

# FactGrid, Eine Wikibase-Installation für die Geschichtswissenschaften

*Mit dem FactGrid – einer Datenbank für die historische Forschung – macht das Forschungszentrum Gotha Geschichtsprojekten weltweit das Angebot, Daten bei uns auf einer einzigen mehrsprachigen Plattform kollektiv und auf offenes Wachstum hin zu organisieren. Gefördert von Wikimedia Deutschland erwies sich dies mit dem Rückenwind des öffentlichen Interesses an Wikibase als überraschend attraktives aber auch problematisches Angebot. Die Software weist ganz eigene Defizite auf, doch ermöglichte sie uns weitreichende strategische Allianzen einzugehen, während unser Angebot noch lange nicht auch tatsächlich schon attraktiv ist in den traditionell am Buch und der abschließenden Publikation orientierten Geisteswissenschaften.*

---

*FactGrid – “a database for historians” – is a Wikibase installation launched in 2018 by the Gotha Research Centre. Running as Wikimedia Germany’s show case of an independent Wikibase we could convince researchers from all over the world to run data on our platform – across all language barriers, collectively, with the aim to link into the upcoming global exchange of authority control data – a success story with drawbacks: Open data is not yet an attractive concept in the humanities; open to growth seems to be even more precarious perspective. The software has its own deficits which we are trying to mend.*

*The article sheds light into the developments in which we were able to enter far reaching strategic alliances.*

Das mit den Frontends als Flaschenhälsen ist natürlich nicht übertrieben. Im Januar 1994 verbrachte ich eine Woche vor einem solchen Frontend in Augsburgs Universitätsbibliothek. Ich hatte mit dem soeben bewilligten Promotionsthema versprochen, alles zu lesen, was Kunden mit Geschmack am Roman in deutscher und englischer Sprache zwischen 1710 und 1720 in den Buchhandlungen frisch gedruckt angeboten erhielten. Auf eine Bibliographie für dieses sehr spezielle Materialcorpus konnte ich nicht zurückgreifen. Es ist dies ein Corpus, das sich erheblich von dem der Nationalliteraturen unterscheidet: In ihm gibt es spätantike Klassiker wie aktuelle Übersetzungen neuester Politromane des niederländischen Marktes französischer Sprache, die entgegen aller Epochenannahmen gleichzeitig in Leipzig und London gelesen wurden.

Die Bibliotheken des deutschsprachigen Raums standen meiner Recherche unvorbereitet gegenüber. Auch die neuen OPAC-Kataloge rechneten ausschließlich mit Nutzern, die bereits wussten, welche Titel sie suchten. Die Suchen nach Publikationsjahren war nirgends vorgesehen. Die British Library bot ihren Katalog zu diesem Zeitpunkt neuartig auf CD an – in Augsburgs UB für mich soweit verfügbar, dass ich mir dort ausgewählte Titel auf 3,5 Zoll

Disketten ziehen konnte. Man musste sie nur noch einzeln anklicken, und in Paketen von 200 Titeln auf Disketten schreiben. 11 Jahrgänge, 22.000 Titel, fünf Tage Kopiervorgänge auf 11 Disketten. Wochen später überraschte ich die Bibliothekare der British Library mit meiner Fähigkeit, selbst bei Zusammenbruch des internen Netzes von meinem Computer aus noch immer Signaturen ziehen zu können. Tatsächlich verschafften mir die Jahrgänge den interessantesten Blick auf die Kundennachfrage, auf Reaktionen des Marktes im engen Untersuchungszeitraum. Ich bestellte mich über ein halbes Jahr hinweg sukzessive durch die Jahrgänge mit der steten Frage, wie Trends im Jahrzehnt spürbar wurden. Indes hätte ein etwas offeneres Frontend mir die anfängliche Arbeit etwas erleichtern können.

Mit dem FactGrid, der von uns 2018 am Forschungszentrum Gotha in Betrieb genommenen Wikibase-Instanz, entfallen solche Miseren. Man kann die Instanz schnell mit beliebigen Daten bespielen; man kann Daten in der Konfiguration des jeweiligen Forschungsinteresses herunterladen. Man kann genauso gut täglich frisch die gesamte Plattform im Data-Dump beziehen, um mit ihren Daten in einer laufend aktualisierten externen Datenbank ganz eigene Fragen zu beantworten.

Wir spielen derzeit mit dem Vorhaben, alle Keilschrift-Artefakte der Antike, 360.000 Datenbankobjekte, über das FactGrid verfügbar zu machen,<sup>1</sup> alle Informationen des Repertorium Germanicum in Datentripeln,<sup>2</sup> alle Freimaurer-Logen des 18. und 19. Jahrhunderts.<sup>3</sup> Soeben läuft die gesamte Nomenklatura des ungarischen Machtapparats der sozialistischen Ära in die Maschine ein, Daten zu 7000 Personen, Dutzenden von Organisationen und Hierarchie-Ebenen.<sup>4</sup> Ein Projekt, auf das ich irgendwann gerne einmal zurückkommen würde, wäre es, die europäische Buchproduktion der frühen Neuzeit in ihrer zunehmenden Kopplung an das Rezensionswesen neu zu denken. So klar ist nicht, wo die technischen Grenzen liegen. Das Wikidata-Projekt, dessen Software wir nutzen, agiert derzeit mit über 110 Millionen Datenbankobjekten. Wir stehen gerade bei 435.000 Datenbankobjekten bei einer anstrengenden jährlichen Verdoppelungsrate.

Über das FactGrid auf einer Tagung zum Frontend als Flaschenhals zu sprechen, war reizvoll. Die gesamte Plattform ist eher eine Teeschale (die eigentliche Frage ist, ob die Forschung eine derart offene Ressource wirklich sucht).

---

<sup>1</sup> Siehe Adam Anderson, *The FactGrid Cuneiform Project: A Wikibase for Cuneiform Sources*. FactGrid Blog 2022-05-12. <https://blog.factgrid.de/archives/2833>.

<sup>2</sup> Siehe die diesbezügliche Projekt-Repräsentanz im FactGrid Wiki: [https://database.factgrid.de/wiki/FactGrid:Middle\\_Ages](https://database.factgrid.de/wiki/FactGrid:Middle_Ages).

<sup>3</sup> Siehe die Projekt-Repräsentanz [https://database.factgrid.de/wiki/FactGrid:The\\_Global\\_Genealogy\\_of\\_Lodges](https://database.factgrid.de/wiki/FactGrid:The_Global_Genealogy_of_Lodges).

<sup>4</sup> Das von László Kiss initiierte Projekt „Network of the political and administrative elite of Hungary's communist era“ [https://database.factgrid.de/wiki/FactGrid:Network\\_of\\_the\\_political\\_and\\_administrative\\_elite\\_of\\_Hungary%27s\\_comunist\\_era](https://database.factgrid.de/wiki/FactGrid:Network_of_the_political_and_administrative_elite_of_Hungary%27s_comunist_era)

Über das FactGrid in einem Tagungsband zu schreiben, ist weit schwieriger. 2020 gaben mir die Herausgeber des Bandes den Tipp, doch vielleicht über die Frage nachzudenken, warum das FactGrid so schnell keine Mittelalter-Sektion aufbieten werde. Das schien 2020 noch beantwortbar: Graph-Datenbanken zu bespielen, ist extrem kompliziert. Hier werden bestehende Objekte miteinander verknüpft – die laufend zuvor vorausschauend und arbeitsökonomisch angelegt werden müssen. Wir standen damals bei etwas über 100.000 Objekten mit einem Focus auf Freimaurerforschung im ausgehenden 18. Jahrhundert. Unser Wachstum geschah fast ausnahmslos an den Rändern durch Projekte, die immer bereits genügend zu verknüpfende Objekte bei uns vorfanden, um hier aus ihrer jeweiligen Perspektive arbeitssparend agieren zu können.

Zwei Jahre später ist klar, dass wir einen Mittelalterschwerpunkt bekommen werden, der im Datenvolumen mit der gesamten bestehenden Ressource mithalten wird können: Das PhiloBiblon-Projekt mit seinem Team aus Berkeley und Barcelona arbeitet seit anderthalb Jahren an einem Import, mit dem man 450.000 Datenbankobjekte zu iberischen Handschriften des Mittelalters generieren wird.<sup>5</sup>

Einige Überlegungen dazu, was die Software, Wikiase – ihr verdanken wir unser massives Wachstum in erster Linie – derzeit so attraktiv macht; zweitens eine Erwägung zu Projekten, die wir derzeit erfolgreich ansprechen – sie weisen neuartige Gemeinsamkeiten auf, sind jedoch eher keine Standardprojekte. Drittens Der Versuch eines Ausblicks – wir stehen derzeit von Etablierungsprozessen, deren Konsequenzen derzeit nicht so leicht zu überschauen sind.

## 1. Wikidata – die globale Graph-Datenbank und ihre Software

### *Eine Graph Datenbank zwischen allen Wikipedien*

Als 2010 erste Diskussionswellen durch das „Wikiversum“ liefen, ausgelöst von Erwägungen der Wikimedia-Foundation, eine Datenbank-Software und eine zentrale Datenressource im Zentrum aller Wikipedia-Projekte zu entwickeln, war Beunruhigung spürbar: Geplant wurde eine Instanz, die harte Fakten in alle Wikipedien einspielen würde. Das Gewicht des Mondes, die Höhe des Mount Everest – einmal definiert, identisch verfügbar in allen Sprachen und Wikipedien der Welt. In den Wikipedia-Communities machten sich Sorgen breit über die kommende Entmachtung, über die zentrale Steuerung, das Ende der Menschen hinter allen Wikipedien, die bislang Quellen recherchierten und aktuelle Daten eigenhändig notierten. Kein Zufall sollte es sein, dass dieses Projekt von Google unterstützt

---

<sup>5</sup> Siehe Charles Faulhaber, „PhiloBiblon: From Siloed Databases to Linked Open Data via Wikibase: Proof of Concept Wikiversum“, FactGrid Projekt-Blog 2021-05-30, <https://blog.factgrid.de/archives/2243>, mit ausführlicher laufender spanischsprachiger Berichterstattung im Blog der Bibliothek der Universität Berkeley: <https://update.lib.berkeley.edu/2021/05/23/pbwiki/>.

werden sollte – Wikipedia schien auf dem gefährlichen Weg der Kommerzialisierung. Es drohte der Sieg des Trägervereins über die Communities, die ihn großgemacht hatten.

Die Sorgen lösten sich überraschend schnell auf. Das Wikidata-Projekt ist seit Jahren das innovativste Projekt des ganzen Felds der Wikimedia-Projekte. Wer eine WikidataCon besuchte, traf ein neues globales Gefüge an Gruppen und Einzelnen, die an derselben Ressource arbeiten – in Wikidata in der Lage, Daten in jeder in seiner Sprache einspeist, global lesbar zu machen.

Man kann Wikidata-Daten in Wikipedia-Artikel einspielen, das geschieht derzeit in den Info-Boxen. Man kann aus Wikidata ganze Artikel-Substitute generieren – es sind dies gar nicht uninteressante Substitute auf neuen externen Plattformen,<sup>6</sup> sie haben eine ganz andere Leistungsfähigkeit als von Menschen verfasste Artikel, sobald es darum geht, große Informationsmengen übersichtlich ohne große Textdarlegung, verfügbar zu machen. Wikidata erweist sich mitten in dieser internen Arbeit zunehmend als Brückenkopf zwischen Datenbanken weltweit.

### *Wikidata Q-Nummern: Normdatenbanken ohne Normierungsprozess*

Die Arbeit an Wikidata begann unter Federführung des in Berlin situierten Entwicklerteams 2012. Das erste Ziel des Teams war es, die Binnenverlinkung zwischen den Wikipedia-Projekten (die Arbeit der bisherigen „Interwiki-Links“<sup>7</sup>) zu übernehmen, mit denen man aus der Wikipedia einer beliebigen Sprache vom jeweils geöffneten Artikel aus in die Parallel-Artikel der anderen Wikipedien gelangt. Hier stand eine von allen Nutzern seit 2002 geschaffene Verweisstruktur übernahmefertig zur Verfügung. Bei der Vernetzung von Personen-Artikeln hatten sich die Communities auf vergleichsweise sicherem Terrain bewegt. Unproblematisch waren auch universelle Wissensgegenstände wie Vater, Mutter, Haus, Hund und Auto vernetzt. Schwieriger sah es aus, wo Sprach-Communities ihre Wissensfelder unterschiedlich organisierten: Die deutschsprachigen „Wissenschaften“ sind auf den ersten Blick anglophone „Sciences“, doch fallen unter die „Sciences“ vor allem dabei die deutschen „Naturwissenschaften“ und en passant noch die „Sozialwissenschaften“. Deutsche „Geisteswissenschaften“ sind anglophone „Humanities“ – und hier wird es für den deutschen „Literaturwissenschaftler“ dann unbequem. Auf Englisch betreibt er „Literary Criticism“ – „Literaturkritik“ ist im deutschen Sache der Zeitungen, keine Wissenschaft... Unterschiedliche Venn-Diagramme stehen einander hier

---

<sup>6</sup> Den Anfang setzte hier Magnus Manske in einem Versuch, ganze Artikel-Substitute aus Wikidata zu generieren. Das Musterbeispiel ist hier die *Reasonator*-Seite zu Johann Sebastian Bach, Wikidata <https://www.wikidata.org/wiki/Q1339> im Reasonator die Seite: <https://reasonator.toolforge.org/?q=1339>. Von Markus Krötzsch stammt alternativ der SQID-Browser, die Bach-Vergleichs-Seite ist hier: <https://tools.wmflabs.org/sqid/#/view?id=Q1339>.

<sup>7</sup> Siehe zu den Interwiki Links die entsprechende Sektion im Hilfs-Bereich der deutschen Wikipedia (die Versionsgeschichte erlaubt es nachzuvollziehen, wie Wikidata hier eingriff): <https://de.wikipedia.org/wiki/Hilfe:Interwiki-Links>.

gegenüber. Durch all diese Unwägbarkeiten hindurch hatten die Communities approximative, pragmatische Schneisen geschlagen. Wikidata erfasste die Links zwischen den am ehesten einander parallelen Artikeln aus gut 300 Wikipedien und generierte daraus mehrere Millionen „Items“, die nun für (weitgehend) dieselben Wissensgegenstände in 300 Sprachen standen.

In den großen Projekten hatten die Communities in Wikipedia-Artikeln unsichtbare Basisdaten hinterlegt, hinzugekommen waren Identifikationen in öffentlichen Datenbanken wie der GND, aus ihnen konnte die Wikidata-Instanz im zweiten Schritt Informationen auf die Items spielen und diese nach außen durch Identitätsbehauptungen vernetzen.

Das Ergebnis war eine Datenbank, die Millionen von Objekten voneinander trennt und in 300 Sprachen verfügbar machte – und die sehr schnell darüber nachdenken ließ, ob man bei allem Streben nach universell akzeptierten Datenstandards nicht laufend auf das falsche Pferd gesetzt hatte – auf einen Einigungsprozess, statt auf die schlichte Übersetzbarkeit von Identitätsbehauptungen. Wikidata ist heute die Datenbank, in der Mitarbeiter der Normdatenbanken weltweit nachsehen, um zu erfassen, unter welcher Kennung ihr jeweiliger Gegenstand in anderen Datenbanken läuft. Man muss nun nicht mehr bei Namensgleichheiten unter dutzenden von Kenndaten Vergleiche anstellen. Man sucht den Wikipedia-Artikel zur jeweiligen Person, zum Thema zu jeweiligen Ort und gelangt über diesen zum Wikidata-Item, das alle Parallelreferenzen listet und das Themen so nuanciert erfasst, wie kein Schlagwortkatalog das auch nur anstrebte.

In der Funktionalität löst Wikidata überraschend lange gehegte Probleme auf; technisch erweist sich die Wikidata-Software dabei als Schritt in die Zukunft. Das Benutzerinterface lässt sich in beliebige Sprachen umschalten, sofort erscheinen die Daten auf den Datenseiten in diesen Sprachen (soweit Begrifflichkeiten dazu hinterlegt wurden).

Die aktuelle GND-Software erlaubt es, zu einem jeweiligen Gegenstand Sekundärliteratur im Komplex zu notieren. Wikidata erlaubt es, zu jedem einzelnen Aussagentripel Quellen anzugeben und diese einzeln zu bewerten – was eminent praktisch ist, da die Software zudem beliebige einander widersprechende Statements zu setzen erlaubt: zehn Geburtsdaten mit zehn verschiedenen Quellenangaben, die einzeln auf- und abgewertet werden können.

Jeder einzelne Editier-Zugriff wird versioniert. Beliebig viele Nutzer weltweit können die Instanz mit Masseneingaben bombardieren und sie protokolliert jede Veränderung Aussage um Aussage.

## *Der weltweite Run auf Wikibase*

Um das Jahr 2016 wurde unter den Normdatenanbietern klarer, was hier geschehen war. Wikidata machte nicht nur einen Vorschlag universelle Identität von Wissensgegenständen und Beziehungen zwischen ihnen herzustellen und dabei Datenbanken aufeinander zu verweisen. Die ganze Modellierung der Daten, die hier vorgelegt wurde war neuartig flexibel und verwirrend modelllos.

Ganz anders als die bisherigen Datenbanken war die neue, kreuz und quer vernetzte, dabei durchsuchbar – in der ganzen planlosen Beliebigkeit von Beziehungen, die sie auftrat, überraschend nah am Verständnis menschlichen Wissens. Man kann beliebige Fragen an diese Datenbank stellen – eine Schablone dazu gibt es nicht, man muss die Fragen auf die Vernetzung ausgerichtet formulieren. Gleichzeitig gibt es dabei gar keinen Grund, warum Fragen dieser Art nicht quer durch Instanzen dieses Typs in Netze von Instanzen gestellt werden sollten. Man muss wissen, was dieselben Objekte und Beziehungen zwischen ihnen in Datenbanken weltweit sind – und Wikidata sammelt genau dieses Wissen. Eine Landschaft föderierter Instanzen derselben Software würde globale Recherchen erlauben – und Daten neuartig verfügbar machen: man kann solche Suchen sofort herunterladen und in traditionellen Umgebungen wie Excel untersuchen. Es ist dies das Projekt das seit 2017 unter dem Stichwort „Federated Wikibase Instances“ diskutiert wird.<sup>8</sup> Seit Ende 2018 steht fest, dass die deutsche Nationalbibliothek, die BNF und die Schwedische Königlichen Bibliothek Wikibase-Instanzen aufbauen werden, um ihre Daten global kompatibel verfügbar zu machen und eine vernetzte Struktur von Ressourcen mit Wikidata als globalem Brückenkopf und Lieferanten der Technologie aufzubauen.<sup>9</sup>

## 2. Wie schafft Wikibase das: beliebige Information beliebig mehrsprachig verfügbar machen?

Wikibase Daten sind verblüffend einfach strukturiert – einen Vorgeschmack davon erhält man, wenn man eine Instanz anlegt (heute geht das im Download eines Docker-Image sehr viel leichter als 2017<sup>10</sup>). Man erhält ein leeres Gefüge, das sich sofort wie eine Wikipedia-

---

<sup>8</sup> Ein erster internationaler Workshop zum Thema Federated Wikibase Instances fand im April 2017 statt, siehe meinen Beitrag „2018-04-23/25, Antwerp — the first Federated-Wikibase-Workshop“, FactGrid Blog 2018-05-03, <https://blog.factgrid.de/archives/835>. Das Thema gewann breite Bedeutung im deutschen NFDI-Prozess und wurde hier am 23 Februar 2021 Thema eines Erste Wikibase Workshops: <https://nfdi4culture.de/de/aktuelles/nachrichten/wikibase-workshop.html>. Siehe hierzu auch den Wikidata:Development plan auf [https://www.wikidata.org/wiki/Wikidata:Development\\_plan#Wikibase\\_ecosystem](https://www.wikidata.org/wiki/Wikidata:Development_plan#Wikibase_ecosystem) und die auf das Praktische ausgerichtete Seite: <https://www.mediawiki.org/wiki/Wikibase/Federation>.

<sup>9</sup> Siehe hierzu meinen Artikel im FactGrid Blog, 2018-12-27 <https://blog.factgrid.de/archives/1243>, nachgedruckt unter dem Titel „GNDCon 2018. Ein Nachklapp aus Forscherperspektive“, *info* 7, 33/3 (2018), S. 43-46.

<sup>10</sup> Siehe die Installations-Seiten <https://www.mediawiki.org/wiki/Wikibase/Installation> und <https://www.mediawiki.org/wiki/Wikibase/Docker>.

Instanz bedienen lässt. Zur Möglichkeit, Internetseiten anzulegen und zu verlinken, kommt nun jedoch das Angebot „Items“ anzulegen und diese mittels „Properties“ zu vernetzen. Es gibt dazu keine Eingabeschablonen, nichts, was auch nur darüber aufklärte, was „Items“ und „Properties“, wenn man sie denn anlegt, überhaupt leisten sollen. Man generiert ein Item unter einem Titel und soll eine Aussage formulieren und das geht nicht eher, als bis man eine erste Property genau dazu definiert hat.

Ebenso leer sind die Seiten, mit denen man Suchen und mittels derer man Masseneingaben durchführen soll. Sie warten auf Eingaben, die man irgendwie frei gestalten muss.

Tatsächlich, lassen sich in einer Wikibase-Instanz beliebige Datenbankobjekte zu beliebigen Wissensgegenständen anlegen. Sie werden unter Q-Nummern hochgezählt. Ihnen stehen P-Nummern gegenüber, mit denen sich Eigenschaften von Objekten, Beziehungen zwischen ihnen deklarieren lassen.

Ein erstes Objekt könnte Johann Sebastian Bach sein – Q1 in der leeren Instanz. Um eine erste Aussage zu setzen, etwa die seiner ersten Ehe mit Maria Barbara Bach, brauchen wir das Aussageformat das auf den Ehepartner verweist, P1, und ein Objekt für Maria Barbara, die erste Ehefrau: Q2. Mit den drei Bausteinen lässt sich das erste Datentripel generieren. In welcher Sprache und mit welchem Wortlaut, ist zweitrangig

Q1 – P1 – Q2

Johann Sebastian Bach – war verheiratet mit – Maria Barbara Bach  
Иоганн Себастьян Бах – был женат на – Мария Барбара Бах  
Johann Sebastian Bach – was married to – Maria Barbara Bach  
Johann Sebastian Bach – était marié à – Maria Barbara Bach

Dass man hier tatsächlich nicht Textfelder befüllt, sondern Objekte, Items, miteinander verknüpft, wird sichtbar, sobald ein Item unter dem gesuchten Namen („Label“) noch nicht zur Verfügung steht. Das Eingabefeld nimmt die Eingabe nicht eher an, als bis das Objekt angelegt ist. Findet die Instanz mehrere Items desselben Namens, so liefert sie ein Pulldown-Menü samt Kurzbeschreibungen, die man zu jedem Objekt mitliefern sollte.

Eher im Ausnahmefall werden Properties so definiert, dass sie auf ein Textfeld münden (bei uns etwa die Property P5, Titelblatt-Transkript), auf ein Datum, eine Mengenangabe, eine Geokoordinate oder eine URL (respektive einen variable URL-Bestandteil).

### *Datentripel, Qualifikatoren und Quellenangaben: Das Wikibase-Datenmodell*

Jede Wikibase-Instanz wird, wenn man sich nicht an anderen orientiert, mit ihren eigenen Objekten und auf sie treffenden Formulierungen unter eigenständigen P- und Q-Nummern befüllt. Eine neue Instanz kann etwa Mitgliederlisten erfassen und dazu mit Items zu Organisationen und Personen hantieren; eine andere kann unter denselben Q-Nummern chemische Elemente und Moleküle generieren mit Properties zu deren Qualitäten und Beziehungen zueinander.

Die Aussagen sind in ihrer Grundform „Datentripel“ deren erster Bestandteil jeweils das Objekt ist, dem eine Eigenschaft oder Beziehung zugeschrieben wird, die Beziehungsaussage und das Bezugsobjekt folgen. Die Aussagen lassen sich danach ihrerseits zum Gegenstand von Aussagen, erneut Datentripeln machen. Dabei gibt es zwei Sorten von Aussagen über Aussagen: die Qualifizierenden Aussagen, sie sind unmittelbar unter dem jeweiligen Tripel sichtbar, und die Quellen-Aussagen, die auf Anklicken hin erscheinen. Das Arrangement ist sinnvoll, da sich mit ihm Bezüge klar setzen lassen – etwa die zum Zeitpunkt des Eheschlusses, bei dem Mann im Falle Bachs mit seinen zwei Ehen, jedem Eheschluss ein eigenes Datum zuweisen muss:

(Q <sub>1</sub> – P <sub>1</sub> – Q <sub>2</sub> )	
—————	P <sub>2</sub> – Q <sub>3</sub>
—————	P <sub>3</sub> – 17. Oktober 1707

Johann Sebastian Bach (Q<sub>1</sub>) – war verheiratet mit (P<sub>1</sub>) – Maria Barbara Bach (Q<sub>2</sub>),  
 Ort des Eheschlusses war (P<sub>2</sub>) Dornheim bei Arnstadt (Q<sub>3</sub>), Datum des Eheschlusses  
 war der 17. Oktober 1707

Die Angabe von „Fundstellen“ funktioniert nach demselben Prinzip. Qualifizierbare Datentripel sind einerseits elementar einfach, sie erlauben andererseits die Abbildung komplexer sprachlicher Aussagen mit eigenem Temporal und Bezugssystemen abzubilden.

### *Neuartig offene Datenbankanfragen*

Relationale Datenbanken bauen Tabellen auf und verweisen aus ihnen heraus spaltenweise auf Tabellen, die die weitere Information zu Objekten spezifischer Spalten bieten – etwa dann, wenn es zu einer Mitgliedertabelle eine Ortstabelle gibt, in der mehr zu den Geburts- und Wohnorten der Mitglieder gesagt ist. Wenn solche Tabellengefüge wachsen sollen, oder wenn neue Suchen in ihnen durchgeführt werden sollen, stehen komplexe Arrangements an, die zumeist nur noch die Datenbankprogrammierer bewerkstelligen können.

Graph-Datenbanken sind eher wie menschliches Wissen gestaltet. In ihnen werden Wissensgegenstände gebaut und aufeinander bezogen. Das ist insbesondere dann praktisch, wenn sich zu den Gegenständen vollkommen unterschiedliche, auch unterschiedlich viele Aussagen machen lassen – etwa, da ihnen ganz unterschiedliche Quellen zugrunde liegen. Es gibt keine Schablonen mit Hunderten von Eingabefeldern (so viele bräuchten wir mittlerweile bei den Aussagen die im FactGrid allein zu Personen und Dokumenten möglich sind. Notiert werden schlicht die Aussagen, die Verbindungen zu anderen Objekten oder Daten oder Werten, die gemacht werden sollen.

Die Informationsabfrage muss sich dem anpassen: Man kann Eigenschaften beliebiger Objekte abfragen, oder erfassen, welche Objekte in einer bestimmten Art mit derselben Eigenschaft verbunden sind – etwa, weil sie diese Eigenschaft aufweisen, oder weil sie mit einem Objekt verbunden sind, dass die gesuchte Eigenschaft aufweisen soll.



Der Illuminatenorden hat im FactGrid die Item-Nummer Q10677. Die Datenbank lässt sich nun befragen, welche Mitglieder er hatte – bei ihnen handelt es sich um Items, von denen ein P<sub>91</sub>-Link (Mitglied von) auf Q10677 verweist. Das Ergebnis ist eine Liste aller Mitglieder, zu denen sich nun weitere Fragen etwa zu, P<sub>77</sub>, ihren Geburtsdaten oder, P<sub>82</sub>, ihren Geburtsorten stellen lassen. Die Abfrage wird im QueryService über ein Skript verfasst; unter dem i-Symbol ist jedoch ein Abfragehelfer verborgen, der Eingabefelder mit Worten zu befüllen erlaubt. Das Ergebnis ist eine Tabelle, die sich sofort in verschiedenen Formaten herunterladen lässt:

The screenshot shows the FactGrid Query interface. On the left, there's a sidebar with filters for 'Mitglied von' (set to 'Illuminatenorden') and 'Anzeigen' (set to 'Geburtsdatum', 'Geburtsort', 'Sterbedatum', 'Sterbeort'). The main area displays a SPARQL query:

```
1 SELECT ?Illuminatenorden ?IlluminatenordenLabel ?Geburtsdatum ?Geburtsort ?GeburtsortLabel ?Sterbedatum ?Sterbeort ?SterbeortLabel WHERE {
2   SERVICE wikibase:label { bd:serviceParam wikibase:language "[AUTO_LANGUAGE],en". }
3   ?Illuminatenorden wdt:P91 wd:Q10677.
4   OPTIONAL { ?Illuminatenorden wdt:P77 ?Geburtsdatum. }
5   OPTIONAL { ?Illuminatenorden wdt:P82 ?Geburtsort. }
6   OPTIONAL { ?Illuminatenorden wdt:P88 ?Sterbedatum. }
7   OPTIONAL { ?Illuminatenorden wdt:P168 ?Sterbeort. }
8 }
```

Below the query, a table titled 'Vorschau des Ergebnisses' shows 20 results in 73 ms. The table has 8 columns: Illuminatenorden, IlluminatenordenLabel, Geburtsdatum, Geburtsort, GeburtsortLabel, Sterbedatum, Sterbeort, and SterbeortLabel.

Illuminatenorden	IlluminatenordenLabel	Geburtsdatum	Geburtsort	GeburtsortLabel	Sterbedatum	Sterbeort	SterbeortLabel
<a href="#">Q237</a>	Justus Johann Dehnhard						
<a href="#">Q459</a>	Hugo Franz Graf von Hatzfeldt-Wildenburg	1. Januar 1755	<a href="#">Q10314</a>	Bonn	1. Januar 1833	<a href="#">Q10312</a>	Berlin
<a href="#">Q460</a>	Johann Joseph Carl Hauk	1. Januar 1743	<a href="#">Q220995</a>	Weinburg (Ungarn)	1. Januar 1819		
<a href="#">Q461</a>	Johann Ernst Christian Haun	21. Juni 1748	<a href="#">Q34639</a>	Siebleben	22. März 1801	<a href="#">Q10279</a>	Gotha
<a href="#">Q461</a>	Johann Ernst Christian Haun	21. Juni 1748	<a href="#">Q43543</a>	Gräfenonna	22. März 1801	<a href="#">Q10279</a>	Gotha
<a href="#">Q462</a>	N.N. Hauptmüller						
<a href="#">Q463</a>	Carl Heinrich Freiherr von Hayn	1. Januar 1738	<a href="#">Q220998</a>	Rodersdorf (Weischütz)	1. Januar 1813	<a href="#">Q10288</a>	Neuwied

Das raffinierte an der SPARQL-Abfrage ist, dass vollkommen offen ist, zu welchen Objekten und Beziehungen zwischen ihnen hier Suchanfragen abgeschickt werden. Offen ist nicht minder, wie verwinkelt diese Abfragen in die Vernetzung der Information hineingehen: Wir haben alle 4200 Absolventen des Gothaer Gymnasiums Illustre erfasst und können nach ihren weiteren Karrieren (P<sub>165</sub>) fragen. Wir können nach ihren Vätern (P<sub>141</sub>) fragen und danach, wo, P<sub>160</sub> diese Ausbildungen machten und was, P<sub>165</sub>, für Karrieren sie machten. Wir können bei den Müttern, P<sub>142</sub>, nach deren Vätern, wieder P<sub>141</sub>, fragen und abfragen was für Ausbildungen, P<sub>160</sub>, und Karrieren, P<sub>165</sub> sie machen. Das Ergebnis wird eine Tabelle sein, in der Zeile sich Zeile um Zeile ablesen lässt, in wieweit Die Karrieren der Familienbeziehungen die der Kinder bestimmten – und ab wann weit eher die Schule Karrieren vorstrukturierte.

Der Tiefe der Abfragen ist an dieser Stelle keine Grenze. Aussagen sind jederzeit mit Zeit- und Ortsangaben verbunden, bei den Orten sind Geokoordinaten hinterlegt, so dass sich umstandslos zwischen Landkarte und Zeitleisten hin und her schalten lässt. Man kann Befunde gruppieren, auszählen und statistisch sortieren, in Balkendiagrammen oder in Linien im Zeitverlauf – das erfordert zum Teil komplexe Skriptkenntnisse, doch erfordert es keine Programmierer, die die Datenbank auf die Suche hin konfigurieren. In der Praxis arbeitet man mit „Mustersuchen“, bei denen man P- und Q-Nummern austauscht. Im

Internet gibt es SPARQL-Handbücher und Communities, die Ideen haben, wie man eine bestimmte Suche „scriptet“.

### *Rohe Software ohne breitentaugliches User-Frontend*

Der extrem komfortable Leistungskatalog: Wikibase-Instanzen können beliebige Tabellen in sich aufnehmen und deren Information in beliebigen Sprachen simultan lesbar machen, sie öffnen die Daten dabei einer beliebigen Durchsuchbarkeit, und sie können dabei von beliebig vielen Nutzern simultan befüllt werden – alle Editiervorgänge werden versioniert festgehalten... – dieser extrem komfortable Leistungskatalog darf nicht darüber hinwegtäuschen, dass die Software optisch und in der Handhabung eher hemdsärmelig daher kommt. Ein Nutzer, der versuchte, Musterabfragen im eigenen Computer zum Laufen zu bringen, kontaktierte uns, da das System bei ihm immer abstürze, wenn er die Anfrage anklickte. Ich gab ihm im Share-Screen den Blick auf meinen Bildschirm und es stellte sich heraus, dass er gar nichts anderes gesehen hatte als ich soeben aufrief, aber das Suchskript des QueryService dann als unbekannte Fehlermeldung gelesen habe, den blauen Knopf zum Abschicken der Suche gar nicht erst betätigt hatte.

Wikibase-Instanzen rechnen mit Projekten, die sich eigene Frontends herstellen, spezifische Frontends, mit denen sich ganz bestimmte Materialbereiche einer großen Instanz einkreisen und auswerten lassen. Man kann von diesen aus Anfragen so vorgestalten, dass die Nutzer mit SPARQL als möglicher Abfragesprache gar nicht in Berührung kommen.

Was als Mangel erscheinen mag, sollte im selben Moment als bewusste Designentscheidung gewertet werden. Wikibase meidet jede einschränkende Oberfläche, jedes einengende Frontend, jede Fixierung der Software. Man bietet hier gezielt den stabilen Kern an, das Objekt, mit dem Communities Masseneingaben und Massenrecherchen bewältigen können, und von dem aus sie ihren eigenen Nutzern mit viel dezidierteren Angeboten gegenüberreten können.

Schien es uns anfänglich noch so, als würden wir erste Browser für das FactGrid erst mit eigenen Software-Integrationen in größere Konsortien gewinnen können, so ist heute klarer, dass Projekte an Bord sich solche Browser bauen und anderen Projekten zur Verfügung stellen. Technisch gibt es dazu verschiedene Optionen von Browsern wie dem von Bruno Belhoste vorgelegten FactGrid Viewer,<sup>11</sup> der ohne Zwischenabbild und ohne Suchabfrage Informationen aus den Datenbankseiten und ihrer Vernetzung generiert, bis

---

<sup>11</sup> Die URL ist hier <https://database.factgrid.de/viewer/>. Technische Informationen gab Bruno Belhoste in „Browsing FactGrid with the FactGrid Viewer“, FactGrid Projekt-Blog, 2022-04-06, <https://blog.factgrid.de/archives/2684>.

zu Michael Ringgaards KnolBase Browser,<sup>12</sup> der aus den täglichen Datenabbildern, die er von Wikidata und aus dem FactGrid bezieht neue Informationsseiten generiert.

Die aus Wikibase gewonnenen Informationen sollten dabei nicht unterschätzt werden. Man kann aus tripelbasierter Information Seiten generieren, die mit Wikipedia-Artikeln mithalten, schlicht, weil sie ihre Informationen weit übersichtlicher gegliedert in Tabellen mit beliebiger Vollständigkeit anbieten können als jeder Artikel dies tun kann, angereichert mit Visualisierungen<sup>13</sup> und Abbildungen<sup>14</sup> sowie Textintegrationen,<sup>15</sup> die jeweils nach dem aktuellen Informationsstand aus der Ressource generiert werden.

### 3. Wikibase-Instanzen als Forschungswerkzeuge nutzen

Wikimedia-Software hat sich zwar schrittweise darauf ausgerichtet, wissenschaftliche Fakten referieren zu können – es gab in den Wikipedien anfangs keine Möglichkeit, auch nur Fußnoten zu setzen; an der wissenschaftlichen Nutzung bleibt Wikimedia-Software jedoch bis in Wikibase hinein vorbeikonzipiert. Es handelt sich hier um eine ausgeprägte Community Software – einerseits vollkommen transparent, andererseits jedoch so gestaltet, dass die Öffentlichkeit an dieser Transparenz gezielt vorbeigeführt wird. In Wikimedia-Projekten etabliert sich Autorschaft intern auf Seiten, auf denen die Community sich in ausgesprochen intern und entsprechend neuartig unabhängig organisiert. Der interne Austausch ist konsequenterweise jederzeit gereizt. Alles, was hier Autoren untereinander auf Nutzerseiten, Diskussionsseiten und Projektseiten besprechen, ist vor der Weltöffentlichkeit gesagt, aber in aller Regel unter Pseudonymen gesagt, die jederzeit außerhalb nachvollziehbarer Verantwortlichkeit agieren.

---

<sup>12</sup> Siehe für mehr Informationen Michael Ringgaards Startseite <https://ringgaard.com/> (abgerufen am 29. 8. 2022). Suchen in der FactGrid-Instanz laufen hier unter der Subdomain: <https://factgrid.ringgaard.com/kb>.

<sup>13</sup> Visualisierungen des SPARQL Query Service lassen sich nicht nur in verschiedenen Formaten herunterladen, sondern auch als fertig formuliertes I-Frame Fragment, das sich in Websites und Blogs integrieren lässt wie in Bruno Belhostes FactGrid-Viewer.

<sup>14</sup> Bilddateien lassen sich ohne weitere Hürde von Wikimedia Commons <https://commons.wikimedia.org/>, dem zentralen Wikimedia Medien-Repository einspielen. Es sieht im Moment danach aus, dass wir langfristig ein eigenes paralleles Repository aufsetzen, um mit ihm insbesondere Digitalisate zu handhaben, die Projekte so einfach nicht in Wikimedia Commons hochladen wollen.

<sup>15</sup> Wikibase ist nicht geschaffen, um Text in größeren Mengen zu integrieren. Zwar lässt sich das Zeichenlimit in Properties, die Text aufnehmen sollen, hochstellen (es liegt im FactGrid derzeit bei 1500 Zeichen), doch bleiben Texteingaben in den Datenbankfeldern ohne weiteres Markup –weder lassen sich hier Schrifttypen variieren, noch lassen sich hier Fußnoten setzen. Die elegante Lösung ist es, für Textpassagen auf das Wiki auszuweichen, das mit der Installation einer Wikibase-Instanz kommt. Wir tun dies zum Beispiel mit der FactGrid Property P251, die auf „Dokumenten“-Seiten mit Transkriptionen verlinkt. Bruno Belhostes FactGrid Viewer integriert die Texte, die dort mit vollem Markup mitsamt Fußnoten generiert werden, nahtlos auf Seiten, die aus Datenbankinformationen generiert werden. Siehe zur Demonstration <https://database.factgrid.de/viewer/item/Q6641> - eine Seite die aus der Item-Seite <https://database.factgrid.de/wiki/item:Q6641> und der Transkriptseite <https://database.factgrid.de/wiki/D-Q6641> amalgamiert wird.

Die Software ist, was diese Herstellung interner Autorschaft anbetrifft suggestiv: Wissenschaftler, die überall auf einem Austausch unter Klarnamen bestehen, melden sich im FactGrid, sobald sie das Antragsformular nutzen (statt per E-Mail um die Eröffnung eines Kontos zu bitten), nahezu grundsätzlich mit einem sich anonymisierenden Kürzel an – die Schablone gibt ihnen suggestiv die Chance, den Nutzernamen anonym und die Identität intern den Administratoren bekannt zu notieren. Wir verweisen in diesen Fällen auf die Nutzungsbedingungen,<sup>16</sup> unter denen auf der Plattform alle unter Klarnamen agieren. Es ist dies die Stelle, an der wissenschaftliche Antragsteller sich auf den Klarnamen einlassen und die Institutsadresse als Kontaktadresse nutzen. Das Internet ist erst einmal nicht der Ort, an dem es Nutzern naheliegt, offen mit der eigenen Person, wie mit dem Forschungsvorhaben und den eigenen Befunden zu agieren. Wir müssen hier auf der Wikimedia-Plattform gegen den Strom schwimmen – gewinnen jedoch im selben Moment einen extrem kooperativen Austausch auf der Plattform.

### *Das FactGrid als Wikibase-Instanz für Original Research*

Wir unterlaufen mit der eigenen Instanz Wikimedia-Usancen: Die Instanz ist zwar offen sichtbar, editieren lässt sie sich jedoch nur durch Konteninhaber, die Konten teilen die Administratoren nach Rücksprache mit den Interessenten zu. Wir gewinnen im selben Moment erhebliche Freiheiten gegenüber den Wikimedia-Plattformen, die sich, in der Teilnahme offener – jeder kann hier schreiben, mit oder ohne Konto – in den Binnenregularien entschieden restriktiver gerieren. Die globale Nutzerschaft ist von den Wikipedien eingeladen, das Wissen der Welt verfügbar zu machen, jeder Beiträger wird dabei jedoch zugleich bis an den Punkt entmachtet, an dem er oder sie nur noch sagen kann, was privilegierte Autoren, die Administratoren, die jederzeit „Wissen“ löschen können, in allgemein verfügbaren Publikationen als allgemein verfügbares wissenschaftlich haltbares Wissen nachweisen können:

1. Die „no original research“<sup>17</sup> Regel – sie schließt jede Erstveröffentlichung von Information auf den Wikimedia-Plattformen aus. Alle in Wikipedia-Artikeln angebotene Information sollte in wissenschaftlich anerkannten Quellen externen publiziert vorliegen.
2. Die Einigung auf „Relevanzkriterien“<sup>18</sup> gibt vor, welcher Art Artikel mit welcher Informationstiefe eröffnet werden können. Die *Relevanz* muss erneut extern notabel sein. Das öffentlich nachweisbare Informationsinteresse balanciert damit das öffentliche Informationsangebot aus und hält „Selbstdarsteller“ aus dem Medium.

---

<sup>16</sup> Die FactGrid Nutzungsbedingungen finden sich mehrsprachig auf der über die Startseite verlinkten URL <https://database.factgrid.de/wiki/FactGrid:Nutzungsbedingungen>.

<sup>17</sup> Siehe die Wikipedia-interne Seite [https://en.wikipedia.org/wiki/Wikipedia:No\\_original\\_research](https://en.wikipedia.org/wiki/Wikipedia:No_original_research) mit den eingehenden Darlegungen zur Redaktionsrichtlinie. Die parallele deutsche Seite firmiert unter [https://de.wikipedia.org/wiki/Wikipedia:Keine\\_Theoriefindung](https://de.wikipedia.org/wiki/Wikipedia:Keine_Theoriefindung).

<sup>18</sup> Im Deutschen gelten hier die Wikipedia-Relevanzkriterien – einer der am heftigsten umkämpften Bereiche in allen Wikipedia-Projekten: <https://de.wikipedia.org/wiki/Wikipedia:Relevanzkriterien>.

3. Die „neutral point of view“<sup>19</sup>-Regel – nach ihr werden in Wikipedia-Artikeln keine Autorenstimmen aktiv, die auch nur methodologische Entscheidungen rechtfertigen könnten – legt fest dass Artikel so geschrieben sein müssen, wie jeder andere sie genauso schreiben könnte.

Alle drei Regeln bescheren den großen Wikimedia-Projekten permanente interne Konfliktlinien: Das Ziel soll es sein, die Nutzer des Internets weltweit ohne Diskriminierung zu befähigen, zum kollektiven Wissen beizutragen. Gerade das Wissen unterprivilegierter Nutzergruppen kann dabei jedoch erst zum Zuge kommen, wenn es privilegiertes Wissen ist: Der Jäger, der über persönliche Erfahrung in der Herstellung von Steinzeitwerkzeugen berichten kann, kann von diesem Wissen nur genau notieren, wie er aus allgemein überprüfbar wissenschaftlich gedeckten Ressourcen im Zweifelsfall zitierbar wird. Die Wissenschaften verhindern im selben Verfahren als externe Instanz, dass die Wikipedia-Projekte von Selbstdarstellern, Verschwörungstheoretikern, politischen Parteien und Firmen gekapert werden.

Das Aparte an diesem Arrangement ist, dass die Wissenschaften im selben Moment in Wikipedia-Artikeln nichts von dem tun können, was außerhalb ihre Aufgabe ist. Sie sind darauf verpflichtet Original Research zu leisten, sie bieten an, Relevanzkriterien mit neuen Faktenlagen und Interpretationen zu verschieben, die auktoriale Präsentation schafft dabei öffentliche Verantwortung.

Das FactGrid ist an dieser Stelle vor allen Dingen eingerichtet, um Forschenden die Positionierung neuer Datenlagen und ihre erste Interpretation zu erlauben. Das Angebot erwies sich nicht nur in der Forschung attraktiv. Es erlaubte uns gerade den Wikimedia-Projekten wie den Nationalbibliotheken das Angebot zu unterbreiten, hier wissenschaftsintern möglich zu machen, was Instanzen beider Anbieter Wissenschaftlern extern verweigern müssen.

### *Wir holen Beiträge von den Versionsgeschichten in die Datensätze*

Ganz praktisch nutzen wir Wikibase anders als Wikidata, indem wir unseren Beiträgern nicht nur Nutzerkonten geben – unter Klarnamen und mit offener Kontaktinformation auf ihren Nutzerseiten –, sondern an diese gekoppelt auch Datenbankobjekte, „Items“. Mit ihnen können Nutzer in Aussagen auf den Item-Seiten auftreten in eigener Person – ob mit der Quellenangabe, dass sie selbst dieses Dokument einsahen und Messungen von Seitenlänge und breite etwa wie angegeben vornahmen – hier erstmals veröffentlicht. Sie können mit den Items der eigenen Person nicht minder als Autoren von Mutmaßungen und Arbeitshypothesen auftreten.

---

<sup>19</sup> Ausführlich hierzu die Seiten [https://en.wikipedia.org/wiki/Wikipedia:Neutral\\_point\\_of\\_view](https://en.wikipedia.org/wiki/Wikipedia:Neutral_point_of_view) für die englischsprachige und [https://de.wikipedia.org/wiki/Wikipedia:Neutraler\\_Standpunkt](https://de.wikipedia.org/wiki/Wikipedia:Neutraler_Standpunkt) für die deutschsprachige Wikipedia.

Zu den Optionen, die eigene Person als Akteur in Aussagen zu handhaben, kommen im FactGrid die Angebote, Projekte zu definieren und über eine eigene Property (P<sub>131</sub> „Forschungsprojekte, die zu diesem Datensatz beitrugen“) an Items zu binden. Mit der Notiz wird nicht die Bearbeitung derselben Items durch Kollegen auf der Plattform eingeschränkt – auf vielen Items notieren mehrere Projekte die gemeinsame Arbeit am selben Forschungsgegenstand. Es wird hiermit jedoch einfach, die Datenbank danach zu befragen, wo Arbeit, die ein bestimmter Geldgeber etwa finanzierte, Datenbeziehungen änderte. Über die Versionsgeschichten lässt sich im selben Moment an jedem Item erfassen, was einzelne Mitspieler im Detail taten.

Zu den Properties, die auf der Plattform Autorschaft, persönlich und im Projekt, generieren kommen weitere, mit denen sich thematische Felder abstecken lassen – sinnvoll etwa, wo Projekte einen eigenen Browser aufbauen, der gezielt Informationen aus einem markierten Forschungsfeld beziehen soll.

Wir gehen bislang mit diesen Optionen extrem konstruktiv um, sie entspannen die Interaktion zwischen Mitspielern, da sie zum einen mehr öffentliche Präsenz schaffen – der Raum tatsächlicher Interaktion vor der Öffentlichkeit wird hier näher an die persönlichen Identitäten der Nutzerseiten gebracht. Die offen gehandhabten Kontaktinformationen erlauben es den Mitspielern gleichzeitig jederzeit miteinander in Kontakt über andere Kanäle zu treten, per E-Mail oder in spontanen Videokonferenzen. Das Interaktionsrepertoire ist bei uns weiter und damit offener und entspannter als im Hintergrund der großen Wikipedia-Projekte.

## 4. Plattform interagierender Forschungsprojekte

*Die Software, die sich beliebig eng an jede historische Datenlänge anpassen lässt*

Das FactGrid Logo steht für das, was sich mit Wikibase neuartig unkompliziert und sehr viel leichter als mit klassischen relationalen Datenbanken herstellen lässt. Wir können jederzeit beliebige Gegenstände generieren und wir können zu ihnen jederzeit auch sehr eigenartige Aussagen treffen – eng an der Sprache der jeweiligen Datenlage entlang. Die Bläue des Himmels vor 200 Jahren können wir in ihren Nuancen wiedergeben, sobald ein Instrument wie Horace-Bénédict de Saussures Cyanometer auf diese Interaktion mit der zukünftigen Auswertung vorbereitete. Das Instrument ist im FactGrid das Objekt Q<sub>424999</sub>; die 52 Farbabstufungen, die auf das Weißfeld Q<sub>425000</sub> folgen, tragen die Nummern Q<sub>425001</sub> bis Q<sub>425052</sub> mit jeweiligen Hexadezimal-Aussagen zu den Farbwerten (in zwei Interpretationen der Datenlage). Mit der Property P<sub>777</sub> lassen sich damit beliebige Messungen, wie sie Alexander von Humboldt mit seinem Cyanometer ausführte, auf Landkarten bringen – warum nicht als Farbpunkte, mit denen Bergbesteigungen in Südamerika nachfahren lassen? Wo die Aussagen vorliegen, können wir sie, ohne andere Datenmodelle zu tangieren, aus dem Stand nachvollziehen.

Für Forschungsprojekte ist Wikibase im selben Moment ein Wagnis. Kann man sich auf ein Produkt einlassen, das von Wikimedia weiterentwickelt wird? Welchen Einfluss hat man auf diese Firma (die nebenbei keine Firma ist, sondern ein Verein, dem jeder beitreten kann)? Kann sich Forschung mit der Arbeit auf einer Community-Plattform arrangieren? Will man überhaupt, dass alles, woran man soeben arbeitet, auf einer solchen Plattform sofort für alle sichtbar wird? Will man, dass sich fremde Nutzer die eigenen Daten frei, wenn sie wollen, auch ohne Nennung der Urheberschaft herunterladen können?

Konkurrenzprodukt wie Nodegoat<sup>20</sup> sind in all diesen Fragen sehr viel geschickter auf traditionelle DH-Projekte zugeschnitten. Hier erhält man eine abgeschlossene persönliche Umgebung, aus der man Daten oder Visualisierungen selektiv online stellen kann. Wie die Daten einzugeben sind, ist weitgehend vorgegeben; sie erscheinen, wenn man alles wie vorgegeben macht, genau in den Visualisierungen, die derzeit en vogue sind – und das heißt: nicht auf OpenStreetMap Oberflächen. Das FactGrid bietet in all diesen Fragen einen massiven Kontroll-Verlust an. Allenfalls eine eigene Wikibase-Instanz könnte da die Alternative sein.

*Für die meisten Projekte im Moment schwer denkbar: die eigenen Daten mitten im Forschungsprozess offen handhaben*

Das FactGrid ist derzeit vor allem für sich breit aufstellende Projekte interessant, die erwägen, sich zwischen den großen Datenanbietern wie Wikidata und der GND, in einem wachsenden Kollektiv als forschungsgetriebener Mitspieler zu positionieren.

Für die meisten DH-Projekte ist diese Möglichkeit in ihrer eigenen Bindung an die Publikation, wie sie Druck perfekt geschieht, kaum denkbar. Mit dem gedruckten Buch wird seit den 1480er Jahren mit jeder Veröffentlichung neu ein Objekt vorgelegt, das in der Hand aller seiner Leser in allen soeben vorgelegten Exemplare der „Edition“, „Ausgabe“, „Publikation“, „Veröffentlichung“ – all diese Worte zelebrieren das disruptive Moment dieses Schritts aus dem Verborgenen an die Öffentlichkeit – Buchstabe für Buchstabe, Seite um Seite gleich ist.<sup>21</sup> Die war die Revolution des Drucks kam mit dieser Wende: Die neuen Publikationen erlaubten es lebenden Autoren, unvermittelt auf demselben Terrain aufzutreten, auf dem bislang nur Autoritäten notabel waren. Auf sie konnte man sich bislang wie auf die Bibel beziehen im Blick auf ein Werk, das in Bücher gegliedert war, mit Angaben von Kapitel, Versen oder Absätzen. Lebende Autoren waren in aller Regel

---

<sup>20</sup> Siehe das exquisit präsentierte Angebot der Nodegoat-Macher auf <https://nodegoat.net/>.

<sup>21</sup> Adrian Johns wartete in *The Nature of the Book: Print and Knowledge in the Making* (Chicago; London: Univ. of Chicago Press, 1998) mit der gegenteiligen Bemerkung auf: Die Autoren des Frühdrucks kämpften gegen Verleger, die keine Datenkonsistenz auch nur über eine Ausgabe hinweg in allen ihren Exemplaren herstellen konnten. Das ist korrekt beobachtet, doch andererseits gerade die Notiz, dass den neuen Wissenschaftlern nun etwas ganz Neues möglich erschien, was mit Handschriften bislang nicht möglich war: Stabile seitengleich identische und global zitierbare Publikationen herzustellen, wenn noch nicht mit Verlegern und Druckern der Zeit, dann doch vielleicht unter Wissenschaftlern, die selbst zu Verlegern und Druckern würden.

unzitierbar – ihre Werke hatten keine Titel, die Handschriften in denen sie kursierten boten den Text jeweils neu foliiert. Das gedruckte Exemplar erhielt zwischen 1470 und 1570 schrittweise einen eigenen Titel einen Publikationsort das Publikationsjahr und endlich Seitenzahlen, mit denen es überall, wo ein Exemplar vorlag, vollständige Referenzen auf minimalste Details, auf Satzfehler, auf Transkriptionsfehler erlaubte. Erst mit dem Internet ist es plötzlich wieder schwieriger geworden sich auf öffentlich verfügbare Information stabiler nachvollziehbar zu beziehen. Die Internet-Seiten wird bei jedem Besuch neu generiert, und kann zwischenzeitlich an ihrem jeweiligen Ausgangsort beliebig neu konfiguriert werden.

Die Digital Humanities distanzieren sich zwar vom Buch als ultimativem Medium. Der Angriff galt dabei bislang jedoch vor allem der linearen Lektüre zugunsten eines nun befreiend interaktiven Zugriffs. Das disruptive Moment der Veröffentlichung, an dem sich Diskussionen entzünden, versuchen die Digital Humanities gerade in die neue Medienwelt hinüberzuretten. Man sieht auf DH-Plattformen Datenpakete mit Publikationsdaten stabilisiert ausliegen. Bibliotheken übernehmen das Angebot dieser Stabilisierung mit der Zusicherung, dass niemand von ihnen verfügbar gemachte Daten unter der Hand verändern kann.

Wo das Moment der plötzlich vorliegenden neuen Informationslage in seinem Impact kostbar ist, muss jedes einzelne Projekt mit aller Macht sicherstellen, dass keine Informationen vorzeitig nach außen dringen. Datendiebstahl ist ein Risiko, doch nicht minder der schlichte Bezug auf Daten im anerkennendem Zitat. Jeder Konkurrent, der auf vorab kursierende Daten Bezug nimmt, lenkt die Debatte von auf seine eigenen Gleise.

Das Problem der irgendwann mit begleitenden Tweets ans Netz gehenden Projekte, die dann für die Diskussion eingefroren werden, ist, dass sie im Internet kaum wirksam werden. Stabilisierte Information wird von allen Suchmaschinen schnell gegenüber Information laufend aktualisierter Medien abgewertet – das aus gutem Grund: Neue Bildschirmformaten lassen alte Websites schnell antiquiert aussehen, wachsende Datenübertragungsvolumen erlauben es, neue Design zu riskieren. Was viel schwerer wiegt: Daten veralten im Netz ganz neuartig – nicht, weil in ihnen Fehler unkorrigiert bleiben. Die Erwartungen an ihre Verlinkung wachsen soeben rasant. Forschungsdaten, die älter als 10 Jahre sind, sind heute kaum mehr nachnutzbar. Wer die Wahl hat, lädt sich vergleichbare Daten aus Wikidata herunter, da sie, aktuell vernetzt, sich sofort gegen eigene (ebenfalls vernetzte) Daten abgleichen lassen.

### *Das FactGrid als Ort der aktuellen globalen Datenlage*

Das FactGrid macht weder ein Angebot an Projekte, hier die „eigene“ Datenlage diskret handhaben zu können – die Software bietet nicht an, dass man auf ihr Projekte hinter Firewalls verstecken kann, sie erlaubt es schon gar nicht die Daten eines Projektes nach deren Publikation zum Zwecke der ewigen Diskussion dieses Wissensstands einzufrieren.



Wir setzen gerade darauf, dass Datensätze von nachkommenden Projekten fortlaufend weiter verbunden und vermehrt werden – und wir setzen dabei auf Werkzeuge, die von uns ganz unabhängig entwickelt werden und die Daten, sobald sie bis auf die atomare Ebene der Datentripel heruntergebrochen sind, im Fluss der weiteren Entwicklungen halten.

Interessant und ist das FactGrid darum vor allem für Projekte, deren Ziel es ist, Datenlagen ihres ganzen Faches öffentlich verfügbar zu machen. Adam Andersons an der University of California, Berkeley, konzipiertes Cuneiform-Project, das in seinem ersten Schritt alle Artefakte mit Keilschriftzeichen verfügbar machen will – im Moment eine Gesamtmenge von 360.000 Items – ist Projekt par excellence dieses neuen Zuschnitts. Auf den ersten Blick ist das die Skizze eines entlegenen Gebietes, eines Nischenfachs. Das Spannende an ihm ist jedoch, dass sich gerade im Nischenfach die Interaktion einer globalen Forschungscommunity bahnbrechend einrichten lässt, einer technisch sehr versierten Community – auf einer Datenbank, die im selben Moment die Nische auflösen würde, denn die multilinguale Software würde Information sofort über das Fach hinaus lesbar machen.

Projekte wie das von Adam Anderson geplante markieren die aktuelle Größenordnung auf uns zukommender Datenvolumina: Die Regensburger Erik Amburger-Datenbank steht mit 138.000 Personen-Einträgen im Raum. Das PhiloBiblon-Projekt zu iberischen Codices plant 450.000 Objekte zu Handschriften, Werken, Personen, Institutionen und Orten.

In dem Maße, in dem große Projekte Daten einbringen, werden die Wünsche kleiner Projekte, unsichtbar arbeiten zu können, rar. Die Frage ist in Anbetracht großer Projekte, wie man die Möglichkeit erhält, an deren Daten Details verändern zu können – und wie man eigene Projekte zeitnah zwischen großen Projekten in der Forschungslandschaft sichtbar macht.

### *Das FactGrid als interaktive Forschungsumgebung*

Wir gingen ursprünglich davon aus, dass das FactGrid auf längere Zeit in der breiteren Nutzung durch einzelne Forschende unattraktiv bleiben würde – schlicht, weil diese laufend viel zu viele Objekten anlegen müssen, um ihre Objekte sinnvoll vernetzen zu können. Die Bespieglung der Ressource mit allen GND-Daten, schien für die deutschsprachige Forschung *state of the art*.<sup>22</sup>

---

<sup>22</sup> Ich resümierte die Diskussion „2017-12-1/2: Erster FactGrid Workshop in Berlin: Ergebnisse“ FactGrid-Blog (4. Dez. 2017) <https://blog.factgrid.de/archives/752>. Im März 2019 gingen wir, aus den Überlegungen resultierend eine Kooperationsvereinbarung mit der Deutschen Nationalbibliothek ein, die jedoch dann in Anbetracht der enormen datentechnischen Hürden folgenlos blieb. Publiziert ist sie mit dem: „Memorandum of Understanding zwischen der Universität Erfurt und der Deutschen Nationalbibliothek – das FactGrid wird in einem Gemeinschaftsprojekt auf GND-Daten aufgesetzt“, *FactGrid-Blog* (9. Apr. 2019), <https://blog.factgrid.de/archives/1475>.

Tatsächlich waren es einzelne Nutzer, die das Experiment mit der neuartigen Software schon nach wenigen Wochen wagten. Sie agierten anfangs ausschließlich an den Rändern unseres Ausgangs-Datensatzes zum Illuminatenorden. Vor- und Nachteile der unbespielten Ressource hielten sich an diesen Rändern die Waage. Einerseits waren laufend neue Objekte anzulegen, andererseits war die Instanz im speziellen Gebiet durchaus interessant vorbestückt und neuartig tief modelliert. Wir setzten schnell nach – nicht mit einer Bespielung durch GND-Daten, sondern mit flächendeckenden geographischen Ausstattungen, die dann nach Nationen voranschritt. Deutschland, Frankreich, Schweden, Großbritannien, Spanien, Portugal und Ungarn sind im Moment flächendeckend mit Orten bis auf Dorfebene in der Datenbank verfügbar, so dass bei neu angelegten Personen hier wenigstens die geographischen Bezugspunkte sofort vernetzbar vorliegen. Der Verzicht auf die GND-Bespielung schafft im selben Moment die Chance, Forschungsdaten nach eigenen Maßstäben generieren zu können, ohne mit dem gewachsenen Datenbestand laufend kritisch umgehen zu müssen.

Auch das Argument, dass man Informationen bis zur geplanten Publikation geheim hält, erwies sich in der Praxis der persönlich vorangetriebenen Projekte überraschend durchlässig. Isabella Schwaderer brach diese Barriere 2020 mit einem Projekt, das sie im Corona Lockdown eher experimentell betrieben haben wollte. Statt in einer „eigenen“ Word-Datei zu arbeiten, legte sie Hintergrundinformationen zu den Mitgliedern der Schopenhauer-Gesellschaft im FactGrid ab – ein Schritt, der, wie sie später beschrieb, angenehm produktiv war. Dort, wo sie nicht weiterkam, konnten immer wieder Andere Verknüpfungen zu Informationen herstellen. Aus dem Experiment wurde ein Projekt, das sich, so wie es aussieht, soeben zu einer Buchpublikation ausweitet. Es sieht nicht danach aus, als ob hier ein größeres Risiko des Datendiebstahls bestünde. Das liegt vor allem daran, dass das anfangs private Projekt auf der Datenbank gerade transparent positioniert ist. Sehr viel wahrscheinlicher ist, dass die Daten als Visitenkarte der eigenen laufenden Arbeit fungieren. Das FactGrid wurde hier sehr schnell eine Option Forschung offen zu positionieren.

Extrem produktiv gestaltete sich die Tatsache, dass sich Information dabei gerade im unfertigen Zustand handhaben lässt im Fall der noch nicht publizierten Dissertation Erik Liebschers. Liebscher arbeitet zu Christian Georg von Helmolt (1728–1805), einem Gothaer Hofbeamten mit Vernetzung in den Illuminatenorden hinein. In einer Selbstauskunft des Jahres 1783 hatte der erwähnt, dass er als Student, Ende der 1740er Jahre, in Jena Mitglied eines „Ordens der grünen Bergmänner“ gewesen sei. In den öffentlichen Archivdatenbanken war zu diesem nichts zu finden.

Das Interessante an der Graph-Datenbank ist, dass wenige vorhandene Aussagen sich in ihr bereits zu für Suchmaschinen interessanten Informationsgefügen zusammensetzen. In diesem Fall ließ sich das Item zur Organisation mit der Universität Jena, der Datierung, drei Personen und dem Quelldokument verbinden. Im Datengraph ein komplexes, vernetzbares

Gefüge. Anderthalb Wochen nach der Eröffnung des Items erhielten wir Post eines US-amerikanischen Forscherteams, das im Archiv der Schwedischen Großloge in Stockholm Dokumente sichtete und dabei auf mehrere Akten der Grünen Bergmänner aus der Hand von Helmolts gestoßen war. Man hatte in der Google-Recherche die Verbindungen erhalten und war neugierig.

Für das Team der Gothaer Illuminatenforschung war der Treffer unmittelbar brisant: Es lagen demnach noch Akten der originalen „Schwedenkisten“ des Jahres 1804, Akten aus dem Umfeld der illuminatischen Dokumentation, in Schweden, die wir dringend sichten mussten. Erik Liebscher verfasste im FactGrid Blog einen eigenen Artikel zur Organisation, über die wir dank des Eingeständnisses des Informationsdesiderats heute spektakulär viel wissen.<sup>23</sup> Eine Gruppe von Studenten veranstaltete in dieser Organisation, was wir heute als gruppendynamische Experimente verbuchen würden.

Forschenden, die im Rahmen von Dissertationen Daten in neue Zusammenhänge bringen, wird man an dieser Stelle den Ratschlag geben können, sich der Plattform vor allem unter strategischen Abwägungen zu bedienen:

- Was von den generierten Daten ist tatsächlich sensibel, gesammelt, um im Moment der Veröffentlichung Diskussionen auszulösen?
- Was von den „eigenen“ Daten ist dagegen Information, die man gefahrlos in Arbeitszusammenhänge bringen kann?
- Was sind Fragen, denen man im akuten Arbeitspensum nicht nachgehen wird, und die man gerade darum einfach einmal ins Netz stellen kann in der Hoffnung, dass sie sich dort weiter vernetzen?
- Inwieweit lässt sich die transparent gehandhabte Recherche gegenüber Geldgebern in Zukunft als Projektfeature anbieten, so dass gerade die Nutzung der eigenen Daten durch andere als Erfolg verbucht werden kann?
- Wann sollte man Rechercheergebnisse gerade ostentativ vorveröffentlichen, um damit Claims abzustecken? Der FactGrid Projekt-Blog bietet sich dabei als Ort an, an dem man Thesen und Dateninterpretationen Jahre vor der anvisierten Buchpublikation fixieren kann.

Entscheidend ist unter all diesen Fragen, ob es uns gelingt, getane Arbeit sichtbar und zitierbar zu machen. Das ist nicht nur eine Frage der Software, es ist auch und ganz entscheidend eine Frage der Positionierung der gesamten Plattform in der Forschungslandschaft. Je klarer diese zur Ressource für die Erstpublikation von Forschungsergebnissen wird, desto größer wird der Anreiz, eigene Befunde hier frühzeitig, weit vor der großen Publikation im Buch für sich zu reklamieren.

---

<sup>23</sup> Erik Liebscher, „Der ‚Orden der grünen Bergmänner‘ – Ein unbekanntes Stück Sozietätsgeschichte aus den Tiefen der ‚Schwedenkiste‘“, FactGrid-Blog (6. 8. 2022), <https://blog.factgrid.de/archives/2970>.

## *Gar nicht so einfach: Mit dem FactGrid Inselprojekte reaktivieren*

Ein guter Teil der Projekte, die mit uns in den letzten Jahren Kontakt aufnahmen, hatten lange Vergangenheiten hinter sich – ob auf privaten Excelblättern oder auf Datenbanken, die sich im Internet nicht weiter bewährten. Spektakuläre Inselprojekte waren darunter: Projekte, die erst vor Jahren ans Netz gingen und deren Software-Module bei Visualisierungen nun nicht mehr griffen. Die Vorteile der offenen Datenbank sind den Interessenten im Moment der Kontaktaufnahme klar. Es erweist sich jedoch als alles andere als einfach, dergleichen Altdaten eingabefertig zu bekommen. Das Problem sind dabei die Felder, in denen stets Text steht – regelmäßig kaum zuzuordnender Text, Städtenamen zu denen es im FactGrid stets gleich mehrere mögliche Städte dieses Namens gibt, Datierungen wie „wohl kurz nach 1670“. Das Problem sind hier fast immer auch Felder, in denen sich enzyklopädische Darlegungen finden, in Personendatenbanken etwa zu Ausbildungsstationen und Erwerbsbiographien. Jede darin genannte Universität, jeder notierte Mäzen muss bei uns als eigenes Datenbankobjekt mit eigenen Informationen im Raum stehen und jeder Bezug muss dabei einzeln hergestellt werden. Sobald das klarer ist, ist zumeist auch klarer, dass wir zwar Daten in Sekundenschnelle importieren können, aber dass diese Daten in dieser Klarheit nie generiert wurden. Man hätte mit einer Ressource wie der unsrigen noch im Moment der Datenerhebung arbeiten müssen.

## *Die eigene Wikibase-Instanz – eine überraschend abgründige Alternative*

Wo die Entscheidung für Wikibase fällt, sind die akuten Bedenken nicht mehr akut. Datentechnisch bewegt man sich hier auf der sicheren Seite. Man wird Daten in Tripeln auf beliebigen Maschinen laufen lassen können, Wikibase ist da nur eine Option – indes eine große und leistungsfähige und absehbar auf Jahre in Entwicklung. Die Frage ist hier, ob man nicht gerade darum auf die eigene Instanz setzen sollte. Das geschieht derzeit mit Instanzen ganz unterschiedlicher Projektreichweiten.

Neu sind, seit es Wikibase in Docker Images gibt, einzelne Projekte, die heute Wikibase so nutzen wie Kollegen bislang Nodegoat nutzen. Man setzt die eigene Instanz bei sich auf dem Server auf und legt ausschließlich die Objekte an, die man im eigenen Projekt benötigt – das können 2000 Romane einer Sprache sein. Personen und Orte legt man exakt im Zuschnitt des unmittelbaren Bedarfs an. Die Installation funktioniert unmittelbar in sich konsistent.<sup>24</sup>

Der Vorteil der eigenen Wikibase-Instanz ist, dass man auf ihr Datenstrukturen ganz für sich fixieren kann. Eine Insellage riskiert man dabei nicht. Spezialinstanzen werden in Zukunft

---

<sup>24</sup> Weit gediehen ist hier etwas über 14.000 Datenbankobjekten und eigener Gestaltung die Ungarische ELTE-Instanz [https://eltdedata.elte-dh.hu/wiki/Main\\_Page](https://eltdedata.elte-dh.hu/wiki/Main_Page). Frisch im Netz ist die Trierer Mimotext-Instanz (zum Projekt mehr unter <https://mimotext.uni-trier.de/aktuelles>) mit einer Installation zu – im Moment – Französischen Romanen [http://data.mimotext.uni-trier.de/wiki/Main\\_Page](http://data.mimotext.uni-trier.de/wiki/Main_Page) und 3900 Items. Zahlen Stand August 2022.

von großen Wikibase-Suchmaschinen (eine solche wird die deutsche Nationale Forschungsdaten-Infrastruktur, vermutlich herstellen) ganz wie Großinstanzen ausgewertet und ins Boot der globalen Datenvernetzung geholt werden.

Knifflig ist jedoch jede fortlaufende Arbeit auf solchen Plattformen. Sobald man den Projektbereich ausweitet, fehlen Datenbankobjekte, und bei deren Anlage tun sich dann laufend die neuen Abgründe auf. Die Arbeit auf der kollektiv breit bespielten Ressource ist an dieser Stelle bequemer – allerdings eben auch komplizierter in allen Abstimmungsprozessen, auf die man sich im größeren Kollektiv einlassen muss, vielleicht aber auch genau darin fruchtbarer, da mit den Kontakten Knowhow und Informationen fließen.

(2) Weitgehend unbemerkt dürften in den letzten drei Jahren Dutzende an verborgene Instanzen entstanden sein – ich hatte 2019/2020 intensiveren Kontakt zu den Projektbeteiligten des Exzellenzclusters Temporal Communities der Freien Universität Berlin.<sup>25</sup> Man hatte hier mit extremer Umsicht und technischer Expertise eine intern arbeitende Wikibase-Instanz aufgesetzt. Wie sich die Instanz entfaltete und wie sich das Feld solcher Instanzen derzeit gestaltet, lässt sich nicht ermessen. Es ist dies ein Feld, das zwar einen spezifischen Schutz im Inneren anbietet, aber im selben Moment Projekte auch ganz eigenen Risiken aussetzt – sie arbeiten hier nicht auf Plattformen, die sich externen Zeugen und der breiten Öffentlichkeit aussetzen.

(3) Die große Entwicklung, die seit 2020 vorankommt, liegt in Nationalen Infrastrukturnetzwerken. Die deutsche Nationale Daten Infrastruktur NFDI ist derzeit mitten im Formierungsprozess. Wikibase nimmt in ihr einen zentralen Platz ein als die Software, die zwischen den sich konstituierenden Konsortien interdisziplinär vermitteln kann. Ich sah vergleichbare Entwicklungen jüngst in der Planungsphase seitens der Nationalen Akademie der Wissenschaften in Polen. Wir befinden uns hier in einer offenen und transparenten Netzworkebildung, in der Kontakte zwischen Projekten jederzeit niederschwellig zustande kommen.

Die große Frage, die mit dieser aktuellen Entwicklung im Raum steht, ist, ob sie das Potential der Software erfassen. Mit Gelassenheit beurteilt, liegen hier keine tieferen Gefahrenzonen. Man wird eines Tages Nutzeroberflächen haben, die all diese Instanzen auswerten und Informationen aus ihnen zu neuen Gesamtbildern des aktuellen Forschungsstands zusammenfügen. Dabei spielt dann keine Rolle, dass einem Item zu einem Bischof aus den ehemaligen deutschen Ostgebieten – aus dem *Germania Sacra* Projekt gespeist – auf Seiten der deutschen Forschungsdateninfrastruktur existiert und diesem ein komplementäres Item zur selben Person, aus denselben Informationsquellen gespeist, auf der Instanz der polnischen Akademie der Wissenschaften gegenübersteht.

---

<sup>25</sup> Siehe das Portal der Unternehmung unter <https://www.temporal-communities.de/>.

Das eminente Risiko dieser Nationalisierung der Landschaft liegt nicht so sehr in der Mehrarbeit, es liegt in der vertanen Chance der unmittelbaren Kommunikation, die Wikidata an selber Stelle zwischen den Wikimedia Communities herstellt. Es sollte interessant sein, mehrsprachige Ressourcen interaktiv und kooperativ zu betreiben, und dabei sollten wir gerade am grenzüberschreitenden wissenschaftlichen Austausch interessiert sein. Hier liegt weit weniger das hehres Ziel einer globalen Respublica Literaria als ein Ziel in der alltäglichen Arbeit jedes Einzelnen. Der deutschen Romanistin ist wenig damit gedient, dass sie ihr Wissen zu französischer Literatur auf Anweisung der deutschen nationalen Forschungsdaten-Infrastruktur auf einer deutschen Plattform unterbringt, während ihre Kollegen im Land, über das sie forscht, auf einer französischen Plattform arbeiten – jener Plattform, die sie bei Recherchen nebenbei bevorzugt benutzt.

Dieselbe Frage nach Integrationsleistung wird im NFDI-Prozess aufkommen. Es wird nicht klug sein wenn Wikibase-Instanzen identische Objekte in Massen anlegen. Die fruchtbringende Infrastruktur wird Felder nicht nach Fachgrenzen, sondern nach Dienstleistungen trennen: Das FactGrid generiert Faktenbeziehungen zwischen Wissensgegenständen – es eignet sich nicht als Editionsplattform. Im klugen Arrangement wird man in Bereichen wie *Text+*<sup>26</sup> hier das interaktive Angebote machen, Texte zu hosten – die historischen Aussagen zu ihnen jedoch eher einer Ressource übergeben, die gerade Aussagen zu Objekten verwaltet. Das Interessante am NFDI-Prozess ist, dass hier mit derselben Software in allen Feldern experimentiert wird und dass die Beteiligten dabei im Moment in einem breiten Austausch miteinander stehen.

## 5. Ausblick. Das Ende der Digital Humanities – ist längst überfällig

Eine Kollegin stellte auf einer Diskussionsveranstaltung 2019 die Frage, die auf einen Schlag die Stühle sichtbar machte, zwischen denen wir aktuelle agieren – von sitzen kann man hier nicht so recht sprechen. Was, so die Frage, könne sie, die „konventionelle“ Historikerin mit dem FactGrid denn aber nun machen, was sie nicht weitaus effizienter mit einer Word-Datei ganz für sich längst tue. In ihr sammelt sie, was ihr soeben aufhebenswert erscheint: Die Chronologie von Ereignissen, später zu zitierende Textpassagen, Vermerke von Sekundärliteratur, schnell hingeworfene Gedanken, die dem parallel geschriebenen Aufsatz vorausseilen. Die Datei funktioniert ohne Datenmodell. Sie ist beliebig durchsuchbar und ein Steinbruch an Information. Netzwerkwolken? Irgend so etwas, was man in den „Digital Humanities“ macht?

Tatsächlich sollte die eigene Word-Datei das einfachste das Mittel zur Datensammlung sein.

---

<sup>26</sup> Das bereits genehmigte und agierende NFDI-Konsortium für Sammlungen, lexikalische Ressourcen und Editionen mit Projektrepräsentanz unter <https://www.text-plus.org/en/home/>.

Auf einer anderen Ebene wird man sich mit selbstbekennenden konventionellen Geisteswissenschaftlern und Geisteswissenschaftlerinnen im komplementären Einvernehmen einigen können: Niemand von uns bewegt sich noch heute in einem vordigitalen Raum. Wir alle suchen zuerst in Google; vielleicht lesen wir einmal in einer Fachbibliothek Artikel einiger Fachlexika zu einem Problem – mehr aus Neugier. Daten bietet Wikipedia verlässlicher, können hier doch Korrekturen von jedem vorgenommen werden. Datenbanken? Nutzt jeder Geisteswissenschaftler in Katalogrecherchen. Digitalisate – lesen wir natürlich (bevorzugt die von Google Books, in denen man schneller blättert als im DFG Viewer und die zudem rudimentäre texterkannt sind).

Was unsere eigene Position gegenüber den neuen Wissensanbietern anbetrifft, ist die Lage überraschend eindeutig. Wir haben auf ihnen nichts zu vermelden. Google sammelt Daten über uns, wir wissen nicht nach welchen Algorithmen es uns im Gegenzug individuelle Suchergebnisse zusammenstellt. In Wikipedia könnten wir Daten ändern, aber nicht länger als Experten der Forschung. Wenn wir in einem Bibliothekskatalog auf einen Fehler stoßen, können wir dessen Sachbearbeiter über den Aufsatz informieren, in dem wir dieses Datum längst richtigstellten. Wir agieren in der digitalen Welt als Kunden, Nutzer und Bittsteller – nicht als die Wissensanbieter.

Sollten wir die konventionellen Geisteswissenschaften abschaffen und Akteure der alternativen Digital Humanities werden? Dagegen spricht im Moment alles. Die breiten Debatten gehen von den Vertretern der konventionellen Geisteswissenschaften aus. Es geht mithin weit eher darum, wo wir in den Debatten generierenden Geisteswissenschaften mitten im Eingeständnis stehen, dass wir den Einfluss auf die neuen Medien nicht weiter angestrebten noch erlangten. Wir erwarten die Rezeption unserer Arbeiten und wir setzen dabei auf Verlage, die nun plötzlich nicht mehr das tun, was sie für uns 400 Jahre lang taten, im eigenen Interesse die Reichweite unserer Arbeiten maximieren. Mit dem Internet hat sich die Position der Fachverlage in neuen Geschäftsmodellen auf die Verknappung von Information hin umgestaltet. Wir unterschreiben komplexe Einverständniserklärungen, in denen wir unsere Rechte umfänglich entäußern. Die unsere Rechte an sich nehmenden Verlage schöpfen im nächsten Schritt auf unserer Seite – eine öffentliche Finanzierung des Privatsektors – Druckkostenzuschüsse ab. Auf Seiten der öffentlichen Bibliotheken, die allein als Kunden der künstlich verteuerten Bücher in Frage kommen findet eine zweite Abschöpfung öffentliche Gelder statt. Sie subscribieren Reihen und ganze Verlagsproduktionen. Digitalausgaben müssen die Verlage im eigenen Interesse unter Maßgaben der offen sichtbaren Verknappung inszenieren. Komplexe Lizenzmodellen und eingeschränkte Zugriffsrechte, die von den Bibliotheken verwaltet werden müssen verlagern hier die Verknappung in den öffentlichen Sektor.

Eine eigene Marktnische entdeckten in den letzten Jahren neue letztlich kommerzielle Websites rund um Academia.edu<sup>27</sup> und ResearchGate<sup>28</sup> als Social Media der Forschung: Hinter eigenen Zugangsbeschränkungen machen sie es den Forschenden möglich, im Privaten, so die Insinuation, ihre Artikel in Form von Word-Dateien und als pdfs einander zukommen lassen.

Unser Problem ist in den Nationen, die sich dem liberalen Markt stellten, nicht staatliche Zensur, sondern eine prekäre Organisation unserer eigenen Diskursteilhabe im Internet, bei der wir uns leichtfertig in die Position von Kunden begaben. Sollte das Internet denn das neue Medium sein, dann müssten dort eben fleißige bibliothekarische Kräfte und unermüdliche Wikipedianer unsere Aufsätze auswerten und Datum um Datum danach abklopfen, ob wir damit soeben den Wissensstand veränderten. Wir manövrierten uns unversehens in ein Aus, in dem niemand für uns speziellen Edit-Knöpfe einrichtete, mit denen wir als Forscher Sachen wie gewohnt noch richtigstellen können.

Die Frage der letzten vier Jahre lautet, welche Position wir nun im soeben laufenden Entwicklungsschub hin zu öffentlich frei verfügbaren Datenlagen jenseits von Google einnehmen wollen. Unsere Forschung verbindet sich mit Datenlagen; wir werden mit Daten sichtbar – oder auch nicht.

Wir sollten hier Wikimedia, die Nationalbibliotheken und die Archivverbünde nicht als Konkurrenten von minderer Wissenschaftlichkeit uns gegenübersehen, sondern als Partner, von denen wir in der täglichen Arbeit gerne weit mehr profitieren würden und denen wir im Gegenzug bieten können, worauf sie sich unter den eigenen organisatorischen Prämissen nicht einlassen können: Wir können theoretisch Forschungswissen anbieten, das auf eigenen Plattformen mitsamt den Debatten ins Spiel bringen können, um die es uns mit unserer Forschung geht. Die Prozesse, in denen wir uns hier bewegen, werden Zumutungen bergen und Chancen Weichenstellungen zumindest zu beeinflussen:

**(1) Die Institutionen der Forschungsförderung werden in den kommenden Jahren Projekten zunehmend mit Forderungen nach einem nachhaltigen Datenmanagement gegenüberreten.** Datenpakete werden sich als nicht weiter nachnutzbare wissenschaftliche Kompensation erweisen. Die Forschungsförderung wird auf die Bespielung von Plattformen drängen, auf denen Information sich sofort als fruchtbar vernetzbar erweist – oder noch vor dem Hochladen als längst vorhanden und in der Vernetzung irrelevant. Apelle, Daten nachnutzbar zu machen, werden verwirrend kontraproduktiv bleiben, lästige Pflichtübungen schaffen.

---

<sup>27</sup> 2008 gegründet mit 71 Millionen Nutzern ausgerichtete non-profit Plattform, die immer wieder wie eine Simulation der Universitätslandschaft agiert und sich auf den Tausch von Artikeln unter Wissenschaftlern ausrichtet: <http://academia.edu/>.

<sup>28</sup> Ebenfalls 2008 gegründete und mit 17 Millionen Nutzern kommerziell ausgerichtet primär auf die Naturwissenschaften ausgerichtet: <https://www.researchgate.net/>



**(2) Erfolgreicher werden erst Forschungsumgebungen agieren, auf denen sichtbar zu sein, in den Geisteswissenschaften unmittelbar interessant wird – da man auf diesen Ressourcen selbst recherchiert und eben hier selbst gesehen werden will.** Derzeit gibt es in der Recherche keine Alternative zu Google. Die aktuelle Hinwendung zu triplbasierter Information schafft gerade einmal die datentechnische Voraussetzung, auf der ein wissenschaftsgetriebenen Knowledge-Graph eines Tages denkbar sein wird. Das Ziel ist dabei eine globale Suchumgebung, die nicht Links zu Websites liefert, auf denen Nutzer die Information möglicherweise finden. Das Ziel ist das Wissen selbst, und was die Forschung anbetrifft: ein Wissen das auf der Tiefenebene aller einzelnen Aussagen mit Forschung und Dokumenten gedeckt ist. Wir werden hier ein Geflecht an Datenbanken erhalten zwischen denen Forschungsdatenbanken es Wissenschaftlern und Wissenschaftlerinnen niederschwellig erlauben mitten im Umgang mit eigenen Daten diese global zu präsentieren.

**(3) Für die Geisteswissenschaften wird der Aufbau einer Ressource interessant sein, in die Informationen der Bibliotheken und Archiven weit hineinragen.** Wir werden Datenressourcen attraktiv finden, in denen wir Archivalien und Digitalisate nicht nur recherchieren, sondern auch unmittelbar vernetzen, bearbeiten, auswerten, mit Transkription belegen – namentlich als Forschende sichtbar.

**(4) Interdisziplinäre Forschungsplattformen sollten dabei in unserem Interesse liegen.** Im Moment scheint hier die Zersplitterung attraktiv: Triplbasierte Information lässt sich ganz beliebig aus getrennten Ressourcen zusammenfügen. Warum soll da nicht eine kunsthistorische Plattform Daten zu Dürer erfassen, und unabhängig von ihr eine historische nicht minder? Die interdisziplinären Ressourcen geben, sobald wir uns auf Wikibase als Software einlassen, den Forschenden unmittelbar Rückmeldung, wo immer jemand ihr Wissen in ganz eigenen Zusammenhängen tangiert. Das Ergebnis ist auf kollektiven Plattformen die Chance, die ResearchGate oder Avcademia.edu, als Social Media-Plattformen der Forschung soeben bieten: sich Information über Fachgrenzen zuzuspielen zu können, gemeinsam an denselben Dingen arbeiten zu können.

**(5) Der Aufbau internationaler Plattform sollte bei alledem ein zentrales Desiderat werden.** Wieder ist korrekt, dass wir mit der neuen Technik tatsächlich Informationen aus Installationen von 195 Nationen amalgamieren können. Der Vorteil von Instanzen, die die Forschung global sammeln, liegt dagegen erneut in der unmittelbaren Sichtbarkeit der Nutzung unserer Daten. Wir sollten es internationalen Kooperationen leichtmachen, auf multilingualen Plattformen aufzutreten. Wir sollten mit wenigen globalem Ressourcen Chancengleichheit unter Forschern und Forscherinnen über Grenzen hinweg herstellen. Mit international Plattformen werden internationale Recherche-Interfaces kommen, die Wissen aller Beteiligten global ohne Achtung der Herkunft der Beiträge nutzbar machen. Es wäre nicht klug uns hier als deutsche Plattform in einer privilegierten Lage zu sehen. Wenn alle Sprachen eigene Forschungsdatennetze aufbauen, wird das deutsche dritte Wahl

sein. Das Ziel sollte die Arbeit jedes einzelnen Forschenden an der globalen Datenlage gegenüber sein – in Anbetracht einer globalen Nutzung der eigenen Arbeit.

Für das FactGrid wird in diesem Gefüge der Handlungsspielräume in den nächsten Jahren darum gehen, dass die Projekte, die die Plattform aufbauten, ihr nun eine eigene Außenvertretung geben. Die Wikimedia Foundation gibt hier das strategisch interessante Modell vor: Es liegt in einer Funktionstrennung zwischen der Plattform und der Außenrepräsentanz, in der sich die Außenrepräsentanz auf eine non-interference policy einlässt und sich primär dem kostenlosen und selbstorganisierten Betrieb der Plattform und ihrer Integration in größere Netzwerke verschreibt. Interessant wird es dabei sein, dass Projekt im experimentellen Raum zu stabilisieren, in dem es bislang agiert. Wir können als Plattform, auf der Nutzer namentlich kenntlich in Verantwortung für ihre Forschung agieren hier Freiräume anbieten, die Wikidata und die Nationalbibliotheken ausschlagen müssen. Hier liegt das Spannende an der Nutzung einer Wikibase durch Forschung.