


# Die Research Collection der ETH Zürich

**Journal Article****Author(s):**

Hirschmann, Barbara 

**Publication date:**

2018-09-04

**Permanent link:**

<https://doi.org/10.3929/ethz-b-000327719>

**Rights / license:**

[In Copyright - Non-Commercial Use Permitted](#)

**Originally published in:**

ABI Technik 38(3), <https://doi.org/10.1515/abitech-2018-3003>

## Fachbeitrag

Barbara Hirschmann

# Die Research Collection der ETH Zürich

<http://doi.org/10.1515/abitech-2018-3003>

**Zusammenfassung:** Im Sommer 2017 lancierte die ETH-Bibliothek nach rund dreijähriger Projektphase die Research Collection, eine neue Publikationsplattform für die Forschenden an der ETH Zürich. Die Plattform vereint die Funktionen einer Hochschulbibliographie, eines Open-Access-Repository und eines Forschungsdaten-Repository unter einem Dach. Sie wurde auf Basis der Open-Source-Software DSpace implementiert und löste zugleich zwei Vorgängersysteme ab. Heute ist die Research Collection ein zentraler Baustein innerhalb der hochschulweiten Informationsinfrastruktur der ETH Zürich.

**Schlüsselwörter:** Institutional Repository, Publikationen, Forschungsdaten

## ETH Zurich's Research Collection

**Abstract:** In the summer of 2017, after a three-year project period, the ETH Library launched Research Collection, a new publication platform for researchers at ETH Zurich. The platform integrates the functionalities of a university bibliography, an open access repository, and a research data repository, all under one roof. It was implemented based on the open source software DSpace and replaced two legacy systems in one go. Today, the Research Collection constitutes a central building block of ETH Zurich's university-wide information infrastructure.

**Keywords:** Institutional Repository, Publications, Research Data

## 1 Einleitung

Im Rahmen des Projekts E-Publications entwickelte die ETH-Bibliothek eine neue Plattform für die Dokumentation, Publikation und Archivierung des wissenschaftlichen Outputs der Angehörigen der ETH Zürich. Hauptziel dieses von Herbst 2014 bis Sommer 2017 dauernden Projekts war die Integration und Ablösung des Open-Access-Repository ETH E-Collection und der Hochschulbibliographie ETH E-Citations. Die neue Publikationsplattform mit

dem Namen Research Collection bietet darüber hinaus auch einen Workflow für die Publikation bzw. Archivierung von Forschungsdaten an, der bisher über eine weitere Applikation der ETH-Bibliothek, das ETH Data Archive, abgewickelt wurde. So entstand aus drei verteilten Applikationen, drei Produkten und drei bibliotheksinternen Prozessen die Research Collection, ein One-Stop-Shop für die Forschenden der ETH Zürich, über den sie

- ihren Forschungsoutput für die akademische Berichterstattung und für die Erzeugung von Publikationslisten auf ihren Webseiten erfassen,
- Dokumente im Open Access publizieren,
- Forschungsdaten öffentlich zugänglich machen oder archivieren.

## 2 Ausgangslage

Im Jahr 2013, als an der ETH Zürich die ersten Vorüberlegungen zum hier vorgestellten Projekt stattfanden, konnte die ETH-Bibliothek bereits auf über ein Jahrzehnt an Erfahrungen mit der Entwicklung und dem Betrieb von Dienstleistungen zur digitalen Verwaltung und Verbreitung des Publikationsoutputs der Forschenden der ETH Zürich zurückblicken. Auch mit der Publikation und Archivierung von Forschungsdaten hatte sie zu diesem Zeitpunkt bereits erste Erfahrungen gesammelt.

### 2.1 Open-Access-Repository und Hochschulbibliographie

Als eine der ersten Hochschulbibliotheken weltweit hatte die ETH-Bibliothek im Jahr 2001 ein Institutional Repository für die ETH Zürich unter dem Namen E-Collection lanciert. Die E-Collection wurde anschließend sukzessive zu einem den damaligen Standards entsprechenden Dokumentenserver ausgebaut. Einen wichtigen Entwicklungsschritt machte das Repository 2008 mit der Migration von einer inhouse entwickelten Lösung auf die Open-Source-Software Fedora,<sup>1</sup> womit zahlreiche funktionale und

<sup>1</sup> <http://duraspace.org/fedora/> (30.05.2018).

technische Erweiterungen umgesetzt werden konnten. Im selben Jahr verabschiedete die Hochschulleitung der ETH Zürich eine Open-Access-Policy für die Hochschule und benannte darin die E-Collection als Instrument für die aktive Umsetzung der Policy: Die Forschenden der ETH waren damit angehalten, Artikel und anderen Forschungsausgaben über den Weg der Zweitveröffentlichung in der E-Collection zugänglich zu machen.<sup>2</sup>

Parallel zu den Aktivitäten rund um die E-Collection tat sich für die ETH-Bibliothek ein weiteres Handlungsfeld auf: In Zusammenhang mit den Prozessen für die akademische Berichterstattung der Professuren war an der ETH Zürich das Bedürfnis nach einer qualitätsgeprüften Hochschulbibliographie entstanden. Diesem Bedarf kam die ETH-Bibliothek mit der Entwicklung von ETH E-Citations nach. Die Bibliographie wurde im Frühjahr 2010 lanciert, ebenfalls basierend auf der Software Fedora; dies einerseits, um Synergien bei der technischen Umsetzung zu nutzen, andererseits stand bereits damals die Idee im Raum, E-Citations und E-Collection zu einem späteren Zeitpunkt zusammenzuführen.<sup>3</sup>

In den Folgejahren setzte die ETH-Bibliothek verschiedene Maßnahmen um, die zum Ziel hatten, den Aufwand zur Bedienung der beiden Systeme auf Kundenseite so gering wie möglich zu halten: So wurde es Benutzenden von E-Citations ermöglicht, im Eingabeformular einen Volltext für E-Collection hochzuladen. Anwender von E-Collection konnten den Nachweis ihrer Publikation in E-Citations ebenfalls bereits im Upload-Formular veranlassen. Die dahinterliegenden Prozesse blieben aber mit erheblichem Koordinationsaufwand für die Mitarbeitenden der Bibliothek verbunden.

## 2.2 Publikation von Forschungsdaten

In den Jahren 2010 bis 2013 entwickelte die ETH-Bibliothek im Rahmen des Projekts „Digitaler Datenerhalt“ eine Dienstleistung für die Langzeitarchivierung von Forschungsdaten, Geschäftsunterlagen und Bibliotheksdokumenten der ETH Zürich. Auf Basis der Software Rosetta von Ex Libris<sup>4</sup> wurde dabei eine technische Infrastruktur aufgebaut, die es den Forschenden der ETH Zürich ermöglichen sollte, ihre Forschungsdaten dauerhaft zu

sichern. Obwohl der Fokus dieses Projekts zunächst auf der Sicherstellung des langfristigen Zugriffs lag, rückten aufgrund sich verändernder Anforderungen zunehmend auch die Themen Publikation und Zitierfähigkeit von Forschungsdaten in das Blickfeld. So wurde schließlich ein Workflow entwickelt, der es ETH-Angehörigen ermöglichte, einzelne oder mehrere Datensätze über das ETH Data Archive (Rosetta) zu publizieren und mithilfe eines Digital Object Identifiers (DOI) deren Zitierfähigkeit sicherzustellen.<sup>5</sup>

## 3 Projektmotivation und Projektziele

Mit den im vorangegangenen Kapitel beschriebenen Dienstleistungen hatte die ETH-Bibliothek ein umfassendes Angebot zur Unterstützung der Forschenden der ETH Zürich bei der Zugänglichmachung, Dokumentation und Archivierung ihres Forschungsausgaben aufgebaut. Je stärker diese Services durch die ETH-Community genutzt wurden, desto deutlicher traten jedoch auch ihre Schwachstellen sowie die Problematik ihrer bisher nur losen Vernetzung zutage.

So bewirkte die Anbindung von E-Citations an die Webseiten der ETH Zürich – als Datenquelle für Publikationslisten – zwar einerseits eine verstärkte Nutzung und Sichtbarkeit der Bibliographie, andererseits akzentuierte sich in diesem Zusammenhang aber auch die Unzufriedenheit vieler Kundinnen und Kunden über die Tatsache, dass Dokumentation und Veröffentlichung ihrer Publikationen nicht in ein und demselben System abgewickelt werden konnten. Auch wurde es zunehmend schwierig, den Kundinnen und Kunden den Unterschied zwischen E-Collection und E-Citations zu vermitteln bzw. den Grund für die Existenz zweier Systeme zu erklären. Gleichzeitig zeigte sich auch bei den ersten Pilotkunden des Projekts „Digitaler Datenerhalt“, dass der Workflow zur Publikation von Forschungsdaten zwar seinem Zweck diene, das User Interface von Rosetta jedoch zentrale Anforderungen an einen kundenfreundlichen, niederschweligen Publikationsprozess nicht erfüllen konnte.

Auf Seiten der Forschungsförderer, allen voran der Europäischen Union, später auch des Schweizer Nationalfonds, war zudem 2013 bereits eine zunehmende Tendenz erkennbar, nicht nur den offenen Zugang zu wissenschaftlichen Publikationen, sondern auch zu Forschungsdaten

<sup>2</sup> Vgl. <http://www.library.ethz.ch/de/Ueber-uns/Projekte/ETH-E-Collection> (30.05.2018).

<sup>3</sup> Vgl. <http://www.library.ethz.ch/de/Ueber-uns/Projekte/Publikationen-Pool> (30.05.2018).

<sup>4</sup> <http://www.exlibrisgroup.com/products/rosetta-digital-asset-management-and-preservation> (30.05.2018).

<sup>5</sup> Vgl. <http://www.library.ethz.ch/de/Ueber-uns/Projekte/Digitaler-Datenerhalt> (30.05.2018).

zu fordern. Es zeichnete sich daher ab, dass Forschungsdaten als eigenständige digitale Objekte in Publikationssystemen an Bedeutung gewinnen würden und die Möglichkeit geschaffen werden musste, diese Objekte über geeignete Metadaten mit Textpublikationen zu verlinken.

Deshalb entschied die ETH-Bibliothek, E-Collection und E-Citations in einer gemeinsamen Publikationsplattform zusammenzuführen und in diese auch einen Workflow für die Publikation bzw. die Archivierung von Forschungsdaten zu integrieren. Ziel war die Entwicklung eines Single-Point-of-Access für die Angehörigen der ETH Zürich, über den sie ihren gesamten Forschungsoutput – seien es klassische Publikationen oder Forschungsdaten – dokumentieren, publizieren und archivieren können.

## 4 Projektorganisation

Für das Management von IT-Projekten ist an der ETH-Bibliothek seit mehreren Jahren ein standardisierter Projektlauf (SPA) etabliert. Das Vorgehensmodell basiert auf der Projektführungsmethode HERMES der Schweizerischen Bundesverwaltung.<sup>6</sup> Ablauf und Organisation des Projekts E-Publications erfolgten entlang der im SPA vorgegebenen Strukturen und Prozesse.

Mit der operativen Umsetzung des Projekts beauftragte der Direktor der ETH-Bibliothek die Projektleitung – bestehend aus Gesamt- und operativer Projektleitung – in Zusammenarbeit mit einem Projektkernteam und einem erweiterten Projektteam. Im Steuerungsausschuss war neben den Sektionsleitungen Kundenservices, Medienservices und IT-Services sowie Innovation und Entwicklung der ETH-Bibliothek auch der Leiter der Abteilung Controlling der ETH Zürich vertreten.

Das relativ kleine Projektkernteam mit sechs Mitarbeitenden aus den Teams E-Publishing, Integration und Entwicklung, Digitaler Datenerhalt sowie Web and Digital Media stellte einen effizienten Arbeitsablauf sicher und ermöglichte in kritischen Phasen unter Einbringung des individuellen fachlichen Knowhows der beteiligten Personen einen gesicherten Projektfortschritt. Gleichzeitig war es über die Organisationsform des erweiterten Projektteams möglich, weitere Personen und Stellen situativ in das Projekt einzubeziehen.

<sup>6</sup> <http://www.hermes.admin.ch> (13.05.2018).

## 5 Anforderungsanalyse

Für die Erstellung des Anforderungskatalogs konnte das Projektteam auf die umfangreichen Erfahrungen der für E-Collection, E-Citations und das Data Archive verantwortlichen Mitarbeitenden zurückgreifen. Zahlreiche Gespräche mit den beteiligten Personen gaben einen guten Einblick in die Stärken und Schwächen der vorhandenen Systeme, sowohl aus Mitarbeitenden- als auch aus Kundensicht. Um die Kundenperspektive nochmals gezielt abzufragen, wurde zudem eine Gruppenbefragung mit Kundinnen und Kunden aus der ETH Zürich durchgeführt. Die aus diesem Prozess hervorgegangenen Erkenntnisse wurden in ein rund achtzigseitiges Pflichtenheft zuzüglich zahlreicher Anhänge überführt. In den folgenden Kapiteln wird ein grober Überblick über die wichtigsten darin spezifizierten Funktionen und die zugrundeliegenden konzeptionellen Überlegungen gegeben.<sup>7</sup>

### 5.1 Rollenkonzept

Das für die Research Collection erarbeitete Rollenkonzept basierte auf Anforderungen, die sich aus dem für die Publikationsplattform vorgesehenen Betriebskonzept ergaben. Für den Qualitätssicherungsprozess (QS) von Publikationen und Forschungsdaten würde demnach ein Kernteam bibliothekarisch ausgebildeter Mitarbeitender verantwortlich sein, das bei der Bearbeitung einfacher „Metadata Only“-Einträge zudem von studentischen Aushilfen (QA) unterstützt würde. Die studentischen Aushilfen sollten dabei über ein kleineres Set von Rechten im System verfügen als die Mitarbeitenden des Kernteams. Eine weitere Besonderheit des Rollenkonzepts ist die Unterscheidung zwischen „normalen“ zum Upload berechtigten Personen (Submitter) und den sogenannten „Submitter Plus“, denen das Recht eingeräumt wird, bestimmte Metadatenfelder,

<sup>7</sup> Der Fokus liegt hier auf den aus fachlich-konzeptioneller Sicht relevantesten Themen. Im Pflichtenheft wurden darüber hinaus die angedachte Systemarchitektur sowie zahlreiche weitere Anforderungen definiert, auf die hier aus Platzgründen nicht näher eingegangen werden kann. Darunter fallen zum Beispiel das User Management, Suchfunktionen und Indexierung, Vorschaufunktionen und Cover Pages, Batch Edit Funktionalitäten, Reports und Downloadstatistiken, datenbank- und objektbezogene Anforderungen, Anforderungen hinsichtlich Datensicherheit, Verfügbarkeit, Performance, Support und Dokumentation, Anforderungen an das User Interface sowie detaillierte Beschreibungen aller elektronischen und manuellen Workflows, u. a. für die Eingabe von Publikationen, Suchen und Browsen, Massenimport, Trefferexport, Duplikatsprüfung und das Embargo Handling.

**Tab. 1:** Rollenkonzept der Research Collection. Die für eine Rolle definierten Rechte gelten jeweils auch für die nächste Berechtigungsstufe, das heißt ein Library Admin verfügt z. B. automatisch auch über alle für QA, Submitter (Plus) und General User definierten Rechte.

Rolle im System	Rolle im Betrieb	Rechte
IT Admin	Applikationsverantwortlicher IT	Kann Änderungen der Systemkonfiguration vornehmen, die Applikation starten und stoppen usw.
Library Admin	Kernteam QS	Kann Publikationen bearbeiten, freischalten und löschen, Reports definieren und erzeugen, Duplikate zusammenführen, Batch Edit Tasks ausführen.
QA	Studentische Aushilfe QS	Kann Publikationen bearbeiten und freischalten.
Submitter Plus	Power User aus der ETH Zürich	Kann Publikationen erfassen, Änderungswünsche melden und ein Subset von Metadaten selbst bearbeiten.
Submitter	User aus der ETH Zürich	Kann Publikationen erfassen und Änderungswünsche melden.
General User	Externer User	Kann suchen, browsen, Treffer exportieren und frei zugängliche Objekte herunterladen.

wie zum Beispiel Tags, Sachgruppen oder Grant-IDs bei vorhandenen Publikationen, selbständig zu bearbeiten bzw. hinzuzufügen. Normale Submitter dagegen müssen Änderungswünsche über ein Formular beantragen.

## 5.2 Qualitätssicherungsprozess

Der Qualitätssicherungsprozess für Publikationen und Forschungsdaten in der Research Collection umfasst u. a. die folgenden Arbeiten:

- Vervollständigung, Korrektur und Vereinheitlichung der Metadaten von Publikationen mittels Recherche der Originalpublikation.
- Anreicherung von „Metadata Only“-Einträgen mit Volltexten, sofern diese beim Verlag unter einer Creative-Commons-Lizenz publiziert wurden.
- Prüfung von Zweitveröffentlichungen hinsichtlich Einhaltung der von den Verlagen vorgegebenen Richtlinien.
- Prüfung von Hochschulschriften hinsichtlich Einhaltung von Vorgaben der entsprechenden ETH-Verordnungen.
- Prüfung von PDFs und gegebenenfalls Forschungsdaten hinsichtlich Lesbarkeit und Schreibschutz.

Abbildung 1 zeigt die Bedeutung und Abfolge der Qualitätsstatus, die ein Eintrag in der Research Collection durchlaufen kann, wobei besonderer Wert auf eine flexible Handhabung des QS-Workflows gelegt wurde. So sollten einzelne Status im QS-Prozess bei Bedarf übersprungen werden können und nicht fertig bearbeitete Publikationen nicht zwingend im Workflow eines einzelnen QA Users pendend bleiben, sondern bei Bedarf auch von anderen Usern weiterbearbeitet werden können. Auch können bereits fertig bearbeitete Publikationen auf einen früheren

Status zurückfallen, zum Beispiel wenn ein Submitter zu einem „Metadata Only“-Eintrag nachträglich einen Volltext hochlädt.

Der Status, den ein Eintrag im Qualitätssicherungsprozess hat, wirkt sich auf dessen Sichtbarkeit aus. Damit die bibliographischen Daten zu neuen Publikationen möglichst bald auf den Webseiten der Forschungsgruppen angezeigt werden können, sind Metadaten zu Publikationen und Forschungsdaten sofort nach dem Import in die Research Collection für Endnutzer sichtbar und stehen via REST API zur Nachnutzung im Web-CMS und in der akademischen Berichterstattung zur Verfügung. Volltexte und andere digitale Objekte sind zu diesem Zeitpunkt jedoch noch nicht nach außen sichtbar. Diese werden erst bei Einträgen mit den Status QS5 und QS6 angezeigt, das heißt nach Prüfung des Volltextes bzw. der Forschungsdaten hinsichtlich Lesbarkeit und Urheberrechten. Dass Publikationen bereits in ungeprüftem Zustand auf den Webseiten der ETH-Angehörigen und im akademischen Bericht angezeigt werden können, federt den zeitlichen Druck auf die Qualitätssicherung insbesondere in der Hochphase der akademischen Berichterstattung ab.

## 5.3 Metadatenschema

Bei der Spezifikation des Metadatenschemas orientierten wir uns an der Metadatenstruktur der Vorgängerapplikation E-Citations, die sich für die Zwecke einer Hochschulbibliographie bewährt hatte. Es handelte sich dabei um eine flache Metadatenstruktur, jedoch mit relativ hoher Granularität auf Ebene der einzelnen Felder, was die Nachnutzung der Daten für die Erzeugung von Publikationslisten auf Webseiten oder in anderen Dokumenten begünstigte. Das Metadatenschema wurde in Anlehnung

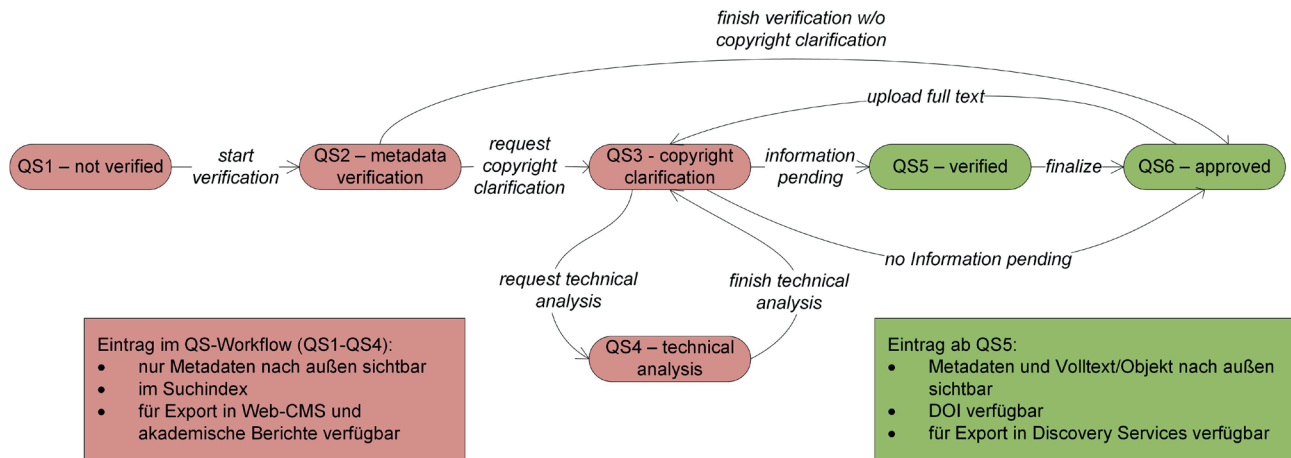


Abb. 1: Qualitätssicherung Statusdiagramm

an das DataCite Metadata Schema<sup>8</sup> um Felder für die Beschreibung von Forschungsdaten erweitert und mit dem Format Dublin Core<sup>9</sup> sowie den OpenAIRE Guidelines<sup>10</sup> abgeglichen, sodass einem Export in diesen standardisierten Datenformaten nichts im Weg stehen würde.

Das vollständige Metadatenschema kann in der Online-Dokumentation der Research Collection eingesehen werden.<sup>11</sup> Bei Feldern, in denen ein kontrolliertes Vokabular zum Einsatz kam, erfolgte dessen Spezifikation unter Rückgriff auf die OpenAIRE Guidelines, das DataCite Metadata Schema, und Elemente der CASRAI Output Types<sup>12</sup> bzw. der COAR Controlled Vocabularies.<sup>13</sup> Zudem wurde für mehrere Felder die Nutzung interner oder extern angebundener Normdateien bzw. Klassifikationen vorgesehen, darunter für Autoren und Autorinnen sowie andere Beitragende, Zeitschriften, Organisationseinheiten, GrantIDs und Sachgruppen.

## 5.4 Zugriffsrechte und DOIs

Während das Vorgängersystem E-Collection ausschließlich als Open-Access-Plattform diente und kein abgestuftes Zugriffsrechtekonzept kannte, wurde für die Research Collection ein flexibles Zugriffsrechtekonzept erarbeitet, das insbesondere die diesbezüglichen Bedürfnisse und

Anforderungen in Zusammenhang mit der Publikation von Forschungsdaten in den Blick nahm. Tabelle 2 gibt einen Überblick über die spezifizierten Zugriffsrechte. Die Zugriffsrechte beziehen sich hier immer auf den Volltext bzw. das digitale Objekt, während die Metadaten zu Publikationen bzw. Forschungsdaten in der Research Collection stets sichtbar sind. Die gewählten Zugriffsrechte wirken sich auch darauf aus, ob für einen Eintrag in der Research Collection ein DOI registriert wird oder nicht.

Da die Research Collection in Bezug auf die Veröffentlichung von Textpublikationen wie ihr Vorgängersystem den Charakter eines Open-Access-Repository behalten sollte, wurde entschieden, den Endkunden die ganze Bandbreite der o. g. Zugriffsvarianten nur für die Hinterlegung von Forschungsdaten anzubieten. Für die Hinterlegung von Publikationen und Hochschulschriften hingegen stehen dem Submitter im Sinn der Open-Access-Policy der ETH Zürich, ausschließlich die Varianten Open Access und Embargoed zur Verfügung.

## 5.5 Aufbewahrungsfristen

Bei der Publikation und Archivierung von Forschungsdaten besteht eine besonders große Herausforderung darin, dass diese von den Datenproduzenten oft in Dateiformaten bereitgestellt werden, die für die Langzeitarchivierung ungeeignet sind. Die Transformation in ein für die Langzeitarchivierung taugliches Format wäre meist ein zeitaufwändiges und technisch komplexes Unterfangen, das für die Datenproduzenten selbst zudem oft eine geringe Priorität gegenüber der zeitnahen Bereitstellung ihres Datensatzes für die Nutzung durch Dritte hat. Um diesem Dilemma zu begegnen, hat sich die ETH-Biblio-

<sup>8</sup> <http://schema.datacite.org/> (30.05.2018).

<sup>9</sup> <http://dublincore.org/documents/dces/> (30.05.2018).

<sup>10</sup> <https://guidelines.openaire.eu/en/latest/literature/> (30.05.2018).

<sup>11</sup> <https://documentation.library.ethz.ch/display/RC/Metadaten-schema> (30.05.2018).

<sup>12</sup> [http://dictionary.casrai.org/Output\\_Types](http://dictionary.casrai.org/Output_Types) (30.05.2018).

<sup>13</sup> <https://www.coar-repositories.org/activities/repository-interoperability/coar-vocabularies/> (30.05.2018).

Tab. 2: Zugriffsrechte- und DOI-Konzept

Zugriffsrecht	Bedeutung	Eintrag erhält DOI
Open Access	Frei zugänglich	Ja
Embargoed	Nach Ablauf einer Sperrfrist frei zugänglich	Ja
ETHZ Users	Nur für eingeloggte ETH-Angehörige zugänglich	Ja
Selected Users	Nur für ausgewählte eingeloggte Nutzer zugänglich	Nein
Closed Access	Nur für den Submitter und QS-Personal zugänglich	Nein

thek entschieden, in Schulungen und Beratungsgesprächen darauf hinzuwirken, dass die ETH-Angehörigen ihre Daten in archivierungstauglichen Dateiformaten bereitstellen. In der Praxis sollten jedoch grundsätzlich alle Dateiformate für den Upload in die Research Collection zugelassen werden. Gleichzeitig sollte den Kunden die Möglichkeit geboten werden, für ihre Daten eine befristete Aufbewahrungsdauer von 10 oder 15 Jahren zu wählen. Dies wird besonders dann empfohlen, wenn die Daten in proprietären oder anderen nicht archivierungstauglichen Formaten vorliegen. Es gibt den Produzenten die Möglichkeit, entsprechend der guten wissenschaftlichen Praxis<sup>14</sup> ihre Daten für einen befristeten Zeitraum sicher aufzubewahren und gegebenenfalls zu publizieren, ohne dass eine unbefristete Aufbewahrung oder Erhaltung der Lesbarkeit durch die ETH-Bibliothek vorausgesetzt wird.

## 5.6 Import- und Exportfunktionen

Um eine möglichst optimale Einbettung der Publikationsplattform in die Systemlandschaft der ETH Zürich zu erreichen, sowie die Prozesse zum Import neuer ETH-Publikationen weitestgehend zu automatisieren, wurden zahlreiche Schnittstellen zu Drittsystemen, sowie Import- und Exportfunktionalitäten vorgesehen, die in das Repository integriert werden sollten. Bei einem Teil davon konnte auf die Erfahrungen mit den Vorgängersystemen zurückgegriffen werden, andere, wie zum Beispiel die Schnittstelle zum Grants Management der ETH, wurden für die Research Collection neu konzipiert. Tabelle 3 gibt einen Überblick über die spezifizierten Schnittstellenmodule.

<sup>14</sup> Vgl. ETH Zürich. *Richtlinien für Integrität in der Forschung*. Zürich: ETH Zürich, 2011. doi:10.3929/ethz-b-000179298.

## 6 Marktanalyse und Ausschreibungsverfahren

Bereits in der konzeptionellen Vorphase zum Projekt hatten erste Analysen des Projektteams sehr schnell gezeigt, dass die Zusammenführung von E-Collection und E-Citations auf Basis eines der bestehenden Systeme (das heißt z. B. durch Integration von E-Collection in E-Citations oder umgekehrt) nicht realisierbar wäre. Grund dafür waren unter anderem die umfangreichen ETH-spezifischen Erweiterungen, die an beiden Systemen vorgenommen worden waren. Auch eine neuerliche Eigenentwicklung oder die Inhouse-Adaptierung eines Open-Source-Produkts wurden als zu risikoreich und nicht tragfähig für die Umsetzung des Vorhabens bewertet.

Aus diesen Gründen entschied sich die ETH-Bibliothek, das Projekt in Zusammenarbeit mit einem externen Dienstleister durchzuführen, der im Rahmen eines öffentlichen Ausschreibungsverfahrens gefunden werden sollte. Auch dieses Vorgehen war mit einem gewissen Risiko verbunden, denn im Unterschied zum Markt der klassischen Bibliothekssysteme mit seinen etablierten kommerziellen Anbietern war der Markt der Repositorien-Software zum Zeitpunkt des Projektstarts von Open-Source-Lösungen dominiert, für die sich bislang nur wenige Dienstleister für Weiterentwicklungen, Integrationsarbeiten sowie Support und Maintenance etabliert hatten. Zudem konnte auf den ersten Blick keine der im Rahmen der Marktanalyse gefundenen Lösungen den Anspruch an ein System, das sowohl als Hochschulbibliographie, als auch als Open-Access-Repository und als Forschungsdaten-Repository dienen sollte, out-of-the-box erfüllen. Tabelle 4 gibt einen Überblick über die zu Projektbeginn Ende 2014 vorgefundene Marktsituation.

Im April 2015 wurde die Ausschreibung für die neue Publikationsplattform der ETH Zürich auf der Schweizerischen Plattform für das öffentliche Beschaffungswesen [simap.ch](https://www.simap.ch)<sup>15</sup> publiziert. Die eingegangenen Offerten bewertete das Projektteam anhand der in der Ausschreibung

<sup>15</sup> <https://www.simap.ch> (30.05.2018).

**Tab. 3:** Übersicht der spezifizierten Schnittstellenmodule

Modul	Zweck/Beschreibung	API/Protokoll
Citation Import	Benutzergesteuerter Massenimport von Literaturlisten im BibTex- und RIS-Format	n/a
Citation Harvesting	Automatischer Import neuer ETH-Publikationen aus Web of Science und Scopus	WoS Web Services Expanded Scopus API
Metadata Retrieval	Metadatenabfrage via DOI oder ORCID ID zum Vor-Ausfüllen des Eingabeformulars	CrossCite API ORCID API
Copyright Policies	Anzeige der Copyright-Richtlinien von Zeitschriften im Eingabeformular	Sherpa/Romeo API
Authentication and Authorization	Single-Sign-On für ETH-Angehörige	Shibboleth (SWITCH AAI)
Citation Export	Export von Trefferlisten im CSV-, BibTex- und RIS-Format	n/a
RSS Channel	Abonnieren von Suchergebnissen als RSS-Feed	RSS 2.0
OAI Harvesting Service	OAI-PMH-Schnittstelle zur Integration von Research Collection Publikationen in Suchmaschinen	OAI-PMH
Preservation Interface	Übermittlung von Metadaten und digitalen Objekten an das Langzeitarchiv (ETH Data Archive/Rosetta)	Rosetta-spezifischer Ingest-Workflow
Grants Management	Import von Daten zu EU- und SNF-geförderten Projekten zur Verknüpfung mit Publikationsdaten (Quelle: SAP)	REST-basierte Schnittstelle
Organizational Structure	Import der Organisationseinheiten der ETH Zürich und deren hierarchischer Beziehungen zur Verknüpfung mit Publikationsdaten und für Browsen via Organisationsstruktur (Quelle: Organisationsdatenbank der ETH Zürich)	SOAP-basierte Schnittstelle
Person Directory	Import der Personenamen und ORCID IDs von ETH-Angehörigen aus der zentralen ETH-Personendatenbank zur Integration in die Autorennormdatei	LDAP
Leitzahl Certification	Abfrage der „Leitzahlberechtigung“ eines Users bei Anmeldung und Hinterlegen der Berechtigung im User Profil. Diese Berechtigung wird im SAP-Modul für die akademische Berichterstattung vergeben und regelt, wer Publikationen einer Organisationseinheit für die Anzeige in deren akademischen Jahresbericht zuordnen darf	SOAP-basierte Schnittstelle
REST	Schnittstelle für den Zugriff auf und die Abfrage von Publikationsdaten; primäre Anwendungsfälle: Einbindung von Publikationslisten in Webseiten der ETH Zürich sowie in die akademische Berichterstattung	REST

**Tab. 4:** Marktanalyse Repository Software

Produkt	Proprietär/Open Source	Anbieter/Dienstleister
<b>Open Access Repository Software</b>		
EPrints	Open Source	EPrints Services
DSpace	Open Source	Atmire, Longsight (seit 2017 Teil von Atmire), Enovation Solutions
Opus	Open Source	KOBV (nur Deutschland)
Fedora	Open Source	Mediashef
Invenio	Open Source	TIND Technologies
Visual Library	Proprietär	Walter Nagel
<b>Data Repository Software</b>		
CKAN	Open Source	LIIP
DataVerse	Open Source	n/a
Figshare Institutional Edition	Proprietär	Figshare (Digital Science)
<b>Forschungsinformationssysteme</b>		
Pure	Proprietär	Elsevier
Converis	Proprietär	Thomson Reuters (heute Clarivate Analytics)
Elements	Proprietär	Symplectic (Digital Science)
VIVO	Open Source	n/a



definierten Kriterien. Aus diesem Evaluationsverfahren ging eine Offerte der belgischen Firma Atmire,<sup>16</sup> die auf der Open-Source-Software DSpace<sup>17</sup> basierte, als wirtschaftlichstes Angebot hervor.

## 7 Implementierung der Softwarelösung

Nach dem Abschluss des Evaluationsverfahrens sowie der Vertragsverhandlungen mit dem Dienstleister Atmire startete das Projekt Anfang 2016 in die Umsetzungsphase. Für die Implementierung der Softwarelösung gingen wir in enger Zusammenarbeit mit Atmire in einem Phasenmodell vor. Hierbei wurde jede im Pflichtenheft spezifizierte Anforderung einer von fünf Entwicklungsphasen zugeteilt. Anschließend wurden für jede Phase (a) die Anforderungen nochmals diskutiert, analysiert und in einer Detailspezifikation dokumentiert, (b) die besprochenen Funktionalitäten von Atmire entwickelt und auf dem Testsystem umgesetzt sowie (c) diese vom Projektteam getestet. Die Anforderungsanalysen wurden mittels wöchentlicher Telefonkonferenzen durchgeführt. Allfällige Issues meldet das Projektteam über Testprotokolle an Atmire zurück.

Zusätzlich zum hier beschriebenen Release Testing durch das Projektteam wurden in einer späteren Phase der Umsetzungsarbeiten auch ausgewählte Mitarbeitende der ETH-Bibliothek, Kunden und Kundinnen sowie Vertreterinnen und Vertreter der nachnutzenden Systeme zum Testen der Applikation eingeladen. Des Weiteren führten Experten des Usability-Lab der HTW Chur<sup>18</sup> Tests mit dem Fokus auf der Nutzbarkeit des User Interfaces durch. Die Ergebnisse dieser beiden Testrunden flossen anschließend wo möglich und angezeigt noch in die letzte Entwicklungsphase ein. Weniger dringende Anregungen aus diesen Tests wurden in einer Ideensammlung für zukünftige Releases der Applikation dokumentiert.

## 8 Datenmigration und Switch to Production

Die Arbeiten zur Vorbereitung der Datenmigration aus den Vorgängersystemen in die Research Collection beschäftigten das Projektteam während der gesamten Projektlaufzeit. Bereits in einer frühen Phase analysierten wir die zu migrierenden Daten aus E-Citations und E-Collection und erstellten ein Mapping der Metadaten der Vorgängersysteme auf das Schema der Research Collection. In der Umsetzungsphase wurde dann das eigentliche Vorgehensmodell für die Datenmigration erarbeitet. Eine Herausforderung bestand darin, dass die Daten aus zwei Vorgängersystemen mit unterschiedlichen Metadatenschemata zusammengeführt und dedupliziert werden mussten, eine weitere darin, dass Metadaten, Volltexte und Access Logs der E-Collection aus drei verschiedenen Quellsystemen bezogen werden mussten (vgl. Abbildung 2).

Um die Komplexität des Migrationsprozesses zu reduzieren, wurde eine Zwischendatenbank erstellt, in welche die Daten aus den Vorgängersystemen zunächst als separate Tabellen importiert wurden. Die hierfür notwendigen Migrationsskripte wurden in einem iterativen Verfahren entwickelt und getestet, bis sie zur Ausführung im Rahmen der finalen Datenmigration bereit waren. Im nächsten Migrationsschritt, dem Import der Daten aus der Zwischendatenbank nach DSpace, wurden die Daten zusammengeführt und dedupliziert sowie auf ein einheitliches Schema gemappt. Über die Nutzung einer Redirect-Applikation wurde sichergestellt, dass der Aufruf einer Publikation über eine „alte“ URL auch nach der Inbetriebnahme der Research Collection noch zum korrekten Eintrag führen würde. Die Logfiles zur Anzeige der Downloadstatistiken einzelner Items lagerten ebenfalls in der Zwischendatenbank, wurden aber erst in einer späteren Phase, nach dem Switch-to-Production, nach DSpace importiert.

Die eigentliche Aufschaltung der Plattform und die zeitgleich stattfindende Datenmigration erfolgten schrittweise: Zunächst wurden E-Collection und E-Citations in einen Read-Only-Modus überführt. Danach wurden die darin enthaltenen Daten bestmöglich für die anstehende Migration vorbereitet. Parallel schalteten wir die Research Collection bereits im Rahmen eines „Soft Launchs“ ohne Inhalte frei und öffneten sie zunächst nur für die Abgabe von Dissertationen. Anschließend erfolgte die finale Migration der ca. 170 000 Items in die Zwischendatenbank und von dort in die Research Collection. Nach Abschluss des Migrationsimports, der sich über mehrere Tage erstreckte, wurden E-Collection und E-Citations vom Netz genommen und alle URLs auf die neue Plattform umgeleitet. Anschlie-

<sup>16</sup> <https://www.atmire.com/> (30.05.2018).

<sup>17</sup> <http://duraspace.org/dspace/> (30.05.2018).

<sup>18</sup> <http://www.cheval-lab.ch/> (30.05.2018).

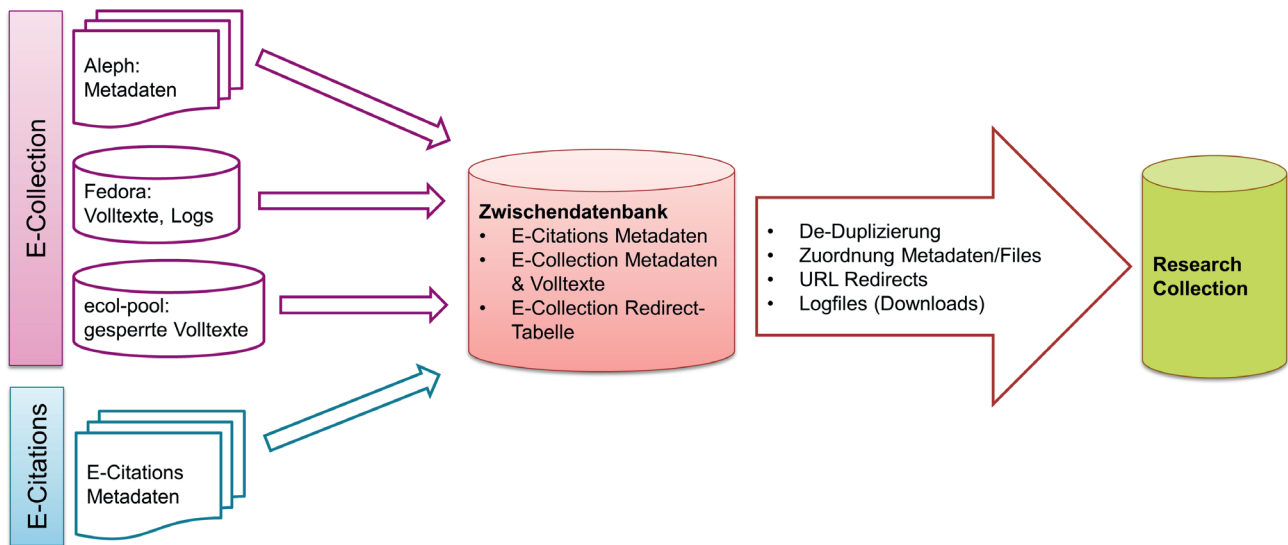


Abb. 2: Schematische Darstellung der Prozesse zur Datenmigration

End öffneten wir die Research Collection für die Eingabe neuer Publikationen und Forschungsdaten. Wenige Tage später wurden die ETH-Angehörigen in E-Mails und auf diversen anderen Kanälen über die neue Dienstleistung informiert.

## 9 Schnittstellenmanagement

Um eine hochschulweite Publikationsplattform wie die Research Collection erfolgreich in Betrieb zu nehmen, ist es von zentraler Bedeutung, dieses Vorhaben nicht als isoliertes Entwicklungsprojekt der Bibliothek zu betrachten, sondern die Plattform als integrales Element der Informationsinfrastruktur der Hochschule wahrzunehmen. Die Verantwortlichen für die nachnutzenden Systeme innerhalb der Hochschule müssen zu jedem Zeitpunkt über den Projektfortschritt informiert sein, damit sie die notwendigen Anpassungen an ihren Applikationen mit genügend Vorlauf planen und ihre Nutzer darüber informieren können. Es ist für den Projekterfolg zentral, ihre Bedenken aufzunehmen und zu adressieren, zugleich gilt es umgekehrt aber auch darauf hinzuwirken, dass Anforderungen der Drittsysteme an die neue Plattform klar und mit genügend zeitlichem Vorlauf dokumentiert und dem Projektteam kommuniziert werden.

Im Fall der ETH Zürich betraf dies insbesondere das zentrale Web-CMS, das den ETH-Angehörigen die Möglichkeit bietet, Publikationslisten auf Basis der Daten der Research Collection (davor E-Citations) in ihre Webseiten einzubinden. Hinzu kamen die Verantwortlichen für die akademische Berichterstattung, die ebenfalls eng in das

Projekt eingebunden waren. Des Weiteren mussten auch die Verantwortlichen für die betrieblich weniger kritischen, aber dennoch für die Außenwirkung zentralen Schnittstellen, wie zum Beispiel das Discovery-System der Bibliothek, eingebunden werden (vgl. Abbildung 3).

## 10 Betriebsüberführung

Durch die geplante Zusammenführung dreier an der ETH-Bibliothek bislang größtenteils separat und in verschiedenen Teams betriebener Prozesse in einer Plattform entstand durch das Projekt automatisch auch Anpassungsbedarf an den betrieblichen Abläufen. In Zusammenarbeit mit den Verantwortlichen für den inhaltlichen Betrieb arbeitete das Projektteam parallel zu den anderen Arbeitspaketen darauf hin, dass zum Zeitpunkt der Inbetriebnahme der Research Collection alle bisher mit dem Betrieb von E-Collection und E-Citations betrauten Mitarbeitenden als Research Collection „Kernteam“ in die Gruppe E-Publishing innerhalb der Sektion Kundenservices (heute: Forschungsdienstleistungen) der ETH-Bibliothek gewechselt hatten, sofern dies nicht bereits der Fall war. Dadurch konnte mit dem Launch der Publikationsplattform auf ein organisatorisch wie räumlich vereintes Team an Mitarbeitenden zurückgegriffen werden, das den fachlichen Betrieb der Research Collection sicherstellen würde. Um den Wissenstransfer innerhalb dieses zukünftigen Research-Collection-Kernteam frühstmöglich anzustoßen, wurde bereits im Herbst 2016 mit der Schulung der beteiligten Mitarbeitenden begonnen. Die Schulungen dieser Phase hatten zum Ziel, einen ersten

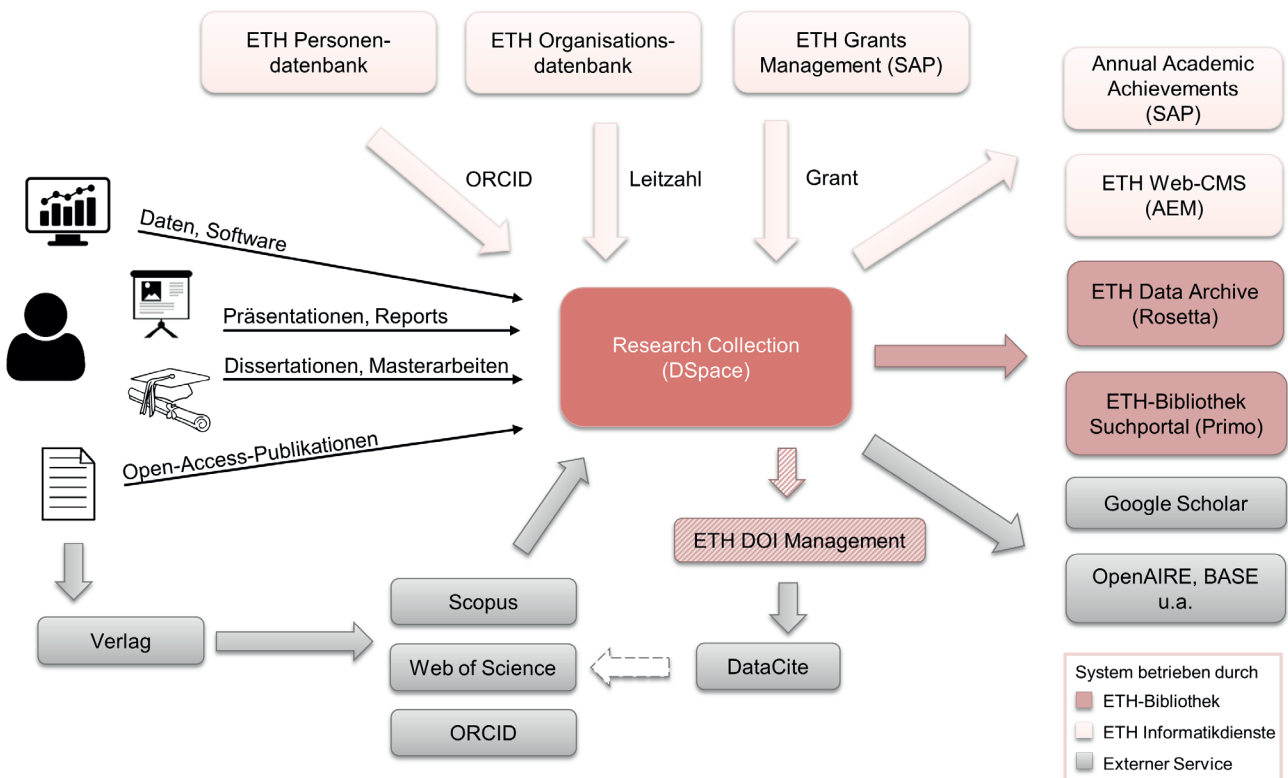


Abb. 3: Schnittstellenübersicht

Wissensaustausch und Kompetenztransfer zwischen den Expertinnen und Experten für bibliographische Qualitätssicherung, Open Access und Digitalen Datenerhalt zu ermöglichen. Eine weitere Schulungsphase im Frühjahr 2017 thematisierte dann die konkreten Arbeitsabläufe für die Bearbeitung von Publikationen und Forschungsdaten in DSpace und beinhaltete praktische Übungen im Testsystem.

Natürlich lag es auf der Hand, dass neben den in einer ersten Phase des Echtbetriebs üblicherweise auftretenden „Kinderkrankheiten“ und den damit verbundenen technischen Anpassungsarbeiten auch die betrieblichen Abläufe sich erst einige Zeit nach Projektabschluss endgültig einspielen würden. Im ersten Jahr nach der Betriebseinführung galt es für das Team zunächst vor allem, die zahlreichen neuen Prozesse für den Betrieb der Plattform im Team umzusetzen, zu erproben, zu verbessern und zu dokumentieren. Insbesondere im Bereich der Bearbeitung von Forschungsdaten, wo das Team auf die wenigsten Vorerfahrungen zurückgreifen konnte, war das erste Jahr mit einem stetigen Lernprozess verbunden: Von der Frage, was die Qualitätssicherung bei Forschungsdaten überhaupt leisten kann, bis zu Details bezüglich empfohlener Dateiformate und Dateigrößen, wurden wir mit zahlreichen herausfordernden Aufgaben konfrontiert.

## 11 Fazit und Ausblick

Im Projekt E-Publications hat die ETH-Bibliothek eine zukunftsweisende Entscheidung getroffen, indem sie sich für die Entwicklung einer „3-in-1“-Lösung für die Hochschule entschieden hat, die die Funktionen einer Hochschulbibliographie, eines Open-Access-Repository und eines Forschungsdaten-Repository integriert; dies zu einem Zeitpunkt, als es für ein solches Vorhaben weltweit de facto noch keine Referenzprojekte gab. Rückblickend ist nicht von der Hand zu weisen, dass der Aufwand für die Umsetzung dieser Vision insbesondere in Kombination mit dem ebenfalls zu bewältigenden Migrationsvorhaben beträchtlich war und zu einem außerordentlichen Komplexitätsgrad im Projekt geführt hat.

Die zahlreichen positiven Kundenfeedbacks seit dem Launch der Plattform bestärken uns jedoch darin, dass wir mit der Research Collection den richtigen Weg für die ETH Zürich eingeschlagen haben. Dass der von der ETH-Bibliothek vorgezeigte Weg auch für andere Hochschulen ein mögliches Zukunftsszenario sein kann, bestätigen aktuell zwei Meldungen aus der Branche der Bibliotheksdienstleister, in denen sowohl Figshare als auch Ex Libris Entwicklungsvorhaben ankündigen, bei denen

eine gewisse Ähnlichkeit mit den Funktionalitäten der Research Collection ins Auge sticht.<sup>19</sup>

Unabhängig davon, wie sich die Marktsituation in Zukunft entwickelt: Die Research Collection ist für die ETH Zürich ein Zukunftsmodell. Als Datenlieferantin für die Webseiten der Institute und Professuren, ebenso wie für die akademische Berichterstattung ist sie zu einem unverzichtbaren Baustein in der hochschulweiten Informationsinfrastruktur geworden. In ihrer Funktion als Hochschulbibliographie und Open-Access-Repository ermöglicht sie ein gezieltes Monitoring des Publikationsverhaltens der ETH-Forschenden, insbesondere im Hinblick auf Open Access. Und als Forschungsdatenrepositorium ist sie ein niederschwelliges Tool für Wissenschaftler, um die Anforderungen der Forschungsförderer im Hinblick auf den offenen Zugang zu Forschungsdaten zu erfüllen.

Ein Blick auf die Roadmap für die weitere Produktentwicklung zeigt, dass es auch nach Projektabschluss zahlreiche Wünsche, Ideen und Verbesserungsvorschläge für die Research Collection umzusetzen gibt. Im nächsten Jahr werden wir uns unter anderem mit der Automatisierung des Ingests von Open-Access-Artikeln aus den Schweizer Nationallizenzen beschäftigen. Im Bereich der Forschungsdaten ist die Entwicklung einer Schnittstelle zu openBIS,<sup>20</sup> einem an der ETH Zürich entwickelten Tool für das aktive Datenmanagement, geplant, ebenso wie die Erarbeitung einer Lösung für den Umgang mit sehr großen Dateien.

## Autoreninformationen



**Barbara Hirschmann**

ETH Zürich

ETH-Bibliothek

Rämistrasse 101

CH-8092 Zürich

[barbara.hirschmann@library.ethz.ch](mailto:barbara.hirschmann@library.ethz.ch)

[orcid.org/0000-0003-0289-0345](https://orcid.org/0000-0003-0289-0345)

<sup>19</sup> So kündigte Figshare Anfang 2018 die Weiterentwicklung ihres Angebots „Figshare for Institutions“ zu einem „all-in-one next generation repository“ an und expandiert damit in den Markt der Open-Access-Repositoryen, vgl. [https://figshare.com/blog/Announcing\\_the\\_figshare\\_Institutional\\_Repository\\_and\\_Data\\_Repository\\_and\\_Thesis\\_Repository\\_really\\_just\\_an\\_all-in-one\\_next\\_gen\\_repository/389](https://figshare.com/blog/Announcing_the_figshare_Institutional_Repository_and_Data_Repository_and_Thesis_Repository_really_just_an_all-in-one_next_gen_repository/389) (30.05.2018). Etwa zeitgleich machte auch eine Meldung von Ex Libris die Runde, wonach das Unternehmen in einer Entwicklungspartnerschaft mit fünf Universitäten an einer neuen „Research Services Platform“ arbeitet, die ebenfalls ähnliche Funktionalitäten wie die Research Collection der ETH Zürich anbieten soll, vgl. <http://markets.businessinsider.com/news/stocks/five-partners-join-ex-libris-in-developing-new-research-services-platform-1015123216> (30.05.2018).

<sup>20</sup> <https://openbis.elnlims.ch/> (30.05.2018).