

# Nachhaltigkeit in langjährigen Erschliessungsprojekten

## FAIR-Data-Kriterien bei Editions- und Forschungsplattformen zum 18. Jahrhundert\*

Peter Dängeli / Martin Stuber

### Abstract

In this article, the FAIR data criteria for digital publications are being tested on the *haller-Net* journal and research platform – alongside a number of other platforms concerned with eighteenth century research. The FAIR approach will prove beneficial beyond simple data management as its fundamental goals can be pursued across other digital research platforms. Given the long history of database-led research, *hallerNet* also lends itself to wider reflections on the sustainability of digital research data.

Keywords: sustainable research data; FAIR criteria; scholarly digital edition; digital platforms; data management plan (DMP); Albrecht von Haller

### Einleitung

Das historisch-kritische Editions Wesen erlebte in den letzten Jahrzehnten eine grundlegende Transformation.<sup>1</sup> In der gedruckten Edition ging es um die Herausgabe eines klar umrissenen, abgeschlossenen Produkts, dessen Rezeptions- und Verarbeitungsweise durch das Publikum weitgehend vorhersehbar war. Die institutionalisierte Aufbewahrung durch Bibliotheken und Archive verstärkte diese Stabilität und schuf zugleich durch Kataloge, Findmittel und in jüngerer

\* Der Beitrag entstand im Rahmen des vom Schweizerischen Nationalfonds finanzierten Projekts *Online-Edition der Rezensionen und Briefe Albrecht von Hallers: Expertise und Kommunikation in der entstehenden Scientific community* (2018–2023); Gesuchstellende: Christophe v. Werdt (Albrecht von Haller-Stiftung der Burgergemeinde Bern), André Holenstein (Historisches Institut der Universität Bern), Hubert Steinke (Institut für Medizingeschichte der Universität Bern), Oliver Lubrich (Germanistisches Institut der Universität Bern), Claudia Engler (Burgerbibliothek Bern).

Wir danken Sarah Baumgartner für wesentliche Anregungen, insbesondere aus ihrer schriftlichen Hausarbeit im Studiengang MAS-ALIS: *Erschliessung historischer Quellen aus der Eidgenossenschaft des 18. Jahrhunderts durch digitale Editions- und Forschungsplattformen*, 2019.

<sup>1</sup> In der jüngeren Forschungsliteratur ist zumeist die Rede von *Scholarly Digital Editions* (SDE). Zu grundlegenden Aspekten des digitalen Publizierens siehe: Paul Eggert, «The Archival Impulse and the Editorial Impulse», *Variants* 14, 2019, S. 3–22; Elena Pierazzo, *Digital Scholarly Editing. Theories, Models and Methods*, Farnham, Ashgate Publishing LTD, 2015; Patrick Sahle, *Digitale Editionsformen. Zum Umgang mit der Überlieferung unter den Bedingungen des Medienwandels*, 3 Bde., Norderstedt, Books on Demand, 2013; Ders., «What is a Scholarly Digital Edition?», in: Matthew Driscoll, Elena Pierazzo (Hg.), *Digital Scholarly Editing. Theories and Practices*, Cambridge, Open Book Publishers, 2016, S. 19–40.

Zeit Datenbanksysteme auch zuverlässige Wege, edierte Quellen zu finden und zu konsultieren. Bei der Digitalen Edition handelt es sich dagegen um einen Publikationsvorgang in einem fluiden und offenen Kontext, der nachträglich Korrekturen oder die Integration zusätzlicher Inhalte ermöglicht. Zudem stellen die Rezipienten neue Ansprüche, etwa jenen eines unmittelbaren Zugriffs auf die publizierten Daten zwecks Einbindung und Verarbeitung in anderen Ressourcen, mitunter in für den Produzenten kaum vorhersehbarer Weise. Der Verlust jener stabilen Festsetzung, die der gedruckten Veröffentlichung eigen war, kann die Verwendung digitaler Publikationsformen zu Forschungszwecken problematisch machen, denn zur Erfüllung des zentralen wissenschaftlichen Qualitätskriteriums der Nachvollziehbarkeit ist die zuverlässige Wiederauffindbarkeit zitierter Befunde und Stellen unabdingbar.

Was mit digitaler Publikation gemeint ist, kann eine grosse Bandbreite abdecken. Am einen Ende des Spektrums steht die digitale Version einer Druckpublikation, die über den gängigen Verlagsweg zustande kommt und alle etablierten Konventionen berücksichtigt, nebst einigen zusätzlichen, durch das digitale Format ermöglichten Funktionalitäten (man denke etwa an eine PDF-Datei mit Titlei, Inhaltsverzeichnis, Paginierung, Apparaten, verlinktem Register und Volltext-Durchsuchbarkeit). Am anderen Ende des Spektrums befindet sich beispielsweise ein Korpus nach bestimmten Massgaben codierter, inhaltsreicher und stark intern und extern verknüpfter Einzeldateien, verstreut über mehrere Forschungsdatenrepositorien. Dazwischen angesiedelt sind Digitale Editionen, die typischerweise eine Datengrundlage umfassen, diese aber in Präsentationsoberflächen mit spezifischen Funktionalitäten vermitteln und, zumindest in den fortschrittlichsten Umsetzungen, auch Möglichkeiten der Versionierung und stabilen Zitierung anbieten. Eine gewisse Bandbreite zeigt sich auch in der Bezeichnung solcher Publikationsformen. Neben «Digitalen Editionen», die Quellen erschliessen, ist die Rede von «Plattformen», zumeist charakterisiert durch zusätzliche Funktionalitäten, «Portalen», die als Einstieg in unterschiedliche Inhalte oder Informationsressourcen dienen, und schliesslich auch von «Datenbanken», die ganz unterschiedlich geartet sein können, immer aber Zugriff auf bestimmte Forschungsdaten geben.

Diesem Variantenreichtum des digitalen Edierens begegnet man auch in einem aktuellen Überblick der *Schweizerischen Gesellschaft für die Erforschung des 18. Jahrhunderts* (SGEAJ). Um die bestehenden und in Entwicklung begriffenen digitalen Editions- und Forschungsprojekte zum langen 18. Jahrhundert besser sichtbar zu machen, veröffentlichte die SGEAJ eine Zusammenstellung von insgesamt achtzehn einschlägigen Plattformen und Datenbanken.<sup>2</sup> Die Sammlung ist betreffend der erschlossenen Quellengattungen und der Tiefe der Aufbereitung sehr heterogen. Zu finden sind sowohl Manuskripte wie auch

---

2 <https://www.sgeaj.ch/dh> (05.02.2020).

gedruckte Werke, die teils als Transkripte, teils aber als gescannte Handschriften, und in einigen Fällen auch in beiden Formen verfügbar gemacht werden. Die (vermutlich) älteste Plattform, *Lumières.Lausanne*<sup>3</sup>, wurde bereits 2008 aufgeschaltet, die jüngsten, *Helvetische Strassenenquête*<sup>4</sup> und *hallerNet*<sup>5</sup>, erst 2019. Während einige die Resultate von weitgehend abgeschlossenen Projekten präsentieren, sind andere explizit darauf angelegt, ständig erweitert zu werden. Die Zusammenstellung der SGEAJ führen wir im Folgenden in ausgewählten Einzelaspekten an die FAIR-Data-Kriterien heran. Ein Schwerpunkt liegt dabei auf den Erfahrungen mit der Plattform *hallerNet*, an der die beiden Verfasser dieses Beitrags mitwirken.

## FAIR-Data-Kriterien für nachhaltige Forschungsdaten

Seit 2016 besteht mit den FAIR-Data-Kriterien ein einprägsamer und allgemein anerkannter Ansatz zum nachhaltigen Umgang mit digitalen Forschungsdaten. Die Kriterien fassen auf dem Vorschlag einer Arbeitsgruppe von mehr als fünfzig ExpertInnen aus der (naturwissenschaftlichen) Forschung, Datenzentren, Universitätsbibliotheken und Verlagen. Der Fokus liegt auf Forschungsdaten i. e. S., also der Publikation von Beobachtungs-, Messdaten und dergleichen.<sup>6</sup> Ein Transfer der den Kriterien zugrundeliegenden Überlegungen auf geisteswissenschaftliche digitale Daten und Ressourcen ist aber durchaus möglich und sinnvoll, weil die den FAIR-Ansatz konstituierenden Ansprüche, die sich direkt aus dem Akronym lesen lassen, auch für geisteswissenschaftliche Forschung zentral sind: Forschungsdaten sollen *findable*, *accessible*, *interoperable* und *reusable* sein. Zur konkreten Evaluierung wurden fünfzehn Unterkriterien formuliert, die jeweils einem der vier Zielbereiche zugeordnet sind. Sie überschneiden sich teilweise und sind daher stets im Hinblick auf den übergeordneten Aspekt, den sie zu klären helfen sollen, zu verstehen. Im Vordergrund des FAIR-Ansatzes steht letztlich nicht die Einhaltung der einzelnen Unterkriterien, sondern die regelgeleitete Gewährleistung der Wiederverwendbarkeit von Daten.

Seit ihrer Veröffentlichung fanden die FAIR-Data-Kriterien rasch viel Akzeptanz und Eingang in die Förderbedingungen nationaler und internationaler Forschungsförderer. So beschloss die *Europäische Kommission* im April 2016

3 <http://lumieres.unil.ch> (05.02.2020).

4 <http://www.strassenenquete.ch/stapfer-enquete> (05.02.2020).

5 <https://hallernet.org> (05.02.2020).

6 Die Arbeitsgruppe *FAIR data publishing group* (2014–2016) war Teil der fortbestehenden Initiative FORCE11 (The Future of Research Communications and e-Scholarship). Die massgebliche Veröffentlichung der Kriterien erfolgte in Mark D. Wilkinson et al., «The FAIR Guiding Principles for scientific data management and stewardship», *Nature: Sci Data* 3, 2016, unpag., DOI: 10.1038/sdata.2016.18. Eine laufend aktualisierte Version der Prinzipien wird unter <https://www.go-fair.org/fair-principles> gepflegt (05.02.2020).

ihre Berücksichtigung für kommende Förderrunden (z. B. *Horizon 2020*). Der *Schweizerische Nationalfonds* SNF fordert für die meisten Förderinstrumente seit Oktober 2017 die Einreichung eines *Data Management Plans* (DMP), in dessen Rahmen der SNF «von den Forschenden [erwartet], dass sie ihre Daten im Sinne der FAIR-Prinzipien in öffentlich zugänglichen, digitalen Datenbanken veröffentlichen.»<sup>7</sup> Gemeint sind dabei zentrale, projektübergreifende Datenbanken institutioneller Träger und nicht die oftmals auch datenbankbasierten und öffentlich zugänglichen individuellen Auftritte einzelner Projekte oder Projektverbünde.

Wenn im Folgenden die FAIR-Data-Kriterien an langjährige Erschliessungsprojekte zum 18. Jahrhundert herangeführt werden, ist erstens zu bedenken, dass die FAIR-Thematik noch sehr jung ist und in diesem Bereich noch viel Pionierarbeit zu leisten ist. Zweitens muss berücksichtigt werden, dass es sich bei digitalen Editions- und Forschungsplattformen nicht um statische Daten handelt, sondern um «lebende» Anwendungen.<sup>8</sup> Als solche können sie nur begrenzt und mittelbar gegen die FAIR-Prinzipien reflektiert werden, und es gilt neben den Daten als Gesamtpaket (z. B. einer Edition) auch ihre atomaren Bestandteile (z. B. ein einzelner Brief) zu berücksichtigen. Ziel dieses Aufsatzes ist es daher nicht, Defizite freizulegen, sondern auf gute Ansätze hinzuweisen und Potenziale aufzuzeigen.<sup>9</sup>

## Das Kriterium der Auffindbarkeit

Die elementarste Anforderung an eine gute Verfügbarkeit von Forschungsdaten besteht darin, dass diese für Mensch und Maschine problemlos auffindbar sind. Die FAIR-Prinzipien verlangen zu diesem Zweck, dass die Daten mittels gängiger Suchinstrumente leicht und replizierbar zu ermitteln sind:<sup>10</sup>

---

<sup>7</sup> Mit *Dryad*, *EUDAT*, *Harvard Dataverse* und *Zenodo* empfiehlt der SNF gegenwärtig vier Daten-Repositorien, die die SNF-Anforderungen erfüllen (Publikation von FAIR-Daten, kein kommerzielles Geschäftsmodell), vgl. [http://www.snf.ch/SiteCollectionDocuments/FAIR\\_data\\_repositories\\_examples.pdf](http://www.snf.ch/SiteCollectionDocuments/FAIR_data_repositories_examples.pdf) (05.02.2020).

<sup>8</sup> Vgl. Patrick Sahle, Simone Kronenwett, «Jenseits der Daten: Überlegungen zu Datenzentren für die Geisteswissenschaften am Beispiel des Kölner 'Data Center for the Humanities'», *LIBREAS*, 23, 2013, S. 76–96, hier S. 89, <https://libreas.eu/ausgabe23/09sahle> (05.02.2020).

<sup>9</sup> An dieser Stelle ist auch auf den Kriterienkatalog für die Besprechung Digitaler Editionen des Instituts für Dokumentologie und Editorik hinzuweisen, dessen Inhalte sich mit den FAIR-Prinzipien teilweise überschneiden, der aber spezifisch auf Digitale Editionen als interaktive Anwendungen ausgerichtet ist. Vgl. <https://www.i-d-e.de/publikationen/weitereschriften/kriterien-version-1-1> (05.02.2020).

<sup>10</sup> In Ermangelung einer deutschen Leitübersetzung paraphrasieren wir jeweils die Kriterien und Unterkriterien aus M. D. Wilkinson et al., *FAIR Guiding Principles*, Box 2. Konzise Auskunft über die Essenz der Kriterien und die daraus folgenden Verantwortlichkeiten der Forschenden liefert die unter [http://www.snf.ch/SiteCollectionDocuments/FAIR\\_principles\\_translation\\_SNSF\\_logo.pdf](http://www.snf.ch/SiteCollectionDocuments/FAIR_principles_translation_SNSF_logo.pdf) (05.02.2020) verfügbare Handreichung des SNF.

- F1: (Meta-)Daten sind mit einem weltweit einmaligen, zeitstabilen Identifikator zu versehen.
- F2: Daten sind mit umfassenden Metadaten zu beschreiben.
- F3: Die Beziehung von Metadaten und Daten ist stets explizit.
- F4: (Meta-)Daten werden in Suchindizes aufgenommen.

Um die Frage zu beantworten, ob und wie gut eine Datenressource auffindbar sei, ist man geneigt, eine *Google*-Suche zu starten und die erste Handvoll Treffer durchzugehen. Das Kriterium *findable* umfasst aber mehr als das bloße Auffinden eines Datenbestandes oder bestimmter Teile davon. Vielmehr geht es darum, dass Informationspunkte in sinnvoller Granularität mit ausreichend Kontextinformationen auffindbar sind, so dass das Gefundene präzise benannt werden kann und zu einem späteren Zeitpunkt oder einer Drittperson genauso auffindbar bleibt. Betrachtet man nun nicht ein Bündel statischer, in einem Repositorium archivierbarer Forschungsdaten, sondern eine auf solchen Daten basierende «lebende» Anwendung wie eine Forschungsplattform, so lassen sich alle vier Postulate gut transferieren.

Ohne klar definierte Identifikation (F1) von Objekten kommt eine Datenbank oder eine Digitale Edition nicht aus. Dabei können solche Identifikatoren automatisiert erzeugt werden (etwa durch die Speicherung in eine Datenbankanwendung) oder nach einer Konvention von Hand informationsreich bzw. sprechend aufgebaut werden. Um die globale Einmaligkeit zu erreichen, finden vor allem zwei Strategien Anwendung: die Vergabe eines 128-stelligen *Globally Unique Identifier* (GUID; auch: *Universally Unique Identifier* UUID) und die Vergabe einer projektspezifischen ID, welche Bestandteil einer Webadresse ist, so dass ein *Uniform Resource Identifier* (URI) gebildet wird, der zugleich das Objekt identifiziert und ein Verweis auf seine digitale Repräsentation ist. Identifikatoren sollen weiterhin klar ersichtlich werden, beispielsweise in einer Zitierempfehlung einer bestimmten Objektansicht. Zu den Plattformen, deren Objekte nach einer klaren Systematik identifizierbar sind, gehören *Lumières Lausanne*<sup>11</sup> und *Viat-images*<sup>12</sup>. Das letzte Beispiel illustriert zugleich eine von mehreren Methoden, wie sich auf Identifikationspersistierung spezialisierte Dienste nutzen lassen: Das Projekt erhält sich die Freiheit einer späteren Umprogrammierung, indem nicht die aktuelle technische Adresse<sup>13</sup> exponiert wird, sondern eine klarer gegliederte permanente URL, die über eine Konkordanz oder eine definierte Logik Aufrufe an die eigentliche Adresse weiterleiten kann.<sup>14</sup>

11 Beispiel: <http://lumières.unil.ch/fiches/trans/1170> (05.02.2020).

12 Beispiel: <https://purl.org/viatimages/fr/image/1711> (05.02.2020).

13 <https://www2.unil.ch/viatimages/index.php?projet=viatocalpes&module=image&action=detail&IDImage=1711&lang=fr> (05.02.2020).

14 *PURL* (<https://purl.org>; 05.02.2020) ist ein Dienst des *Internet Archive* (<https://archive.org>; 05.02.2020). Vergleichbare Funktionalität kann über Systeme wie DOI, Handle und URN erzielt werden, die allerdings etwas weniger niederschwellig angelegt sind.

Die umfassende Beschreibung mit Metadaten (F2) wird von vielen Projekten geleistet, besonders dann, wenn die Metadaten zur Beantwortung von Forschungsfragen interessant sein können. Ein unmittelbarer Nutzen einer tiefen Metadatenerfassung ist die Möglichkeit, Daten sehr spezifisch filtern zu können. Das Projekt *Mapping the Trade of the Société Typographique de Neuchâtel*<sup>15</sup> ist ein gutes Beispiel dafür. Beachtet werden sollte, dass Metadaten nicht nur zu den Quellen und den darin enthaltenen Informationen erhoben werden, sondern auch zum Kontext der Erhebung. Es sollte also ebenfalls ersichtlich gemacht werden, wer die Daten wann und in welchem Rahmen erarbeitet hat. Des weitern sollten die Metadaten vermitteln, was bei der Dateninterpretation zu berücksichtigen sei, also etwa Angaben dazu, wie umfassend eine Recherche ausgeführt wurde, wie Bewertungsskalen definiert sind oder wie eine Auswahl zustande kam.

Das Postulat der Explizitmachung der Beziehung von Metadaten und Daten (F3) ist in einer laufenden Anwendung weniger zentral als etwa bei Messdaten, weil der Zusammenhang zumeist klar gegeben ist, sei es durch Datenbankbezüge oder durch Ansichten, die neben Daten auch Metadaten aufführen. In vielen Editionsprojekten werden Informationen nach den Richtlinien der *Text Encoding Initiative* in einem XML-Format codiert. Wenn diese Daten objektzentriert aufgebaut sind, ist der Bezug in der Regel direkt gegeben: Für ein Objekt werden in der gleichen Datei Metadaten und Daten erfasst (im *teiHeader*- bzw. *body*-Element).

Bezüglich des Kriteriums der Suchindizes (F4) lässt sich zunächst festhalten, dass Suchmaschinen die frei im Web zugängliche Ressourcen in der Regel von selbst erfassen. Der Vorgang und die Bewertung der Suchergebnisse lässt sich aber optimieren, indem beispielsweise eine *Sitemap* erstellt und für jede Ansicht genau eine kanonische Adresse definiert wird. Eine noch bessere Auffindbarkeit wird durch maschinenlesbare semantische Beschreibungen der Daten möglich, weil Suchmaschinen die Informationen so zielgenauer und sinnhafter verknüpfen können. Damit eine Edition oder Plattform auch in themen- oder gattungsspezifischen Verzeichnissen und Suchdiensten gelistet wird, können Normdaten-URL-Konkordanzen (sog. *Beacons*) erstellt und deren Adresse den jeweiligen Diensten mitgeteilt werden. Die Aufnahme in Such- und Verknüpfungsdienste wie *Metagrid*<sup>16</sup>, *histHub*<sup>17</sup> oder *CorrespSearch*<sup>18</sup> kann über eine zu diesem Zweck eingerichtete Schnittstelle oder einen einmaligen oder periodischen Datenaustausch erfolgen. Bereits auf dem Weg zur Metadatenverlinkung mit *histHub* sind *Les points de coutume neuchâtelois*. Erwähnte Personen lassen sich in *histHub* bereits finden und aufrufen (mit Quellenangabe «Rechtsquellenstif-

---

15 <http://fbtee.uws.edu.au/stn/interface> (05.02.2020).

16 <https://metagrid.ch> (05.02.2020).

17 <https://histhub.ch> (05.02.2020).

18 <https://correspsearch.net> (05.02.2020).

tung des Schweizerischen Juristenvereins»), allerdings fehlt noch der kleine Schritt zur direkten Verlinkung der Daten.

## Das Kriterium der Zugänglichkeit

Neben der Auffindbarkeit von Daten bildet ihre freie Zugänglichkeit eine zweite fundamentale Säule der FAIR-Prinzipien. Die Anforderungen zur Zugänglichkeit zielen dabei wiederum auf statische Forschungsdaten in einem Datenspeicher ab. Das ist insofern zu beachten, als Zugänglichkeit (engl. *accessibility*) im Kontext von Webressourcen gemeinhin anders aufgefasst wird, nämlich als Barrierefreiheit.<sup>19</sup> Zugänglichkeit nach den FAIR-Prinzipien zielt demgegenüber auf die freie, einem offenen Standard folgende Abrufbarkeit von Metadaten und Daten ab:

- A1: (Meta-)Daten müssen ID-basiert über ein standardisiertes Protokoll abrufbar sein.
- A1.1: Das Kommunikationsprotokoll ist offen, frei lizenziert und universell implementierbar.
- A1.2: Wo nötig, ermöglicht das Protokoll eine Authentifizierung und Autorisierung des Zugriffs.
- A2: Auch wenn Daten nicht länger verfügbar sind, bleiben die zugehörigen Metadaten zugänglich.

Die Anforderung der Nutzung eines offenen und freien Kommunikationsprotokolls (A1) ist bei Webressourcen in aller Regel *eo ipso* gegeben, weil die Infrastruktur des Internets auf solchen Protokollen beruht. Abgesehen davon, dass auch vermittelte Zugriffe wie URN, PURL oder DOI die Basiskommunikation über *http/https* nutzen, basieren sie nicht auf proprietären oder kommerziellen Protokollen, so dass die A1-Kriterien vollumfänglich erfüllt bleiben. Indem diese Mechanismen auf einmaligen Identifikatoren beruhen (s. o.), sind sie im Gegenteil ausgesprochen FAIR-freundlich. Eine auf CD-ROM (oder DVD) publizierte Digitale Edition, wie es in einer Übergangsphase durchaus üblich war, vermag die Zugangskriterien dagegen nicht zu erfüllen.

Die in Unterkriterium A1.2 geforderte Authentifizierung und Autorisierung ist für die historische Erforschung des 18. Jahrhunderts nicht relevant, zumindest nicht, solange es sich um Forschungsdaten zu historischen Akteuren handelt. Der Punkt müsste dann Beachtung finden, wenn die Daten persönlichkeitsrechtlich heikle Informationen zu lebenden Personen enthalten (z. B. zu Projektbeteiligten), jedoch nicht, wenn es sich um öffentliche Informationen wie

---

<sup>19</sup> Vgl. zu dieser Unterscheidung Marisa Martinez et al., «Refining our conception of ‘access’ in digital scholarly editing: Reflections on a qualitative survey on inclusive design and dissemination», *Variants* 14, 2019, S. 41–74, Absätze 26–33 und 39–43.

Publikationen handelt. Anders verhält es sich mit unveröffentlichten Forschungsdaten und -ergebnissen. Diese lassen sich schützen, ohne dadurch den FAIR-Status einer Ressource zu gefährden.

Mit der oben beschriebenen Verschränkung von Daten und Metadaten in digitalen Ressourcen mindert sich auch die Bedeutung der Priorisierung der Metadaten (A2). Das Kriterium könnte für Webressourcen so interpretiert werden, dass dort, wo frühere Versionen unterschieden werden, zumindest die Metadaten der älteren Versionen weiterhin abrufbar wären. In der Praxis ist dies auf allgemeiner Ebene in einer Versionsbeschreibung (*changelog*) relativ leicht leistbar. Angestrebt werden sollte aber idealerweise ein Versionierungssystem, über das nicht nur Metadaten älterer Zustände, sondern auch die Daten selber aufrufbar sind. Ein bekanntes System dieser Art ist die Versionsgeschichte der Wikipedia. Ähnliche Ansätze finden allmählich auch in der digitalen Editorik Verbreitung, sie sind aber noch rar.

## Das Kriterium der Interoperabilität

Nach FAIR-Data-Kriterien geteilte Daten sollen für Mensch und Maschine operationalisierbar sein. Dies bedingt, dass das Modell, nach dem die Daten erfasst wurden, auf geeignete Weise kommuniziert wird:

- 11: Die (Meta-)Daten-Repräsentation erfolgt in formalisierter, konsensualer Sprache mit weitem Anwendungsbereich.
- 12: Die solchen Sprachen zugrundeliegenden (kontrollierten) Vokabulare folgen ihrerseits FAIR-Data-Prinzipien.
- 13: Abhängigkeiten zwischen (Meta-)Daten werden durch eindeutige Referenzen offengelegt.

Eine sinnvolle Interpretation und Verarbeitung von Daten ist nur möglich, wenn klar dokumentiert ist, welche Information auf welche Weise codiert wurde. Eine beschreibende Dokumentation in Prosaform ist ein erster Schritt, gefordert wird allerdings eine soweit formalisierte Beschreibungsart, dass auch eine (semi-)automatische Auswertung der Daten möglich wird. Ein Projekt, das sich an den TEI-Richtlinien orientiert wie etwa *SSRQ online*<sup>20</sup> oder die Edition der Lavater-Briefwechsel *JCLB*<sup>21</sup>, deren Metadaten überdies nach *MODS*<sup>22</sup> erfasst werden sollen, wird diesem Ziel schon relativ nahe kommen. Wahre Interoperabilität unter verschiedenen TEI-Projekten ist aber eine Illusion, weil es sich bei den TEI-

---

<sup>20</sup> <https://www.ssrq-sds-fds.ch/projekte/ssrq-online> (05.02.2020).

<sup>21</sup> <https://www.lavater.com/briefwechsel>. Die Online-Edition von *Johann Caspar Lavater: Historisch-kritische Edition ausgewählter Briefwechsel* (JCLB) wird 2021 aufgeschaltet.

<sup>22</sup> *Metadata Object Description Schema*; XML-basiertes bibliographisches Beschreibungsschema der *US Library of Congress*.



Richtlinien um keinen harten Standard handelt. Community-Bestrebungen wie die *EpiDoc*-Initiative<sup>23</sup> für (antike) epigraphische Texte oder das *Deutsche Textarchiv*<sup>24</sup> kommen dem Ziel deutlich näher, weil sie ein stark reduziertes Schema mit geteilten kontrollierten Vokabularien nutzen. Andere Projekte<sup>25</sup> wiederum verfolgen die Strategie ontologischer Datenmodelle, deren Subjekte, Objekte und Prädikate sich (bei hinreichender Übereinstimmung) projektübergreifend in Beziehung setzen lassen. Eine weitere Möglichkeit, Daten interoperabel zu verbinden, liegt in der Verwendung eines gemeinsamen Referenzmodells, auf welches das jeweils interne Modell gemappt werden kann.<sup>26</sup> Bestrebungen wie das *Correspondence Metadata Interchange Format* (CMIF)<sup>27</sup> oder das *prosopographi*-Projekt<sup>28</sup> entwickeln Methoden der Interoperabilität für bestimmte Bereiche wie Korrespondenzmetadaten oder prosopografische Daten. Unbesehen davon, welche Interoperabilitätsstrategie ein Projekt verfolgt, sollen die formalisierte Beschreibung des Datenmodells veröffentlicht und die Daten mit ihr in Bezug gesetzt werden.

Während die FAIR-Data-Prinzipien vorsehen, dass Daten in interoperationalisierbarer Weise passiv zur Verfügung gestellt werden, kann man bei digitalen Ressourcen einen Schritt weiter gehen, indem sie untereinander in aktiven Austausch gesetzt werden. Ein Beispiel dafür ist die auf Normdaten wie *GND* oder *Geonames* basierende Datenabfrage ab einer externen Quelle in Echtzeit wie die Einbindung von Wikipedia-Informationen in der *Carl-Maria-von-Weber-Gesamtausgabe*<sup>29</sup> oder die Abfrage des *Metagrid*-Dienstes, die punktgenaue Verweise zu anderen Ressourcen zurückgibt, in denen die betreffende Entität beschrieben wird.

23 <http://epidoc.sf.net> (05.02.2020).

24 <http://www.deutschestextarchiv.de> (05.02.2020).

25 Hier ist z. B. an die Nationale Infrastruktur für Editionen / Infrastructure nationale pour les éditions (NIE INE) zu denken, vgl. <https://www.nie-ine.ch> (05.02.2020).

26 Das bekannteste Beispiel in diesem Bereich ist das *CIDOC Conceptual Reference Model* aus dem Kulturerbe- und Museumsbereich, vgl. <http://www.cidoc-crm.org> (05.02.2020).

27 [https://correspsearch.net/index.xml?id=participate\\_cmi-format](https://correspsearch.net/index.xml?id=participate_cmi-format) (05.02.2020).

28 Georg Vogeler, Gunter Vasold, Matthias Schlögl, «Von IIF zu IPIF? Ein Vorschlag für den Datenaustausch über Personen», vorgetragen an der DHd-Konferenz 2019 in Mainz/Frankfurt (25.–29.03.2019), <http://web.archive.org/web/20200205004404/https://dhd2019.org/programm/do/1100-1230/session-datenmodelle-und-schnittstellen/vortrag-269> (05.02.2020).

29 Vgl. beispielsweise den Eintrag «Abelardus, Petrus» im *Weber-Personenregister*, für den Bild und Text direkt ab dem aktuellen Wikipedia-Artikel zu Abelard eingefügt werden: <https://webergesamtausgabe.de/de/A002750.html> (05.02.2020). Der sehr konzise projekteigene TEI-Code zu Abelard, auf dessen Basis die Abfrage geschieht, lässt sich ebenfalls einsehen.

## Das Kriterium der Wiederverwendbarkeit

Eng verbunden mit dem Interoperabilitäts-Postulat ist das vierte und letzte FAIR-Data-Kriterium der Wiederverwendbarkeit. Die Umsetzungskriterien hängen wiederum mit Anforderungen an Datenformate und Dokumentation zusammen, ein weiterer wesentlicher Aspekt kommt aber mit der offenen Lizenzierung dazu:

R1: (Meta-)Daten sollen umfassend beschrieben werden, sowohl als Ganzes (Herkunft, Umfang, Vorverarbeitungen usw.) als auch einzelne Einheiten betreffend (Datenmodell).

R1.1: (Meta-)Daten sind in eindeutiger Weise mit einer permissiven Lizenz zu versehen.

R1.2: Die Herkunft und Genese der Daten wird bestmöglich offengelegt.

R1.3: (Meta-)Daten werden nach domäne- und community-relevanten Standards modelliert.

Dieses Kriterium lässt sich uneingeschränkt auch auf Digitale Editionen und Plattformen und ihre Inhalte anwenden. Die TEI-Richtlinien stellen mit den fünf hochrangigen Metadaten-Elementen *fileDesc*, *encodingDesc*, *profileDesc*, *xenoData* und *revisionDesc* Möglichkeiten bereit, alle geforderten Aspekte gut zu beschreiben.<sup>30</sup> Auf einer tieferen Ebene von *fileDesc* können im Element *licence* Angaben zur Lizenzierung gemacht werden. All diese Möglichkeiten bestehen auf verschiedenen Ebenen: Sie können sich auf die gesamte Ressource beziehen, nur einzelne Dateien betreffen oder nur bestimmte Teile einzelner Dateien. Auch Beschreibungsstandards wie *METS* oder *MODS* sehen vergleichbare Möglichkeiten vor.

Informationen zur Datenherkunft, zu erfolgten Verarbeitungsschritten und zu Spezifika des verwendeten Modells lassen sich auch in weniger formalisierter Art beschreiben, etwa in Transkriptionsrichtlinien wie es beispielsweise das SSRQ-Projekt macht, dessen Teil *Les points de coutume neuchâtelois* ist.<sup>31</sup>

Gute Beispiele einer FAIR-konformen Lizenzierung der konsultierten Plattformen sind SSRQ und die *Literary Tour de France*<sup>32</sup>, die beide ihre Daten unter einer freien *Creative Commons*-Lizenz teilen (CC). In beiden Fällen handelt es sich um die *CC Attribution-NonCommercial*-Lizenz, wobei *Les points de coutume neuchâtelois* zusätzlich die Verwendung einer kompatiblen (freien) Lizenz zur Bedingung für die modifizierende Nachnutzung macht (*ShareAlike*). Nach wie vor scheuen fast alle Projekte (und Verlage) eine freie Lizenzierung, die auch eine kommerzielle Nachnutzung ermöglichen würde. Mit einer *non-commercial*-

<sup>30</sup> Vgl. Kapitel 2 *The TEI Header*, <https://tei-c.org/release/doc/tei-p5-doc/en/html/HD.html> (05.02.2020).

<sup>31</sup> Vgl. <https://www.ssrq-sds-fds.ch/wiki/Transkriptionsrichtlinie> (05.02.2020).

<sup>32</sup> <http://www.robertdamton.org> (05.02.2020).

Klausel können allerdings die Bedingungen der *Budapest Open Access-Initiative* (BOAI) von 2002 nicht erfüllt werden.<sup>33</sup>

## Die Editions- und Forschungsplattform *hallerNet*<sup>34</sup>

Unsere Auseinandersetzung mit den FAIR-Prinzipien ergab sich aus der Entwicklung und dem Ausbau von *hallerNet*. Diese Editions- und Forschungsplattform und die ihr zugrundeliegenden Daten, die aus verschiedenen, teils viele Jahre zurückliegenden Forschungsprojekten stammen, bieten sich an, die Notwendigkeit einer kontinuierlichen Pflege und Fortnutzung digitaler Daten zu thematisieren. Im Bewusstsein, dass die Zeit nicht stillsteht und sich für die Daten der Plattform auch in Zukunft neue Nutzungsarten ergeben werden, erwuchs schnell ein genuines Interesse an Fragen der guten wissenschaftlichen Praxis im Zusammenhang mit Forschungsdaten und ihrer Langzeitverfügbarkeit. In diesem Abschnitt werden zunächst die Projektgeschichte kurz skizziert und die Charakteristika der Plattform vorgestellt, bevor wir im nächsten Abschnitt an diesem Beispiel etwas ausführlicher auf die Frage nach der Berücksichtigung der FAIR-Kriterien eingehen.

Die Plattform *hallerNet* baut auf umfangreichen prosopografischen und bibliografischen Strukturdaten, die in Form einer relationalen Verbunddatenbank seit anfangs der 1990er-Jahre im Rahmen der beiden SNF-Projekte zum Universalgelehrten Albrecht von Haller (1708–1777)<sup>35</sup> und zur aufgeklärten Reformsozietät *Oekonomische Gesellschaft Bern* (gegr. 1759)<sup>36</sup> erarbeitet wurden. Das Transformationsprojekt *haller online* (2016–2019), das im Auftrag der *Albrecht von Haller-Stiftung der Burgergemeinde Bern* als Kooperation des Historischen Instituts der Universität Bern und des *Cologne Center for eHumanities CCeH* erfolgte,<sup>37</sup> überführte diese untereinander vielfältig verknüpften Daten zu

<sup>33</sup> Vgl. dazu Open Knowledge Foundation, Sharing the results of publicly funded research. Definition of Budapest compliant open access, <https://access.okfn.org/definition/index.html> (05.02.2020).

<sup>34</sup> Dieser Abschnitt folgt Martin Stuber, Peter Dängeli und Christian Forney, «Vom Stellenkommentar zum Netzwerk und zurück: grosse Quellenkorpora und tief erschlossene Strukturdaten auf *hallerNet*», vorgetragen an der DHd-Konferenz 2019 in Mainz/Frankfurt (25.–29.03.2019).

<sup>35</sup> *Albrecht von Haller und die Gelehrtenrepublik des 18. Jahrhunderts* (1991–2003); Gesuchsteller: Urs Boschung (Institut für Medizingeschichte Universität Bern).

<sup>36</sup> *Nützliche Wissenschaft, Naturaneignung und Politik. Die Oekonomische Gesellschaft Bern im europäischen Kontext* (2004–2011); Gesuchstellende: André Holenstein, Christian Pfister; Koordinator: Martin Stuber (Historisches Institut der Universität Bern).

<sup>37</sup> Unter der Gesamtleitung von Christophe v. Werdt (Albrecht von Haller-Stiftung der Burgergemeinde Bern) waren an der Finanzierung (Kreditvolumen 1.18 Mio CHF) zudem die folgenden Institutionen beteiligt: Burgergemeinde Bern, Ernst Göhner Stiftung, Fondation Johanna Dürmüller-Bol, Gesellschaft zu Ober-Gerwern Bern, Lotteriefonds Kanton Bern, Schweizerischer Nationalfonds zur Förderung der wissenschaftlichen Forschung (SNF), Stiftung Pro Scientia et Arte, Universität Bern, Ursula Wirz-Stiftung, Institut für Kulturforschung Graubünden. Projektleitung: Martin Stuber, Peter Dängeli, Patrick Sahle.

rund 48'000 Publikationen, 25'000 Personen, 20'000 Briefen, 3'000 Orten, 2'900 Pflanzenarten, 1'000 Versammlungen und 850 Institutionen in eine TEI-konforme XML-Datenstruktur.<sup>38</sup> Auf der Grundlage der neuen Plattform, dem ihr zugrundeliegenden Datenmodell und der mit ihr in Zusammenhang stehenden Editionsworkflows startete anfangs 2018 ein vom *Schweizerischen Nationalfonds* (SNF) im Rahmen der Infrastrukturförderung unterstütztes Editions- und Forschungsprojekt, das im geplanten Projektzeitraum von sechs Jahren zum einen die Gesamtedition der rund 9'000 Buchbesprechungen Albrecht von Hallers erarbeitet. Zum anderen wird eine begründete Auswahl von inhaltlich damit zusammenhängenden rund 7'000 Briefen ediert, dies als Zwischenetappe zur längerfristig angestrebten Gesamtedition von Hallers Korrespondenz, die insgesamt rund 17'000 Briefe umfasst. Über Haller und die *Oekonomische Gesellschaft Bern* hinaus enthält *hallerNet* weitere gelehrte Sammlungen aus diesem Zeitraum. Integriert wurden beispielsweise eine Auswahledition des Briefwechsels von Johann Jakob Scheuchzer (1672–1733)<sup>39</sup> sowie umfangreiche Metadaten, die im Rahmen des SNF-Projekts *Kulturen der Naturforschung* (2013–2017) erarbeitet wurden, namentlich zu Johannes Gessner (1709–1790) und zur *Naturforschenden Gesellschaft Zürich* (gegr. 1746).<sup>40</sup> Rückblickend lässt sich feststellen, dass die Ausweitung der ursprünglich ausschliesslich an der Zentralfigur ausgerichteten Haller-Datenbank die Nachhaltigkeit der Daten entscheidend erhöht. Um die Integration von zusätzlichen Daten möglich zu machen, mussten vorgängig die Datenstrukturen verallgemeinert werden. Zudem führten die neuen Daten zu Abgleichungs- und Homogenisierungsvorgängen, da man die Generierung von Dubletten unbedingt vermeiden wollte.

Eine Spezialität von *hallerNet* sind die systematischen Verknüpfungen zwischen den neu entstehenden umfangreichen Texteditionen und Quellendigitalisaten mit den über Jahrzehnte erarbeiteten Metadaten, die sich in ihrem bibliografischen Teil auf zwei autoptisch erarbeitete umfassende Grundlagenwerke stützen.<sup>41</sup> Dadurch wird erstens die Editionsarbeit erheblich erleichtert, namentlich die Referenzierung von Akteuren, Institutionen, Orten und Publikationen.

---

38 Ute Recker-Hamm, Martin Stuber, *Haller Online – Konzept für den Umbau, Ausbau und die langfristige Sicherung der Haller-/ OeG-Datenbank* (8.6.2015); Christian Forney, Peter Dängeli und Antonio Rojas Castro, «Vom geschützt zugänglichen Datenbankverbund zur offenen Editions- und Forschungsplattform. Kritischer Rückblick auf halber Strecke», Poster an der DHd-Konferenz in Köln (26.02.–02.03.2018).

39 Simona Boscani Leoni (Hg.), «*Lettres des Grisons*»: *Wissenschaft, Religion und Diplomatie in der Korrespondenz von Johann Jakob Scheuchzer. Eine Edition ausgewählter Schweizer Briefe (1695–1731)*, *hallerNet* 2019, <https://haller.net/edition/scheuchzer-korrespondenz> (05.02.2020).

40 Gesuchstellerin: Simona Boscani Leoni, Mitarbeiterinnen: Sarah Baumgartner, Meike Knittel (Historisches Institut der Universität Bern), <http://p3.snf.ch/Project-144731> (05.02.2020).

41 Maria Teresa Monti (Hg.), *Catalogo del Fondo Haller della Biblioteca Nazionale Braidense di Milano*, 13 Bde., Milano, F. Angeli, 1983–1994; Hubert Steinke, Claudia Profos, *Bibliographia Halleriana. Verzeichnis der Schriften von und über Albrecht von Haller*, Basel, Schwabe, 2004.

Zweitens liefern die daraus generierten Forschungsdaten sowohl in der raumzeitlichen Entwicklung als auch in der inhaltlichen Qualität der Netzbeziehungen eine hohe Präzision und erlauben Antworten auf wichtige Postulate der aktuellen historischen Netzwerkforschung.<sup>42</sup>

Solche «datenzentrierte» Digitale Editionen sind bisher die Ausnahme gegenüber den vielen existierenden und sich in Entwicklung befindlichen «textzentrierten» Digitalen Editionen. Wo es sie gibt, verfolgen sie zumeist ganz spezifische Forschungsfragen und legen dazu ein entsprechend verengtes Datenmodell zugrunde.<sup>43</sup> Daneben existieren biografische, prosopografische oder bibliografische Datenbanken und Forschungsportale, die sich einzig auf Metadatenerhebung und -analyse konzentrieren.<sup>44</sup> Selbstredend lagen auch der Datenerfassung in den verschiedenen Projekten um Albrecht von Haller seit den 1990er-Jahren immer spezifische Forschungsinteressen zugrunde. Auf eine erste Phase, welche die Gesamterschließung von Hallers Korrespondenz<sup>45</sup> mit einer umfassenden prosopografischen Strukturanalyse<sup>46</sup> kombinierte, folgte eine zweite Phase, in der Netzwerkanalysen im Vordergrund standen und die in den Blick genommenen Akteure über Hallers Korrespondenznetz hinaus erweitert wurden.<sup>47</sup> Die dieser Forschung zugrundeliegenden Daten wurden aber von Anfang an in der gleichen Datenbankanwendung (FAUST) und mit möglichst viel struktureller Homogenität gepflegt. Vergleichsweise früh hatte sich eine – über die Projektzyklen hinausweisende – digitale Langzeitperspektive entwickelt. In

42 Siehe z. B. Otto Sonntag, Martin Stuber, Christian Forney: «Göttingen and Its Learned Institutions in Albrecht von Haller's European Network: The Example of His Correspondence with Gerlach Adolph von Münchhausen», in: Karsten Engel (Hg.), *Wissenschaft in Korrespondenzen*. Göttinger *Wissensgeschichte in Briefen*, Göttingen, Vandenhoeck & Ruprecht, 2019, S. 59–76.

43 Beispiele hierfür sind etwa die Projekte *Proceedings of the Old Bailey* ([www.oldbaileyonline.org](http://www.oldbaileyonline.org)), *buckhardtsource.org* (<http://burckhardtsource.org>), *Intoxicants and Early Modernity* (<https://www.intoxicantsproject.org>), *Jahrrechnungen der Stadt Basel 1535 bis 1610* (<https://gams.uni-graz.at/context:srbas>), *Urfehdebücher der Stadt Basel* (<https://gams.uni-graz.at/context:ufbas>), *Sound Toll Registers* (<http://soundtoll.nl>) oder das leider dem digitalen Zerfall anheimgefallene *Alcalá account book*.

44 Als Beispiele zur Gelehrtenwelt lassen sich hier *Amore Scientiae Facti sunt Exules* (<http://asfe.unibo.it>) oder das *Repertorium Academicum Germanicum* (<http://rag-online.org>) nennen. Bemerkenswert sind auch die der *Symogih*-Plattform (Système modulaire de gestion de l'information historique, <http://symogih.org>) angegliederten Projekte, die teilweise auch edierte Texte umfassen.

45 Urs Boschung et al. (Hg.), *Repertorium zu Albrecht von Hallers Korrespondenz 1724–1777*, 2 Bde., Basel, Schwabe, 2002.

46 Martin Stuber, Stefan Hächler, Luc Lienhard (Hg.), *Hallers Netz. Ein europäischer Gelehrtenbriefwechsel zur Zeit der Aufklärung*, Basel, Schwabe, 2005.

47 Regina Dauser et al. (Hg.), *Wissen im Netz. Botanik und Pflanzentransfer in europäischen Korrespondenznetzen des 18. Jahrhundert*, Berlin, Akademie Verlag, 2008; Martin Stuber, Lothar Kreppe, «Las redes académicas de Albrecht von Haller y la Sociedad Económica: un análisis de redes a varios niveles», *REDES. Revista hispana para el análisis. De redes sociales*, 24/1, 2013, S. 1–26, DOI: 10.5565/rev/redes.450 / REDES Online English: «The Scholarly Networks of Albrecht von Haller and the Economic Society of Bern – a Multi-Level Network Analysis»: [http://revista-redes.rediris.es/webdes/novedades/Stuber\\_Kreppe\\_scholarly\\_networks.pdf](http://revista-redes.rediris.es/webdes/novedades/Stuber_Kreppe_scholarly_networks.pdf).

der Folge wurde diese zum einen von langjährigen personellen Kontinuitäten getragen, zum anderen von der Albrecht von Haller-Stiftung der Burgergemeinde Bern.<sup>48</sup>

## FAIR-Data-Kriterien in *hallerNet*

Sowohl in der Projektierungsphase des Transformationsprojekts *haller online* als auch in der Antragsphase des laufenden SNF-Projekts waren Fragen des Forschungsdatenmanagements noch nicht gleichermassen präsent wie heute. Das Ausführungsreglement zum Beitragsreglement des SNF, das die FAIR-Prinzipien explizit erwähnt, datiert vom 9. Dezember 2015. Erst seit Oktober 2017 ist die Einreichung eines *Data Management* Plans für SNF-Projekte verpflichtend. Die lange Kontinuität mit viel Aufwand erhobener, digitaler Daten der Berner Hallerforschung rückte das Thema aber auch ohne diese äusseren Einflussfaktoren vom ersten Tag weg ins Zentrum. Schon früh im Transformationsprojekt fiel der Entscheid, dass die Masterfassung aller Forschungsdaten künftig in Dateiform geführt wird und zwar in offenen Volltextformaten (sowie für Digitalisate in kompressionsfreien Rasterformaten). Indem Masterdaten nicht in datenbankspezifischen Binärformaten abgespeichert werden, vermindert sich die Abhängigkeit von bestimmten Werkzeugen. Zugleich lassen sich die Daten sehr einfach und zuverlässig versionieren.

Aus diesem Entscheid hat sich eine Doppelstrategie entwickelt, die das Projekt hinsichtlich der FAIR-Prinzipien in eine relativ komfortable Lage bringt. Zum einen existiert *hallerNet* auf der Ebene eines kommunen Dateisystems, aus dem Ablieferungen an ein FAIR-kompatibles Repositorium direkt möglich sind. Zum anderen steht *hallerNet* als digitale Plattform im Web, die auf den gleichen Einzeldateien beruht. Selbstverständlich bemühen wir uns, die Plattform so auszugestalten, dass möglichst viele der FAIR-Data-Kriterien erfüllt werden. Dass eine laufende Anwendung viele Vorteile gegenüber einem statischen Datendepositum aufweist, muss nicht erst erklärt werden. Allerdings werden diese Vorteile mit einer grösseren Fragilität über die Zeit erkaufte. Im Wissen um diese Nachhaltigkeitsproblematik digitaler Anwendungen stellt das vollumfänglich an spezialisierte institutionelle Repositorien ablieferbare Dateisystem ein beruhigendes Sicherheitsnetz dar: Seit Mitte Januar 2019 werden alle Forschungsdaten des

---

48 Siehe Daniel Flückiger, Martin Stuber, «Vom System zum Akteur. Personenorientierte Datenbanken für Archiv und Forschung», in: André Kirchhofer et al. (Hg.), *Nachhaltige Geschichte. Festschrift für Christian Pfister*, Zürich, Chronos, 2009, S. 253–269; Hubert Steinke, «Archive databases as advanced research tools: the Haller Project», in: Maria Teresa Monti (Hg.), *Antonio Vallisneri. L'edizione del testo scientifico d'età moderna*, Firenze, Olschki, 2003, S. 191–204; Martin Stuber, «Findmittel und Forschungsinstrument zugleich. Die Datenbank des Berner Haller-Projekts», *Arbedo*, 4/14, 1999, S. 5–10.

Projekts in ein versionskontrolliertes Git-Repository gespeichert.<sup>49</sup> Die Erweiterungen und Mutationen, die das Projektteam im *oXygen-XML-Editor* vornimmt, werden täglich nach Typ gesammelt als *Git-Commit* erfasst. Zusätzlich erfolgen umfassende datenkurative Bearbeitungen in der gleichen Form. So ergibt sich eine relativ übersichtliche zeitliche Versionierung der Daten, die im ersten Betriebsjahr zu etwas mehr als 1'300 Bearbeitungsschichten führte. Jede einzelne dieser Schichten kann mit einem einfachen Befehl wiederhergestellt werden. Gegenwärtig läuft ein interner Testbetrieb mit einer teilautomatisierten Übergabe an das am *CERN* angesiedelte Datenrepositorium *Zenodo*, das Teil der *Open Access Infrastructure for Research in Europe* (OpenAIRE) ist und bezüglich Open-Access-Kriterien sehr gut abschneidet.<sup>50</sup> Zugleich prüfen wir Optionen für ein wissenschaftsnahes externes Hosting der *hallerNet*-Dateirepositories.<sup>51</sup> Der Zugriff auf die *hallerNet*-Daten über *Zenodo* wird in den nächsten Monaten möglich sein.<sup>52</sup>

Hinsichtlich der oben diskutierten FAIR-Aspekte fällt ein Fazit der gegenwärtigen Umsetzung auf *hallerNet* insgesamt gemischt aus. Manche der geforderten Kriterien können als hinreichend umgesetzt betrachtet werden, für andere bestehen erst die Grundlagen und wiederum andere werden realistischerweise kaum leistbar sein. Die vier Kriterien werden im Folgenden gesondert diskutiert.

**Auffindbarkeit:** Seit ihrem Launch am 8. Mai 2019 ist die Plattform *hallerNet* online verfügbar. Die Plattform und ihre Inhalte können mit Suchmaschinen grundsätzlich gefunden werden, aber weil diesbezüglich noch keinerlei Optimierungen vorgenommen wurden, gibt es viel Verbesserungsspielraum. Die Aufnahme von (Meta-)Daten in externe Suchindizes ist teilweise bereits erfolgt (z. B. *Metagrid*), anderswo ist sie in Vorbereitung (z. B. *histHub*, *CorrespSearch*). Jedes auf der Plattform verzeichnete Objekt ist mit einem einmaligen und zeitstabilen Identifikator der Form «[https://hallernet.org/person\\_01200](https://hallernet.org/person_01200)» versehen. Diese Identifikatoren kommen auch in den Zitierempfehlungen zum Ausdruck. Im Zusammenhang mit dem Kriterium *findable* sind zudem die Retroeditionen bereits bestehender Druckeditionen zu nennen. Deren Integration in *hallerNet*, die mit viel Homogenisierungs- und Vernetzungsaufwand einhergeht, macht die grundlegenden Arbeiten der Vorgänger sehr viel «auffindbarer». Bereits erfolgt ist dies bei der *Bibliographia halleriana* (Druck 2004),<sup>53</sup> in Vorbereitung befind-

49 Git ist eine freie Software zur verteilten Versionsverwaltung von Dateien, vgl. <https://git-scm.com>.

50 Vgl. <https://about.zenodo.org/principles> (05.02.2020).

51 Eine interessante Option ist die frei zugängliche Renku-Instanz des *Swiss Data Science Centers*, einer Verbundinstitution von *EPF Lausanne* und *ETH Zürich*, <https://renkulab.io> (05.02.2020).

52 Diese Vorgehensweise bezweckt neben der langfristigen Zugänglichkeit der Forschungsdaten und der Vergabe eines zeitstabilen Identifikators über das *Digital Object Identifier-System* (DOI) auch ihre Auffindbarkeit über spezialisierte Suchmaschinen wie *DataCite* oder die *Google Dataset Search*.

53 Vgl. Anm. 41.

den sich die Retroeditionen von Hallers umfangreichen Briefwechseln mit Bonnet (1983), Pringle (1999), Saussure (1990), Trew (1999) und Werlhof (2014).<sup>54</sup>

Zugänglichkeit: Der Abruf von Daten der Plattform *hallerNet* erfolgt über stabile, FAIR-konforme Identifikatoren und das offene http-Protokoll. Sollten Daten wie Transkriptionen oder Digitalisate eines Objekts auf der Plattform nicht mehr angeboten werden (können), sind auch die direkt damit verbundenen Metadaten nicht mehr verfügbar. Von einer unwahrscheinlichen, aber prinzipiell möglichen umfassenden Löschung eines Objekts von der Plattform wären auch dessen Metadaten betroffen (inkl. Referenzen auf das Objekt an anderer Stelle). Weiterhin vorhanden wären die Metadaten aber auf jeden Fall in früheren Versionen des Dateirepositoriums. Inwiefern die Anzeigemöglichkeit älterer Versionen auch für die Plattform umgesetzt werden kann, lässt sich gegenwärtig nicht abschätzen.

Interoperabilität: Den verschiedenen Datentypen der *hallerNet*-Plattform liegt jeweils ein eigenes TEI-Modell zugrunde (Personen, Briefe, Publikationen, Orte, usw.). Durch die Semantik des verwendeten TEI-Vokabulars sind die Daten zu einem gewissen Grad auch für Projektexterne interpretierbar.<sup>55</sup> Zusätzlich sind alle wichtigen Elemente mit einem Nummer-Attribut versehen, über das Informationen präzise adressierbar sind. Eine Dokumentation, die Aufschluss gibt über die spezifische Bedeutung dieser systematischen Einheiten, ist allerdings noch zu leisten. Für eine wachsende Anzahl von Bereichen besteht ein kontrolliertes Vokabular in einer Taxonomie-Datei, die auch Teil der Projektdaten ist. Das Kriterium *Interoperable* wird auf *hallerNet* zudem wesentlich unterstützt durch die plattforminterne Homogenisierung und Vernetzung. Ausgangspunkt ist das Prinzip, dass jede Entität nur genau einmal im System vorhanden ist, darauf aber von überall her darauf verwiesen werden kann. Damit verbunden ist das Bemühen, historische Namensvarianten und Entitätsverschiebungen auf der Plattform selber kontrolliert und im Detail nachvollziehbar zu homogenisieren.

Aktive Interoperabilität wird aktuell in den Bereichen der Verlinkung von Entitäten und der Einbindung von Digitalisaten genutzt:

---

54 Otto Sonntag (Hg.), *Paul Gottlieb Werlhofs Letters to Albrecht von Haller*, Basel, Schwabe, 2014; Ders. (Hg.), *John Pringle's Correspondence with Albrecht von Haller*, Basel, Schwabe, 1999; Ders. (Hg.), *The Correspondence between Albrecht von Haller and Horace-Bénédict de Saussure*, Bern, H. Huber, 1990; Ders. (Hg.), *The Correspondence between Albrecht von Haller and Charles Bonnet*, Bern, H. Huber, 1983; Hubert Steinke (Hg.), *Der nützliche Brief. Die Korrespondenz zwischen Albrecht von Haller und Christoph Jakob Trew*, Basel, Schwabe, 1999.

55 Auf die enge Verzahnung mit den TEI-Standards verweisen auch die Bemühungen, Erfahrungen aus *hallerNet* in die TEI-Community einzubringen: Christian Forney, Susanne Haaf, Linda Kirsten, «Letter Openers and Closers», in: Stefan Dumont, Susanne Haaf and Sabine Seifert (Hg.), *Encoding Correspondence. A Manual for Encoding Letters and Postcards in TEI-XML and DTABf*, Berlin 2019. URL: <https://encoding-correspondence.bbaw.de/v1/openers-closers.html>; Diess., «Postscripts of Letters», in ebda. URL: <https://encoding-correspondence.bbaw.de/v1/postscripts.html> (05.02.2020).



- Basierend auf Normdatenangaben wie des GND- oder VIAF-Identifikators wird über die *Metagrid-API* eine Liste verbundener Dienste abgefragt, die Informationen zum betroffenen Objekt bereitstellen. Umgekehrt werden auf an *Metagrid* angeschlossenen Portalen auch Verweise auf die *hallerNet*-Entitäten ausgegeben.
- Die Entitätsverweise der *hallerNet*-Plattform können auch über Beacon-Listen abgefragt werden. Die Objektansichten lassen sich aber zusätzlich auch direkt über die Normdatenidentifikatoren abrufen.<sup>56</sup>
- Wo verfügbar, werden extern angefertigte und bereitgestellte Briefdigitalisate über die IIF-Schnittstelle eingebunden, aktuell vor allem für Scheuchzer-Briefe, die durch die Plattform *e-manuscripta*<sup>57</sup> oder im Fall von Scheuchzers Einladungsbrief<sup>58</sup> auf dem Server des *Münchner Digitalisierungszentrums der Bayerischen Staatsbibliothek*<sup>59</sup> bereitgestellt werden. Indem die im Rahmen des Projekts angefertigten Briefdigitalisate ebenfalls über IIF-Schnittstellen angeboten werden, ist die gleiche Vorgehensweise prinzipiell auch andersrum möglich.<sup>60</sup>

Als nächste Entwicklung in diesem Bereich wird die Aufnahme der *hallerNet*-Briefe in *CorrespSearch* angestrebt. Durch die Nutzung des *CMIF*-Formats bildet auch hierfür ein formalisiertes, konsensuales Modell die Grundlage.

Wiederverwendbarkeit: Durch die Nutzung des TEI-Instrumentariums und einer freien Lizenzform erfüllen die *hallerNet*-Daten das Kriterium der Wiederverwendbarkeit weitgehend. Die Datenherkunft ist auf übergreifender Ebene gut beschrieben, und es ist nachvollziehbar, im Rahmen welcher Projekte und zur Vorbereitung welcher Publikationen bestimmte Daten kompiliert wurden. Auf der tieferen Ebene einzelner Aspekte der Daten trifft das nicht gleichermassen zu. Eine derart detaillierte Aufzeichnung wäre kaum praktikabel gewesen bzw. aufgrund des grossen Zusatzaufwands fast zwangsläufig zu Lasten der jeweiligen Projektziele gegangen. Sowohl die in der Vergangenheit erarbeiteten Daten (in ihrer heutigen Gestalt) als auch die in laufenden Projekten entstehenden Transkriptionen und Metadaten werden unter der Lizenz *CC Attribution 4.0 International (CC BY 4.0)* veröffentlicht. Während die Lizenzierung der im Projekt angefertigten Digitalisate noch nicht abschliessend geregelt ist, wird auch aller Programmcode unter freier Lizenz veröffentlicht und ist prinzipiell nachnutz-

---

<sup>56</sup> Diese Funktionalitäten basieren auf Adressen der Form <https://hallernet.org/lod/gnd> bzw. <https://hallernet.org/lod/resolve/gnd:118545140> (analog für *VIAF* und *Geonames*).

<sup>57</sup> <https://www.e-manuscripta.ch> (05.02.2020).

<sup>58</sup> <https://hallernet.org/edition/scheuchzer-einladungsbrief> (05.02.2020).

<sup>59</sup> <https://www.digitale-sammlungen.de> (05.02.2020).

<sup>60</sup> Die Umsetzung ist vorerst auf die *Image API* eingeschränkt. Nach Möglichkeit werden später auch *IIF Manifests* angeboten (*Presentation API*), so dass sich die Dokumente im jeweils relevanten Kontext abfragen lassen, beispielsweise eine Seitenfolge eines Briefes.

bar.<sup>61</sup> Wir schätzen das Privileg, unsere Daten so frei teilen zu können und bemühen uns, auf der Plattform künftig deutlicher auf die Möglichkeiten der Nachnutzung hinzuweisen.

## Fazit

Die meisten Editions- und Forschungsplattformen zum 18. Jahrhundert enthalten umfangreiche Handschriftenkorpora, deren Aufbereitung nur mittels grossem Personalaufwand und über einen längeren Zeitraum zu realisieren ist. Schon allein deshalb ist die Anwendung der FAIR-Data-Kriterien ein Gebot der ökonomischen Vernunft. Zwar verbessert sie die Alterungsaussichten einer «lebenden» Anwendung nicht direkt; ihre Umsetzung ermöglicht aber einen zuverlässigen Zugriff auf Forschungsdaten und stellt klar, unter welchen Bedingungen sich diese nachnutzen lassen. In der Doppelstrategie von versionskontrollierten Daten in einem spezialisierten, den FAIR-Data-Kriterien genügenden Repository auf der einen und einer hinsichtlich der Punkte *findable*, *accessible*, *interoperable* und *reusable* optimierten Webplattform auf der anderen Seite besteht ein zweckmässiger Ansatz. Für *hallerNet* gibt es dabei auf beiden Seiten noch einiges zu tun und sicherlich auch viel zu lernen. Es steht ausser Frage, dass die Nachhaltigkeit konzeptionell und in ihrer technischen Umsetzung ein zentrales Qualitätsmerkmal digitaler Erschliessungsprojekte bildet und als integraler Bestandteil digitaler Editionstätigkeit zu sehen ist.

Peter Dängeli, lic. phil., MA, Historisches Institut der Universität Bern, *hallerNet*,  
Länggassstrasse 49, CH-3012 Bern, peter.daengeli@hist.unibe.ch

Dr. Martin Stuber, Historisches Institut der Universität Bern, *hallerNet*,  
Länggassstrasse 49, CH-3012 Bern, martin.stuber@hist.unibe.ch

---

<sup>61</sup> Zum Einsatz kommt die ISC-Lizenz des *Internet Software Consortiums*, vgl. <https://opensource.org/licenses/ISC> (05.02.2020). Diese Lizenzierung trifft auch auf die Datentransformationsskripte des Projekts *haller online* zu, die aufgrund ihrer hohen Spezifität für Dritte aber nur konzeptuell-methodisch interessant sein können.