise einer OCR-Bearbeitung zu unterziehen. In den so entstandenen Volltexten der einzelnen Handschriftenkataloge werden anschließend IT-unterstützt die jeweils enthaltenen Beschreibungen ausgezeichnet, über die eindeutige Kombination mit Katalog und Signatur identifiziert und damit im Volltext recherchierbar gemacht.

Darüber hinaus werden die in Überschrift und Schlagzeile des erzeugten Beschreibungsvolltextes enthaltenen Attributwerte ausgezeichnet und - ebenfalls über die Kombination mit Katalog und Signatur - softwaregestützt in die vorhandenen Beschreibungsdokumente übernommen. Diese verbesserten Dokumente werden in einem nächsten Schritt ebenfalls für die Befüllung bzw. Ergänzung des Kerndatensets der KODs genutzt.

Erzeugung und Befüllung der KODs

Um im Sinne einer verbesserten Recherchequalität Kerndaten auf einem homogenen Level für alle Handschriften im System zur Verfügung zu stellen und gleichzeitig für alle Informationen zu einer Handschrift eine gemeinsame Referenz zu schaffen, sind aus den vorhandenen Beschreibungsdokumenten automatisiert und iterativ KODs zu generieren. Die Neuerstellung oder Aktualisierung von KODs im späteren Regelbetrieb erfolgt entweder aus neu hinzukommenden Beschreibungsdokumenten oder über redaktionell qualifizierte Annotationen. Für die Beschreibungsdokumente wird dieser Prozess im Regelfall als eine durch rollenbezogene Rechte gesteuerte Routine ablaufen; im Einzelfall ist auch hier eine manuelle redaktionelle Qualifizierung vorgesehen.

Für die initiale Erzeugung der KODs und die Befüllung der Attribute bieten die vorhandenen Beschreibungsdokumente Daten in unterschiedlicher Dichte und Qualität. Um das Nutzungserlebnis und den Nutzwert gegenüber MM deutlich zu verbessern, werden diese Beschreibungsdaten in der ersten Projektphase einem komplexen Veredelungsprozess unterzogen, der die beiden Aspekte der Ergänzung sowie der Normalisierung bzw. Normierung von Attributwerten umfasst.

Zu Beginn werden für alle in MM nachgewiesenen Handschriften KODs im Umfang des verpflichtenden **Minimaldatensets** (s. o. 1.2 "Datenmodell") erstellt. Benötigt werden also Angaben zu Ort, Namen und GND-ID der besitzenden Entität, zur Signatur sowie zur ID des Schriftdenkmals der GND, die aus den vorhandenen Beschreibungsdokumenten, aus einem Datenabgleich mit der GND sowie aus einer ersten Umfrage bei den besitzenden Einrichtungen gewonnen werden. Bei letzteren liegt grundsätzlich das Recht auf die verbindliche Ansetzung der Signaturen in den KODs.

Anschließend werden die Attribute des **Kerndatensets** des KODs befüllt (s. o. 1.2). Die hierfür notwendigen Informationen sind in den meisten gedruckten Beschreibungen zwar vorhanden, jedoch elektronisch z. T. nicht erfasst bzw. strukturiert worden. Je nach Ausgangskatalog der einzelnen Beschreibung und zu befüllendem Attribut sollen die fehlenden Werte nun mittels unterschiedlicher, aneinander anknüpfender Methoden gewonnen werden: 1) durch die Übernahme der bereits bei der Volltextgenerierung gewonnenen Attributwerte aus den vorhandenen Beschreibungsdokumenten (s. o. "Volltextgenerierung"), 2) über eine pauschale Attributbefüllung anhand katalogspezifischer Charakteristika (z. B. Katalog der deutschsprachigen Handschriften / der lateinischen Quarthandschriften / der Pergamenthandschriften), 3) durch die Auswertung impliziter Attributwerte in den Volltexten.

Schließlich werden die besitzenden Einrichtungen in einer zweiten Umfrage um kontrollierende bzw. ergänzende Informationen zu allen bzw. bestimmten Handschriften aus ihrem Bestand gebeten. Um die auf diesen Wegen gewonnenen Attributwerte für eine verbesserte Recherche in den KODs und ihren Beschreibungsdokumenten und für eine sinnvolle Facettierung von Suchergebnissen anbieten zu können, müssen diese in einem kontinuierlichen Bearbeitungsprozess sowohl in ihrer Schreibung korrigiert als auch normalisiert bzw. normiert werden.

Eine nicht unwesentliche Zahl von Beschreibungsdokumenten im Datenbestand hat ihren primären Nachweisort in anderen Systemen und wird auch künftig im Portal nur sekundär zur Verfügung gestellt. Hierbei handelt es sich um Daten zu Handschriften der ÖNB Wien (in QuickSearch der ÖNB), der UB Basel und der weiteren Partner im HAN-Verbund sowie um die Katalogisate aus den Erschließungsvorhaben der ÖAW (manuscripta.at). Für die dort erschlossenen Handschriften werden nach Absprache mit den betreffenden Einrichtungen KODs im Minimaldatenset angelegt, für das Kerndatenset aber werden mit den entsprechenden Einrichtungen in der zweiten Projektphase Datenaustauschroutinen erarbeitet, die eine regelmäßige Übernahme und Verwendung möglichst aktueller (Kern-)Informationen über das Importmodul des Handschriftenportals erlauben.

Trotz der angestrebten Homogenisierung der Datenqualität werden sich die Beschreibungsdokumente naturgemäß auch künftig in Alter und Beschreibungstiefe unterscheiden. Diese Unterschiede werden dem Nutzer in der Recherche durch ein entsprechendes Ranking mit erklärenden Informationen vermittelt.

Um die Aufgaben dieses APs im vorgegebenen Zeitraum bewältigen zu können, werden die Arbeiten im Bereich der Volltextauszeichnung, Attributzuweisung und -veredlung parallel an allen vier Antragsteller-Institutionen durchgeführt.

AP 3.1: Technischer Support für die Datenanalyse und -bereinigung

Um die fachlichen Aufgaben der Datenanalyse und -bereinigung angemessen umsetzen zu können, bedarf es eines individuell auf diese Aufgaben zugeschnittenen und flexibel anpassbaren technischen Supports. Mit Hilfe von automatisierten Verfahren werden die Analyse und Bereinigung der Altdaten unterstützt. In der ersten Projektphase liegt der Schwerpunkt dabei auf der Unterstützung der Auswertung der neuen wie auch der bereits vorhandenen Volltexte. Hierfür werden einfach bedienbare Tools benötigt, die eine klassifizierende Auszeichnung der betreffenden Strings für die einzelnen Attribute und deren Integration in die dazugehörigen Beschreibungsdokumente ermöglichen (s. AP 3.1). Außerdem müssen für die Normierung bzw. Normalisierung der auf diesem Weg gewonnenen sowie der bereits vorliegenden Daten Werkzeuge erstellt werden, mit deren Hilfe die redaktionellen Kräfte einen qualifizierten Datenabgleich zur GND und zu weiteren Indexlisten vornehmen können (s. o. AP

3). Für diesen Zweck wird voraussichtlich das sogenannte APPER-Tool nachgenutzt und in seinen Funktionalitäten erweitert. Diese Werkzeuge müssen über die gesamte Bearbeitungsdauer hinweg immer wieder kontinuierlich angepasst werden.

AP 4: Technische Migration der MM-Daten

Die Migration der Altdaten wird aus technischer Sicht als schrittweises Vorgehen konzipiert und durchgeführt. Basierend auf dem neuen Datenmodell und einer Poolbildung verschiedener Dokument-Kategorien werden verschiedene Migrationspfade erstellt und implementiert. Die Migration der Altdaten wird also in einem iterativen Prozess erfolgen, in dem die Migrationsstrategie und das neue Datenmodell kontinuierlich angepasst und erweitert werden.

In der ersten Migrationsphase von zwei Jahren werden alle Daten in das Zielsystem überführt. Dort werden die KODs mindestens im Umfang des Minimaldatensets zur Verfügung stehen, das Kerndatenset der Beschreibungsdokumente wird so weit wie möglich befüllt werden. Bis zum Ende der ersten Projektphase werden die Beschreibungsdokumente und mit ihnen die KODs als Ergebnis der iterativen Bearbeitung des Kerndatensets weiter vervollständigt.

2.3.3 IT-Entwicklung

AP 5: Definition der systeminternen Schnittstellen und Review der Gesamtarchitektur

Um die Qualität der zu entwickelnden Software zu gewährleisten, muss auf der Basis des Architekturmodells im Projekt ein technisches Detailkonzept des Gesamtsystems erstellt werden. Die Schnittstellen und Kommunikationsbeziehungen zwischen den einzelnen Komponenten müssen im Einzelnen beschrieben und zwischen den Entwicklerteams vereinbart werden. Ein Systemarchitekt führt ein regelmäßiges Review der Entwicklungsergebnisse in Bezug auf die Gesamtarchitektur durch und kommuniziert dieses mit den Entwicklerteams.

2.3.3.1 Nachweis- und verbundene Module

AP 6: Nachweismodul

Als fachliche Grundlage des Nachweismoduls dient das neue logische Datenmodell (s. o. 1.2). Dieses Datenmodell wird im Projektverlauf kontinuierlich verfeinert und vervollständigt.

Im Nachweismodul werden auch diejenigen im Portal verwendeten potentiellen Normdaten abgelegt, die noch nicht in die GND aufgenommen werden können (z. B. noch nicht individualisierte Personennamen) oder die einen gegenüber der GND abweichenden Attributumfang aufweisen (Kerndatenset des KOD).

Konzipiert wird das Nachweismodul als eine zentrale Serveranwendung, die aus mehreren Microservices besteht und von allen anderen Komponenten des Systems zur Ablage der deskriptiven Daten und zur Datenmanipulation verwendet werden kann. Die Integration des Erfassungssystems wird als Client über eine HTTP-REST-Schnittstelle realisiert. Auf der Grundlage des Datenmodells werden ein neues Datenschema entwickelt und die Integritätsbedingungen für die Daten definiert, deren Einhaltung bei Änderungen durch das Nachweismodul geprüft wird. Da das Nachweismodul als Grundlage für die verteilte Erfassung der Daten in vielen parallelen Verarbeitungsschritten dienen soll, ist als Datenablage

ein relationales Datenbankmanagementsystem (RDBMS) vorgesehen (Gewährleistung der ACID-Eigenschaften). Die im Nachweismodul entstandenen Daten werden in der Message-Queue zur Darstellungssynchronisation zur Verfügung gestellt. Durch die Trennung von Geschäftslogik und Persistenzschicht wird die konkrete physikalische Ablage der Daten flexibel steuerbar und kann in Zukunft ohne große Migrationsaufwände verändert werden.

AP 6.1: Erfassung

Für die Erstellung neuer Beschreibungen unterschiedlicher Tiefe im System ist ein Erfassungsmodul erforderlich, das bis zum Ende der ersten Projektphase die MM-Erfassungssoftware MXML abgelöst haben wird. Grundlage der Erfassung ist dabei das unter 1.2 beschriebene Datenmodell. Die inhaltliche Qualität künftig neu zu erfassender Daten wird durch klare Katalogisierungsvorgaben, definierte Redaktionsrichtlinien sowie die Einbindung von Normdaten wirksam sichergestellt. Die Entwicklung des Erfassungsmoduls richtet sich an den DFG-Richtlinien Handschriftenkatalogisierung aus, berücksichtigt die Erfahrungen mit MXML und geschieht in engster Abstimmung mit dem aktuell entstehenden Regelwerk für eine RDA-konforme Katalogisierung von Handschriften in Bibliotheken (s. o. 1.2).

Das User-Interface des Erfassungsmoduls wird den unterschiedlichen Notwendigkeiten der Erfassung komplexer Sachzusammenhänge gerecht werden. Einerseits wird es je nach zu beschreibendem Gegenstand einen individuellen flexiblen Dokumentaufbau und damit eine strukturell angemessene Repräsentation jedes Kulturobjekts ermöglichen. Andererseits wird ein pro Beschreibungstyp bzw. -abschnitt definierter Attributumfang die Nutzer bei ihrer Erfassungstätigkeit unterstützen. Im Unterschied zu der reduzierten Zahl und den begrenzten Wertelisten der Attribute für die retrospektive Datenbearbeitung wird die künftige Katalogisierung in den unterschiedlichen Beschreibungstypen ein erweitertes Datenset nutzen können, das in den ersten beiden Projektjahren definiert und grundsätzlich den Vorgaben der DFG-Richtlinien folgen wird. Dabei werden die Erfahrungen mit der bisherigen Erfassungssoftware MXML sowohl hinsichtlich der Struktureinheiten der Beschreibungen als auch der Kategorien für normierte Recherchedaten (z. B. Initien, Autoren, Werktitel, Provenienzangaben) ausgewertet. Bei zu normalisierenden oder normierenden Attributen findet eine technische Validierung statt, die das Risiko inhomogener Daten begrenzt. Der Nutzer wird außerdem auf der Grundlage einer Vorschau festlegen können, welche Inhalte aus der Erfassung in der Weboberfläche dargestellt werden sollen. Außerdem behält er über individualisierte Dashboards jederzeit den Überblick über seine Arbeitsumgebung. Hier werden ihm auf Wunsch seine letzten Aktionen, anstehende Aufgaben sowie Benachrichtigungen eingeblendet.

Die Erfassungskomponente wird als Web-Anwendung (JAVA-Enterprise-Projekt) umgesetzt, die den Nutzern in Abhängigkeit von ihren Rollen und den damit verbundenen Rechten unterschiedliche Eingabemasken zum Erfassen der deskriptiven Daten zur Verfügung stellt sowie Normdaten in diese integriert. Es werden Validierungsregeln und Eingabehilfen für die Erfassung definiert und implementiert, die bei der Eingabe der Daten aktiviert werden können. Diese Validierung führt über eine Prüfung des Minimaldatensets auch den Abgleich neuer Beschreibungsdokumente mit bereits vorhandenen KODs durch. Sollte keine Identität zu ermitteln sein, so wird dieses Beschreibungsdokument über die Importschnittstelle in das Nachweismodul übergeben und auf diesem Weg einer redaktionellen Prüfung unterzogen, ob ein neues KOD aus dem Beschreibungsdokument zu generieren ist oder dem Erfasser eine Fehlermeldung zurückgeliefert wird. Die Erfassung wird formbasiert als Multi-User-Lösung im Browser zur Verfügung gestellt.

AP 6.2: Versionierung

Im Nachweismodul werden die Autor*innen von Beschreibungen oder die bestandshaltenden Institutionen Änderungen an einem Beschreibungsdokument nachvollziehen und bei Bedarf die Dokumentversion eines früheren Zeitpunktes wiederherstellen können. Um ältere und möglicherweise an anderer Stelle bereits referenzierte Beschreibungszustände verifizieren zu können, wird diese Versionierung nicht nur im Nachweismodul, sondern auch dem Endnutzer in der Weboberfläche zur Verfügung gestellt. Die Versionierung aller deskriptiven Daten muss bei der Datenablage im Nachweismodul realisiert und für die Workflows der Datenerfassung, -aktualisierung und -anzeige berücksichtigt werden. Technisch wird diese Versionierung voraussichtlich auf Basis des weit verbreiteten Versionierungssystems Git umgesetzt, da auf diese Weise der Entwicklungsaufwand reduziert werden kann.

AP 7: Normdaten

Die Verknüpfung der Handschriftenerschließungsdaten des Nachweismoduls mit akzeptierten Normdatenbeständen ist die Voraussetzung für die Einbindung des Portals in Linked-Open-Data-Zusammenhänge sowie für die Datenkonsistenz im Portal selbst. Daher ist hinsichtlich der Normdaten zu Personen, Körperschaften, Orten, literarischen Werken und Schriftdenkmälern eine konsequente Anbindung an die GND vorgesehen. Zu diesem Zweck wird das Handschriftenportal ein als Microservice konzipiertes Normdatenmodul erhalten. Dieses Modul realisiert die Integration der Normdaten auf Basis der OAI-PMH-Schnittstelle des Portals. Die Daten werden permanent mit den Daten der GND synchronisiert, bei Bedarf für die Nutzung im Portal angereichert und lokal abgelegt. Für die Speicherung der Normdaten ist eine NoSQL-Dokumentendatenbank vorgesehen. Werden Normdaten für die Erfassung benötigt, so können diese in gewünschter Form auf Basis einer HTTP-REST-Schnittstelle abgefragt werden. Für die Aktualisierung des entsprechenden SOLR-Indexes werden die Normdaten in der Message-Queue zur Verfügung gestellt.

AP 8: Import

In der ersten Projektphase wird ein Modul zum maschinellen (Massen-)Import von einschlägigen Bibliotheksdaten aus Verbundsystemen sowie definierten externen Datenbeständen besitzender Einrichtungen in das Nachweismodul des Handschriftenportals realisiert.

Dieser Microservice stellt die Schnittstelle für den Import in unterschiedlichen Formaten zur Verfügung (insbesondere TEI, MXML und MARC21, eventuell künftig BibFRAME). Technisch wird die Komponente je nach Datenmenge als HTTP-Schnittstelle oder via Netzwerkfreigabe (z. B. ftp/WebDAV) realisiert werden. Die Dokumentation der Schnittstelle wird veröffentlicht. Beim Import in das Nachweismodul werden die Daten gegen ein definiertes Schema validiert, der Nutzer erhält vom System eine Rückmeldung. Die automatisierte Prüfung importierter Beschreibungen auf das Vorhandensein korrespondierender KODs hin führt entweder zu einer automatischen Verknüpfung oder zu einer redaktionell gesteuerten Anlegen neuer KODs. Im Fehlerfall (KOD vorhanden, aber kein Matching, da abweichende Syntax im importierten Dokument) erhält der Nutzer eine Rückmeldung der Redaktion.

AP 9: Export

Das Handschriftenportal soll eine möglichst vielfältige weitergehende Nutzung der in ihm enthaltenen Daten ermöglichen. Hierbei ist sowohl an die Datenübernahme durch nationale oder internationale

Portale (DDB/Archivportal-D, Europeana) als auch an die gezielte Verwendung von Erschließungsinformationen durch Forschungsprojekte zu denken. Der betreffende Microservice setzt hierfür den maschinellen Export von Daten aus dem Nachweismodul in die gewünschten Datei- und Fachformate um. Voraussichtlich wird die Exportkomponente als eine HTTP-Schnittstelle (z. B. OAI-PMH) realisiert werden. Notwendige Exportformate sind TEI-XML (internationaler Standard und zentrales Austauschformat von MM), METS/TEI (Format des DFG-Viewers) und Dublin Core (für die Realisierung eines OAI-PMH-Angebots für das Kerndatenset). Weitere mögliche Exportformate sind u. a. METS/MODS, EDM, EAD, MARC21 oder RDF sowie eventuell BibFRAME.

2.3.3.2 Präsentation und verbundene Module

AP 10: Suchmaschinensystem für Weboberfläche

Der Aufbau eines zentralen Suchmaschinenservers stellt das wesentliche Bindeglied zwischen Nachweismodul und Weboberfläche dar. Zu dessen Realisierung wird eine Infrastrukturkomponente auf der Basis von Apache Solr entwickelt und implementiert. Apache Solr basiert auf der Programmbibliothek Apache Lucene. Beide Produkte sind weit verbreitet und bilden in vielen Projekten die Grundlage für die effiziente Erstellung moderner Suchmaschinenanwendungen. Es handelt sich um quelloffene Software, die von einer großen und zuverlässigen Gemeinschaft gepflegt und geprüft wird. Apache Solr und Lucene ermöglichen die Anwendung moderner Technologien auf der Basis erprobter und ausgereifter Softwarekomponenten.

Die Entwicklung einer eigenen Suchmaschinenanwendung ist erforderlich, um dem im vorliegenden Projekt adressierten Anwendungsfall vollständig entsprechen zu können. Die Verwendung generischer kommerzieller Angebote kommt nicht in Betracht, da dort strukturierte Norm- und Metadaten nicht in entsprechendem Umfang zur Herstellung individueller Oberflächenfunktionen eingesetzt werden können. Hierzu zählt insbesondere die Navigation in den Beständen und die Auswahl (Facettierung, Data drilling) von Teilmengen auf der Basis strukturierter Meta- und Normdaten. Um die für die Realisierung des Portals notwendigen Funktionen anwendungsspezifisch herstellen zu können, werden Open Source-Softwarekomponenten eingesetzt, wodurch gleichzeitig volle Kontrolle über das Endprodukt gewährleistet ist.

Die in der zweiten Phase vorgesehene Integration von Annotationen und den damit verbundenen Zugangsberechtigungen erfordern eine datenschutzkonforme Implementierung der Suchfunktion ohne Zugänglichkeit für Dritte. Eine pure Volltextsuche über alle Felder soll eine einfache und ergonomische Suchmöglichkeit bieten, die heutigen Suchgewohnheiten entgegenkommt und um die genannten zielgerichteten Funktionen zu ergänzen ist.

AP 10.1: Indexschema und Hostingkonzept

Zur Realisierung der Suchfunktion ist zunächst ein auf den Anwendungsfall zugeschnittenes Indexschema zu entwerfen. Um die angestrebte hohe Nutzerfreundlichkeit und flexible Nutzung der Weboberfläche mit einem effektiven und wirtschaftlichen Betrieb der Suchmaschine zu vereinen, ist nach einer detaillierten Anforderungserhebung ein entsprechendes Schema zu entwickeln (Apache Solr Reference Guide, Schema Design <https://cwiki.apache.org/confluence/x/Z4DxAQ>). Dieses umfasst alle relevanten Datenarten, insbesondere die strukturierten Metadaten und die unterschiedlichen Volltextquellen. Die reine bedarfsorientierte Feststellung und Differenzierung von Durchsuchbarkeit, Facettierbarkeit, Mehrwertigkeit und Wiedergabe in der Oberfläche ist in einem zweiten Schritt um

zusätzliche Prozessierungsschritte, die eine Vervielfältigung der Felder und die Erweiterung der Feldtypen erfordern, zu ergänzen. Zu diesen Prozessierungsschritten zählen beispielsweise die Stammformreduktion, Segmentierung von Zeichenketten in Wörtern, automatisierte Anreicherung von Synonymen (Verweisungsformen), Anpassung von Groß- und Kleinschreibung und ähnliche Verfahren. Um den Prinzipien der agilen Softwareentwicklung Rechnung zu tragen, ist davon auszugehen, dass die Weiterentwicklung und Anpassung des Indexschemas ein dauerhaftes Arbeitspaket über die gesamte Projektlaufzeit sein wird.

Um der Verschiedenartigkeit der Daten und insbesondere ihrer Veränderlichkeit im produktiven Betrieb gerecht zu werden, ist ein angepasstes Hostingkonzept erforderlich, welches verschiedene Kriterien wie Performance, Aktualisierbarkeit und redundanzbasierte Absicherung der Verfügbarkeit des Systems vereint. Das Hostingkonzept soll Abläufe zu Konvertierung, Speicherung und Aktualisierung von Bild- und Metadaten umfassen, ein Schema zur Modellierung von URIs vorgeben und implementieren, den Datenaustausch mit dem Nachweissystem bei der SBB-PK regulieren sowie ein anwendungsbereites Geschäftsmodell inklusive Dienstleistungsvereinbarungen, Klärung der Nutzungsrechte an den Digitalisaten und ein Modell zur Planung und Bereitstellung der Ressourcen umfassen.

AP 10.2: Indexierungskomponente

Zum Befüllen und Aktualisieren des Suchmaschinenindex ist eine Softwarekomponente (SolrUpdate-Service) zur Umwandlung der Daten aus dem Nachweissystem und gegebenenfalls aus weiteren Quellen erforderlich. Diese soll den Suchmaschinenindex ereignisgesteuert über die Message-Queue mit neuen Daten befüllen sowie vorhandene Daten aktualisieren oder entfernen. Die Veröffentlichung der Indexierungskomponente als Open Source ist ein wichtiger Schritt, um die Nachhaltigkeit der Entwicklungsergebnisse zu verbessern.

Für die Übernahme der Daten aus dem Nachweissystem sind diese für die Indexierung in Echtzeit in indexierungsfähige Datenformate zu konvertieren und an den Index zu schicken. Erst nach erfolgter Indexierung kann die Message-Queue um den entsprechenden Vorgang entlastet werden.

AP 11: Weboberfläche

Die Weboberfläche umfasst alle Komponenten, die den Endnutzer*innen zur Verfügung gestellt werden. Hierzu zählen die Präsentation von digitalen Bildern, Texten und statischen Inhalten sowie die Suchfunktionen zur Recherche in Metadaten, Volltexten und später Annotationen.

Für die Anzeige von Digitalisaten, Beschreibungsdokumenten und strukturierten Daten ist zum Zeitpunkt der Antragstellung der Einsatz des IIIF-Viewers Mirador vorgesehen, wobei der DFG-Viewer mit eingebunden wird.

AP 11.1: Web Content Management System

Das Web Content Management System (CMS) wird die Grundlage für die Implementierung aller portalspezifischen Funktionen bilden. Es ist vorgesehen, ein etabliertes System mit großer Community zu wählen, um eine langfristig nachhaltige Lösung und Weiterentwicklung des Systems zu gewährleisten. Zum Zeitpunkt der Projektbeantragung ist hierfür das CMS TYPO3 vorgesehen, welches insbesondere in Europa über eine große Community verfügt.

Die portalspezifischen Funktionen für Suche und Anzeige sollen in Form von separaten TYPO3-Extensions implementiert werden. Die Kommunikation der verschiedenen Komponenten der Weboberfläche wird überwiegend clientseitig im Browser des Nutzers realisiert.

Grundfunktionen und redaktionelle Anwendung

Nach einer konzeptionellen Phase, die der Erhebung konkreter Anforderungen auf redaktioneller Ebene dient, werden die redaktionellen Komponenten des CMS definiert und gegebenenfalls softwaretechnisch um die erforderlichen Funktionen zur Abbildung der redaktionellen Anforderungen erweitert. Hierzu können News-System, Kalender und viele andere einschlägige Funktionen gehören, zu deren Umsetzung möglichst auf bestehende Extensions zurückgegriffen werden soll.

Um die Weboberfläche nahtlos als Komponente im gesamten System integrieren zu können, ist es erforderlich, eine Microservice-Integration zu implementieren. Die Details zu den konkreten Anforderungen sind zu Projektbeginn festzulegen und deren Implementierung im Anschluss auszuführen.

Design und Template

Zum vorliegenden Arbeitspaket gehört die Erstellung eines Web- und Screendesigns sowie dessen Implementierung als TYPO3 Template. Der Weboberfläche soll ein eigener Styleguide zugrunde gelegt und um Spezifikationen zur Gestaltung der Weboberfläche unter ergonomischen Gesichtspunkten ergänzt werden. Dies soll unter Berücksichtigung der Anforderungen des Responsive Webdesigns geschehen. Die für die Nutzerführung wesentlichen Bereiche der Web-Präsentation werden auch auf Englisch angeboten werden.

Diese Aufgaben sollen per Ausschreibung extern vergeben werden, da es sich um fachlich einzugrenzende Aufgaben handelt, die besondere Kenntnisse und Fähigkeiten erfordern, die mit anderen Arbeitspaketen keine Überschneidungen haben. Die Umsetzung soll in enger Kooperation mit einem Auftragnehmer iterativ über einen längeren Teil des Projektzeitraums gestaltet werden. Die erarbeiteten Designs werden frei von Rechten Dritter sein, so dass sie für die weitere Öffentlichkeitsarbeit im Projekt nachgenutzt werden können. Die englische Übersetzung erfolgt in Eigenleistung.

AP 11.2: Integriertes Discovery System

Die TYPO3-Extension zur Suchfunktion implementiert eine vollwertige Discovery Oberfläche, die auf das Suchmaschinensystem zugreift. Es ist vorgesehen, eine vorhandene Extension zur Integration von Solr-basierten Suchfunktionen unter generischen Gesichtspunkten so weiterzuentwickeln, dass die Projektziele damit erreicht werden können. Neben verschiedenen, auf die unterschiedlichen Datenarten abgestimmten Suchmöglichkeiten sollen die Ergebnisanzeigen individuell sortiert und facettiert werden können. Um den differenzierten Suchgewohnheiten und Bedürfnissen der Anwender*innen entsprechen zu können, ist es erforderlich, die Verfügbarkeit folgender Oberflächenfunktionen sicherzustellen:

- Eingabefeld mit Auswahlmöglichkeit der zu durchsuchenden Datenarten
- Browsing
- Facettierbarkeit der Trefferliste nach unterschiedlichen Kriterien aus strukturierten Meta- und Normdaten
- Zuschaltung der Volltextsuche

Gefundene Suchtreffer sollen bequem in die Anzeige übernommen werden können, hierfür ist ein entsprechender Austausch mit der Anzeige Komponente zu realisieren, wobei verschiedene Lösungsmöglichkeiten in Frage kommen, wie Übernahme aus der Trefferliste per Drag and Drop, API-Aufruf in Javascript oder andere Mechanismen. Im Vordergrund steht dabei stets, eine ergonomische Oberfläche mit hoher, intuitiver Nutzbarkeit zu gewährleisten.

AP 11.3: Einbettung von Mirador in TYPO3

Zur Integration von Mirador in das CMS TYPO3 soll eine separate TYPO3-Extension erstellt werden. Diese Extension soll im Layout des TYPO3-Frontends den IIIF-Viewer Mirador als individuelles Content Element zur Verfügung stellen. Die besondere Herausforderung des Arbeitspaketes besteht darin, die Webanwendung Mirador so zu kapseln, dass sie auch im laufenden Betrieb mit der offiziellen Softwareversion von Mirador aktualisiert werden kann und zugleich der Datenfluss zwischen Mirador und den weiteren Komponenten der Oberfläche stattfinden kann.

Die Extension soll im offiziellen Extension-Repository von TYPO3 abgelegt werden. Die Veröffentlichung als Open Source ist ein wichtiger Schritt, um die Nachhaltigkeit der Entwicklungsergebnisse zu verbessern.

AP 11.4: Anzeige von Textdokumenten

Um die integrierte Präsentation der verschiedenen Medientypen, die im Handschriftenportal vorgehalten werden, zu ermöglichen, soll im vorliegenden Projekt Mirador als universelle Präsentationskomponente eingesetzt werden. Es handelt sich um den einzigen etablierten mehrdokumentenfähigen Viewer für IIIF-kompatible Digitalisate, der zudem von einer großen und stabilen Community entwickelt wird. Zu den relevanten Medientypen zählen neben den Digitalisaten verschiedene Texte, insbesondere Beschreibungen, aber auch Transkriptionen und Übersetzungen im Kontext der Annotationen. Diese textlichen Medientypen werden von Mirador bislang noch nicht unterstützt. Mirador wurde primär entwickelt, um zweidimensionale digitale Bilder in mehreren Fenstern interaktiv zu präsentieren, ist insofern also für die Realisierung der Anzeige Komponente des Handschriftenportals hervorragend geeignet. In den vergangenen Jahren wurde von der Community um Mirador die Unterstützung weiterer Medientypen angestrebt, insbesondere die Darstellung von Videos und 3D-Modellen wurde hier verfolgt. Das vorliegende Arbeitspaket soll diese Bestrebungen unterstützen und ein Modul zum Rendern von XML/TEI, und gegebenenfalls weiteren Textformaten, als zusätzlich unterstützte Medientypen in Mirador implementieren.

Die reine Anzeige der Dokumente soll dabei um die Adressierbarkeit mittels URNs auf der Basis von Canonical Text Services (CTS)¹⁶ erweitert werden. Dies ist erforderlich, um auch diese Medientypen mit Annotationen versehen zu können. Möglicherweise sind hierzu auch ergänzende Spezifikationen zu den relevanten APIs des IIIF Frameworks zu erstellen. Alle Softwareentwicklungen und Spezifikationen sind im Sinne ihrer nachhaltigen Entwicklung mit der internationalen Community um IIIF und Mirador abzustimmen.

¹⁶ Canonical Text Services, Verfahren zu permanenten Adressierung von Textbestandteilen mittels URNs, siehe u. a. DOI: 10.13189/lls.2017.050209.

AP 11.5: Interaktion zwischen Dokumentfenstern

Mirador ist zwar grundsätzlich in der Lage, verschiedene Dokumente auf einer Oberfläche in separaten Fenstern darzustellen. Zwischen diesen Visualisierungen besteht in der aktuellen Version der Software jedoch keine Verknüpfbarkeit, die Dokumente werden völlig unabhängig nebeneinander dargestellt. Um das Potential der parallelen Darstellung verschiedener Medientypen auf der gleichen Oberfläche ausschöpfen zu können, wird angestrebt, aus einem Fenster der Mirador-Oberfläche die Darstellung in anderen Fenstern beeinflussen zu können. Beispielsweise soll es möglich sein, in Annotationen oder Textdokumenten vorhandene Verknüpfungen, entweder über die Definition geometrischer Bereiche auf dem Digitalisat oder spezifischer Textabschnitte per CTS in benachbarten Mirador-Fenstern zu öffnen. Hierzu ist es zunächst erforderlich, in Mirador eine entsprechende Komponente zu implementieren, die es ermöglicht, Daten über Ereignisse, die aus Nutzerinteraktionen hervorgehen, in Echtzeit an andere Dokumentenfenster zu übertragen. Zur Modellierung der Daten für die Übertragung dieser Informationen ist in enger Abstimmung mit der Community ein entsprechender Standard zu definieren und zu implementieren. Im Anschluss kann die konkrete Adressierung von geometrischen oder inhaltlichen Abschnitten der darzustellenden Dokumente erörtert und in geeigneter Weise implementiert werden.

Die Realisierung dieses Teilarbeitspaketes erfordert eine anspruchsvolle konzeptionelle Phase, die in enger Abstimmung mit der Mirador- und IIF-Community stattfinden muss. Die anschließende Implementierung wird durch die Übernahme der Projektergebnisse in die offiziellen Softwareversionen der entsprechenden Open Source Produkte abgeschlossen.

AP 11.6: JSONBLOB Service

Die Anwendung Mirador benötigt einen eigenen Webservice zur Zwischenspeicherung von JSON-Fragmenten. Per Voreinstellung wird hier bisher ein kostenloser Service eines Anbieters in den USA verwendet. Aus Gründen des Datenschutzes ist im vorliegenden Projekt von den Betreibern jedoch ein eigener Service einzusetzen, der zudem die Unabhängigkeit des Konstruktes insgesamt sicherstellt. Die Software ist unter einer Open Source-Lizenz verfügbar und muss für den Einsatz angepasst werden. Im vorliegenden Teilarbeitspaket ist daher neben der infrastrukturellen Konzeption festzulegen, welche zusätzlichen Schritte für die Microservice-Integration erforderlich sind.

AP 11.7: Interoperabilität der Präsentationskomponenten

Mit der Entscheidung, IIF als Basistechnologie für die Weboberfläche des neuen Handschriftenportals zu wählen, ist impliziert, dass strukturierte Metadaten und digitale Bilder über Webschnittstellen auf interoperable Weise zur Verfügung gestellt werden.

Die Informationen zu Digitalisaten von Handschriften bilden eine Präsentationsentität, auf die aus dem KOD referenziert wird. Diese Präsentationsentität modelliert die UBL. Auf diese Weise sollen perspektivisch Einsprungpunkte in Präsentationen (für die Referenzierung aus Beschreibungsdokumenten heraus) differenziert werden. Faktisch werden Meta- und Bilddaten durch ihre IIF-kompatible Bereitstellung mit Identifiern (URIs) versehen, deren persistente Verfügbarkeit sicherzustellen ist. In enger Zusammenarbeit mit der Datenmodellierung im Nachweismodul (AP 6) und dem Konzept für die IIF-Hostingdienstleistungen (AP 16) sind zur Modellierung der IIF-Daten (Manifeste) entsprechende technische Designentscheidungen zu treffen: Um die zu gestaltenden URIs im Idealfall als URLs auflösen zu können, haben diese Designentscheidungen unmittelbare Auswirkungen auf die Weboberfläche und

insbesondere einzusetzende Proxyserver. Per Content Negotiation können dann visuelle Präsentationen und maschinell verarbeitbare Daten unter identischen URIs/URLs abgerufen werden.

Um die bestmögliche Nutzbarkeit und Interoperabilität der nach IIIF-Standard modellierten Daten sicherstellen zu können, ist es vorgesehen, zu zwei Zeitpunkten im Projektverlauf Best Practice Workshops innerhalb der IIIF-Community durchzuführen, mit denen die Modellierung der Daten und die Implementierung der Infrastrukturen auf ihre bestmögliche Verknüpfbarkeit hin untersucht und optimiert werden. Auf diese Weise können auch Entwicklungen gebündelt und zusammengeführt werden, die derzeit an verschiedenen Stellen angedacht oder geplant sind (vgl. z. B. die IIIF-Konferenz an der Bibliotheca Apostolica Vaticana: content.fromthepage.com/text-beyond-annotations-at-iiif-vatican).

AP 12: Integration von Digitalisaten

Da das Handschriftenportal sich als zentrale Nachweismöglichkeit für digitalisierte Handschriften versteht, ist von Antragstellerseite eine möglichst hohe Abdeckung der enthaltenen KODs mit Digitalisatpräsentationen angestrebt. Diese Digitalisate liegen in der Regel dezentral bei den besitzenden Einrichtungen. Die Zulieferung der Digitalisate kann in Form der Bereitstellung über Webschnittstellen oder einfachen Dateien erfolgen.

Zur webbasierten Anzeige der Digitalisate stehen grundsätzlich zwei Technologien zur Verfügung: die Bereitstellung im METS-Format über OAI-PMH, kompatibel zum DFG-Viewer, oder als Bereitstellung über Webschnittstellen nach dem IIIF-Standard.

Im Hinblick auf die Nachnutzung integrierter Workspacelösungen (Mirador), die Integrierbarkeit in internationale Strukturen und die Anwendung von Linked Data Prinzipien wird zur Realisierung des neuen Handschriftenportals IIIF als primäre Lösung zur Präsentation der Digitalisate auf der Weboberfläche angestrebt.

Die konventionelle Bereitstellung von Digitalisaten nach den mit dem DFG-Viewer kompatiblen Standards OAI-PMH und METS ist jedoch bei vielen zuliefernden Einrichtungen standardmäßig implementiert und soll daher stets zur Anzeige angeboten werden und im Falle der Nichtverfügbarkeit von IIIF-Daten als Anzeige verwendet werden.

Die Verfügbarkeit von Digitalisaten nach den vom DFG-Viewer verwendeten Standards ist in den strukturierten Metadaten im Nachweismodul verzeichnet und findet über den Suchmaschinendindex in die Weboberfläche. Dort kann zunächst nach dem üblichen Verfahren die Ansicht des Digitalisates im DFG-Viewer aufgerufen werden. Die dem DFG-Viewer zugrundeliegende Open-Source-Software Kitodo.Presentation wird unter anderem von der SLUB Dresden und der UB Leipzig im Rahmen eines EFRE-Projektes um IIIF-kompatible Komponenten erweitert. Im vorliegenden Arbeitspaket soll zunächst die konventionelle Integration des DFG-Viewers im Handschriftenportal ermöglicht werden. Im Weiteren sollen dann mögliche Zusammenhänge zwischen einer erneuerten, IIIF-fähigen Version von Kitodo.Presentation und der Oberfläche des Handschriftenportals geprüft und mögliche Synergien genutzt werden.

2.3.3.3 Funktionsübergreifende Module

AP 13: Identity-Management

Die Weboberfläche steht allen Endnutzern für die Recherche offen. Alle schreibenden Zugänge zum Backend des Systems und seinen Daten setzen jedoch die Autorisierung und Authentifizierung an einer zentralen mehrstufigen Identity-Management-Komponente voraus.

Diese als Microservice angelegte Komponente steuert differenziert Identitäten, Rechte und Rollen von Administratoren, Redakteuren und angemeldeten Nutzer*innen im Nachweismodul. Außerdem werden mit ihrer Hilfe die Berechtigungen externer Partner für den Zugriff auf die Im- und Exportkomponente geregelt. Hinsichtlich der Weboberfläche werden über sie die Arbeit der Administratoren und Redakteure mit dem CMS gesteuert.

Das Identity-Management wird als Service umgesetzt, der die Möglichkeit des Single-Sign-On unterstützt. Die Autorisierung und Authentifizierung gegenüber dem Identity-Management wird daher entweder durch eine direkte Anmeldung an diesem oder durch die Übernahme von Nutzerinformationen aus anderen Identity-Providern wie Shibboleth oder LDAP geregelt. Die Auswahl kann für jedes System beliebig konfiguriert werden. Die Konfigurations- und Steuerungsdaten werden in einer relationalen Datenbank abgelegt. Als externe Schnittstelle ist das OAuth2-Protokoll mit der Integration des JSON Web Tokens implementiert. So können alle Komponenten des Gesamtsystems die Nutzeridentität und die Nutzerrechte prüfen und entsprechend reagieren. Die dazu benötigten Daten werden durch die Identity-Management-Komponente persistiert. Hierbei kann die SBB-PK auf eine im Zusammenhang der Verbesserung der Software für den Gesamtkatalog der Wiegendrucke entwickelte Lösung zurückgreifen, die aber angepasst und erweitert werden muss.

AP 14: Schema- und Skriptpflege

Alle Daten-Import- und -Exportprozesse werden durch eine Überprüfung der Datenstrukturen begleitet. Zu diesem Zweck sind Schemata zu erstellen und pflegen, welche die Validität der Daten gemäß den zu bedienenden Standards überprüft. Hier werden vor allem die Formate TEI, METS und Dublin-Core zu berücksichtigen sein. Im Rahmen der Kooperation mit der DDB, dem Archivportal-D und Kalliope wird auch EAD für die Abgabe von Daten an die Europeana das Format EDM eine Rolle spielen müssen.

Der Im- und Export von Daten erfordert die Konversion von Daten aus anderen Formaten in das Format des Nachweismoduls und umgekehrt. Die Konversion erfolgt über Schnittstellenskripte. Diese werden in der Regel in XSLT realisiert. Da sich Regelwerke durchaus ändern und durch veränderte Kooperationszenarien neue Bedarfe entstehen können, besteht dauerhafter Bedarf an der Pflege der Schnittstellenskripte. Die Pflege der Schemata und Skripte wird über den Projektzeitraum hinaus von der HAB gewährleistet.

2.3.3.4 Aufbau der technischen Infrastruktur

AP 15: Aufbau und Betrieb der Infrastruktur und Administration

Für das gesamte System des neuen Handschriftenportals muss an der SBB-PK eine Infrastruktur aus VM-Servern, Datenbanken, Netzwerken, Applikations- und Indexservern sowie einer Message-Queue-

Lösung aufgebaut und betreut werden. Diese Infrastruktur umfasst alle drei Umgebungen: Entwicklung, Test und Produktion. Notwendig sind die Konzeption und Betreuung der produktiven Architektur unter Berücksichtigung von IT-Sicherheit (Firewall) und Storage-Planung.

Die benötigte IT-Infrastruktur wird als Nutzung virtueller Server im größeren Rechnerverbund der SBB-PK realisiert. Dieser Aufbau ist kostengünstiger als eine separate Hardware-Infrastruktur, da er auf die optimale Nutzung vorhandener Rechnerkapazitäten hin ausgelegt ist. Die beantragten Mittel decken die der SBB hierfür tatsächlich entstehenden Kosten und liegen deutlich unter denjenigen für die Beschaffung von separater Hardware oder für die Anmietung der notwendigen Rechenleistung bei privaten Anbietern.

Für den Betrieb des Gesamtsystems werden außerdem Plattform-Services aufgebaut und betreut. Dazu gehören: zentrales Logging (ELK-Stack), zentrales Monitoring (Nagios), Service-Discovery (Netflix-Eureka) mit Routing und zentrales Konfigurationsmanagement (Archaius). Diese Infrastruktur ist in der SBB-PK bereits im produktiven Einsatz und wird für das Handschriftenportal angepasst und erweitert.

AP 16: IIIF-Hosting und Geschäftsmodell

Im Rahmen der Projektrealisierung soll ein Hostingkonzept für IIIF-Digitalisate entwickelt und Einrichtungen zur Verfügung gestellt werden, die selbst keine IIIF-Server betreiben können oder möchten. Dies schließt insbesondere auch (Klein-) Sammlungen ein, die über keinerlei eigene Präsentation von Bilddaten verfügen.¹⁷ Basis dieses Angebotes soll die IIIF-Infrastruktur der UBL sein, die seit 2016 alle Digitalisierungsprojekte automatisch auch in IIIF konvertiert¹⁸ und eine entsprechende Datenverarbeitungs- und Hostinginfrastruktur betreibt.

Gegenstand des Arbeitspaktes soll die Definition des tatsächlichen Angebotes sein. Wichtige Aspekte sind dabei, welche Aufgaben bezüglich der Konvertierung der Bild- und Metadaten von der UBL übernommen werden können, welche Vereinbarungen mit den zuliefernden Einrichtungen getroffen werden müssen und welche Möglichkeiten bei der Lizenzvergabe für Bilddaten zur Verfügung stehen. Dabei sind unterschiedliche technische und rechtliche Aspekte abzuklären und ein entsprechendes Geschäftsmodell zu entwickeln, welches auch die Konditionen der Dienstleistung nach dem Ende der Projektförderung regelt. Ziel ist es, die Bereitstellung möglichst vieler Digitalisate zu ermöglichen, um die Abdeckung der Bestände im Handschriftenportal auf hohem Niveau zu realisieren.

Im ersten Arbeitsschritt soll ein Grundkonzept des Geschäftsablaufes zur Übernahme von Fremddaten erstellt werden. Hierzu sind organisatorische Grundlagen des Datenflusses und der Datenkonvertierung sowie Festlegungen und Abläufe zu Dienstleistungsvereinbarungen zu treffen. Dabei sind verschiedene organisatorische, rechtliche und technische Aspekte zu berücksichtigen.

Die getroffenen Festlegungen sollen in einer zweiten Phase mit Pilotkunden praktisch erprobt und die technischen und organisatorischen Prozesse entsprechend angepasst werden. Neben der Überprüfung der Geschäftsabläufe auf ihre Praxistauglichkeit sollen in diesem zweiten Arbeitsschritt auch produktiv

¹⁷ Nach einer Erhebung der Handschriftenzentren im Auftrag der DFG aus dem Jahr 2006 ist mit ca. 1.700 Handschriften aus Kleinsammlungen zu rechnen, wobei nach den Erfahrungen aus den beiden seit 2010 durchgeführten Kleinsammlungsprojekten der UB Leipzig eine deutliche Dunkelziffer veranschlagt werden muss, so dass die Zahl sich auf 2.000-3.000 Handschriften erhöhen könnte.

¹⁸ Vgl. <https://digital.ub.uni-leipzig.de/mirador/index.php>.

nutzbare Digitalisate von Dritten für die Darstellung im Handschriftenportal dauerhaft zur Verfügung gestellt werden.

Nach der erfolgreichen Bewältigung der ersten beiden Arbeitsschritte wird das Arbeitspaket durch die Überführung in den regulären Betrieb abgeschlossen. Die organisatorischen Grundlagen sowie die erforderliche technische Infrastruktur werden nun so in der UBL verankert, dass sie als festgelegte Geschäftsabläufe im Regelbetrieb stattfinden können. Dies umfasst Implementierung von Routinen zur Datenübernahme von Dritten einschließlich der Klärung und vertraglichen Absicherung von Nutzungsrechten, URI-Modellierung, Art und Umfang des Dienstleistungsangebotes, Festlegung von Datenaustauschformaten und die Überstellung und Konvertierung von Metadaten und digitalen Bildern.

AP 17: Entwicklungsunterstützende Systeme

Als Voraussetzung für eine effiziente Softwareentwicklung in verteilten Teams wird zentral bei der SBB-PK eine Entwicklungsinfrastruktur aufgebaut und während der gesamten Projektdauer gepflegt.

AP 17.1: Versionskontrolle und Continuous Integration (CI) (Aufbau und Betreuung)

Alle Komponenten des Gesamtsystems werden im Laufe des Projekts an der SBB-PK zusammengeführt und integriert. Zu Beginn des Projekts wird an der SBB-PK eine gemeinschaftlich konzeptionierte zentrale Integrations- und Testumgebung aufgebaut, wobei bereits vorhandene Infrastruktur (Jenkins- und Nexus-Server) der SBB-PK für diese Zwecke teilweise nachgenutzt werden kann. Die verteilte Entwicklung von Komponenten wird in einem gemeinsamen Versionskontrollsystem (Git) erfolgen und der Workflow in der Entwicklung durch ein Versionsmanagementtool (GitLab) unterstützt. Alle Projektpartner werden den entwickelten Quellcode permanent in ein zentrales Repository überführen, das als Basis für die kontinuierliche Integration des gesamten Systems dient. Jede zusätzliche Funktionalität soll möglichst zeitgleich mit einem dazugehörigen Test entwickelt werden. In regelmäßigen Abständen (in einem Nightly-Build-Verfahren) wird automatisch ein Build und Deployment durchgeführt. Integrationstests, die teamübergreifend konzipiert werden, werden automatisiert ausgeführt und die Test-Reports werden für alle beteiligten Entwickler zur Verfügung gestellt.

Seitens der SBB-PK wird eine Continuous-Integration und -Delivery-Pipeline für alle Komponenten des Systems implementiert. Die Integrations- und Auslieferungsprozesse werden überwacht und optimiert. Im Fehlerfall werden eine Voranalyse durchgeführt, Deployment-Fehler behoben und Fehler in den Softwarekomponenten an die Entwickler zurückgemeldet.

Bei einer erfolgreichen Integration des Gesamtsystems werden alle Artefakte auf ein dediziertes Test-System automatisiert ausgerollt (deployed), auf dem eine Qualitätskontrolle durch registrierte Nutzer erfolgen kann (zur Information der Community hierüber s. AP 2). Ein lauffähiger Stand für Demo- und Testzwecke ist damit ständig verfügbar.

AP 17.2: Weitere Werkzeuge zur Unterstützung der Entwicklung (Aufbau und Betreuung)

Für den Austausch binärer Artefakte zwischen den Projektpartnern wird ein Artefakt-Repository benötigt. Zu diesem Zweck soll Nexus als eine effektive Plattform eingesetzt werden. Sie bietet eine zentrale Speicherung und den Zugriff auf Artefakte und Metadaten für Entwickler. Als Projektmanagementsoft-

ware und Ticketsystem zur Releaseplanung, Ticketverwaltung und Dokumentenablage für alle Projektpartner wird Redmine eingesetzt. Der Einsatz der Plattformen wird unter den Projektpartnern abgestimmt, Aufbau und Betreuung finden an der SBB-PK statt.

2.3.4 Meilensteinplanung

Das Arbeitsprogramm der ersten Projektphase ist in sechs Abschnitte von je sechs Monaten gegliedert. Diese Abschnitte unterteilen sich ihrerseits in je sechs Sprints von vier Wochen. Für den Abschluss jedes Sechsmonatsblocks sind Meilensteine definiert, die die zu diesem Zeitpunkt erreichten Projektergebnisse beschreiben. Alpha- und Beta-Versionen dieser Ergebnisse sind nach Ausreifungsgrad und Zugänglichkeit unterschiedene Vorstufen zur Produktivität eines Moduls. Am Ende der ersten Phase sollen alle bis dahin entwickelten Module in einer Version 1 produktiv sein. Folgende Meilensteine und Projektergebnisse sind vorgesehen:

1. Meilenstein (nach 6 Monaten):

- Testworkflow modellhafter KODs in prototypischer Präsentation ist funktional
- CI-Infrastruktur und Testumgebung stehen bereit
- Imagedigitalisierung, OCR-Bearbeitung ist abgeschlossen
- Blog zur Information künftiger Nutzer ist eingerichtet
- CMS und Mirador sind prototypisch integriert
- Entwurf des Indexschemas wurde definiert
- Suchmaschinen-Infrastruktur für die Entwicklung ist nutzbar
- Deployment Workflow von Leipzig nach Berlin ist nutzbar
- Infrastruktur für Übernahme externer IIF-Daten ist konzipiert
- Erster IIF-Workshop ist durchgeführt
- Erste Umfrage ist versandt

2. Meilenstein (nach 12 Monaten):

- Volltextauszeichnung der Kataloge abgeschlossen
- Minimaldatenset eines Teils der echten KODs wird in Alpha-Version der Präsentation im Browser dargestellt
- Alpha-Version des Nachweismoduls steht
- Alpha-Version des Normdatenservices für die Präsentation steht. Verweisformen aus Normdaten der GND werden in den Suchmaschinenindex der Weboberfläche eingespielt
- Migration produktiver Daten hat begonnen
- Identity-Management für Nachweismodul und CMS steht bereit
- Indexierungskomponente für die Suchmaschine existiert im Alpha-Stadium
- CMS-Extension für die Suche existiert im Alpha-Stadium
- Infrastruktur für Übernahme externer IIF-Daten ist installiert
- Usability-Konzept und Webdesign des Dienstleisters liegen vor

3. Meilenstein (nach 18 Monaten):

- Prototypische Erfassung mit GUI steht
- Workshop zu Nutzerakzeptanz der Alpha-Version des Systems durchgeführt (Treffen von Stakeholdern, die nicht Projektbeteiligte sind)
- Anzeige von Textdokumenten im Mirador befindet sich im Alpha-Stadium

- Zweiter IIF-Workshop ist durchgeführt
- CMS-Extension für die Suche befindet sich im Beta-Stadium
- Indexierungskomponente für die Suchmaschine existiert im Beta-Stadium
- Anbindung des IDM an das CMS ist implementiert
- Webdesign und Usability-Konzept sind implementiert
- DFG-Viewer kann aus der Weboberfläche angesprochen werden

4. Meilenstein (nach 24 Monaten):

- Beta-Version der Präsentation geht online
- Alle KODs sind mindestens mit ihrem Minimaldatenset in der Beta-Version der Präsentation recherchier- und sichtbar
- Alpha-Version der Erfassung steht; Anbindung der Erfassung an den Normdatenservice ist realisiert
- Alpha-Version des Imports steht; Import von Testdaten aus Bibliotheksverbänden möglich
- Interaktion von Dokumentenansichten im Mirador befindet sich in der Alpha-Version
- Geschäftsmodell für die Übernahme von Digitalisaten in den IIF-Server ist anwendungsbereit
- JSONBLOB-Server ist produktiv nutzbar
- Echtzeit-Abgleich zwischen Nachweis und Suchmaschine ist produktiv implementiert

5. Meilenstein (30 Monate):

- Beta-Version Nachweismodul und Erfassung
- Beta-Version Im- und Export (zentrale Formate) stehen
- Workshop Beta-Version des Systems durchgeführt
- Daten von ersten Pilot-Einrichtungen sind in den IIF-Server eingespielt und produktiv nutzbar
- Alle Erweiterungen des CMS sind produktiv und die Extensions als Open Source veröffentlicht
- Alle Erweiterungen des Viewers sind produktiv und in die offizielle Version des Quellcodes übernommen
- Die Indexierungskomponente ist produktiv und wurde als Open Source veröffentlicht
- Die zweite Umfrage ist versandt

6. Meilenstein (36 Monate):

- Alle Module in Version 1 produktiv
- Dokumentation abgeschlossen
- Wertebearbeitung der Kerndaten so weit wie möglich abgeschlossen
- Geschäftsprozesse zur Übernahme von Digitalisaten in den IIF-Server sind in der UB Leipzig verstetigt

Die Projektergebnisse können von allen Interessierten kontinuierlich geprüft und erprobt werden. Sobald die Entwicklung eine testbare Reife erlangt hat (etwa nach einem Jahr), werden zu diesem Zweck die Arbeitsergebnisse der jeweiligen Sprints auf eine dedizierte Testumgebung übernommen, wo sie nach einer Registrierung für jedermann zugänglich sind. Daneben wird die Entwicklung von kontinuierlichen Fachtests der Einzelkomponenten begleitet und durch fachliche Tests des Gesamtsystems beschlossen.