

Tambora – die Entstehung einer virtuellen Forschungsumgebung

Franck Borel und Heike Steller

Im Rahmen des Projektes „HISCLIDCore“ hat die UB Freiburg zusammen mit Klimaforschern und dem Leibniz-Institut für Länderkunde eine virtuelle Forschungsumgebung unter dem Namen „Tambora“ konzipiert und programmiert. In diesem Artikel wird die Entstehung von Tambora beschrieben und das Ergebnis dieser zwei Jahre andauernden Entwicklung präsentiert. Weiterhin wird auf die aus diesem Projekt entstandenen neuen Aufgaben der UB eingegangen.

Within the project „HISCLIDCore“ the UB Freiburg, together with climate scientists and the Leibniz Institute for Regional Geography, created a virtual research environment under the name „Tambora“. This article describes the development of Tambora and presents the results of these two years of development. Furthermore, the new role of the university library derived from this new project, will be discussed.

» Moderne Forschung benötigt eine Plattform, mit der Forscher über Institutionsgrenzen hinweg gemeinsam arbeiten können. Diese Idee ist nicht neu, aber erst die Online-Wissenschaft macht kollaborative Forschung möglich. Dabei geht es nicht nur darum, gemeinsam Daten zu sammeln, sondern auch um den Austausch von Ideen und Informationen. Dies ist dann umso wichtiger, wenn Wissenschaftler eine Fragestellung nicht alleine beantworten können. Ein weiterer wichtiger Aspekt moderner Forschung ist die Sicherung und Veröffentlichung der Forschungsdaten. Bislang landeten Forschungsdaten nach der Publikation im Papierkorb.

Forschungsdaten sind die Grundlage für die wissenschaftliche Forschung. Das bisherige „Papierkorbmodell“ führt dazu, dass Daten mehrfach erhoben werden. In den Publikationen sind in der Regel die Forschungsdaten nicht abgedruckt und die interpretierten Daten geben die Details der Rohdaten nicht richtig wieder¹. Ziel ist es daher, das Papierkorbmodell durch ein Modell zu ersetzen, welches eine langfristige Archivierung ermöglicht und Forschungsdaten zitierfähig macht.

Aus diesen Gedanken heraus haben die wissenschaftlichen Einrichtungen IPG Freiburg, IfL Leipzig, IGUA Augsburg und die UB Freiburg **Tambora** konzipiert; eine virtuelle Forschungsumgebung für die historische Klimatologie, welche im Rahmen des von der DFG finanzierten Projektes HISCLIDCore entwickelt wurde.

Forschungsgegenstand

Forschungsgegenstand ist die historische Klimatologie. Die historische Klimatologie untersucht die Wirkung des Klimas und seiner Veränderungen auf historische Gesellschaften². Der typische Arbeitsablauf sieht wie folgt aus:

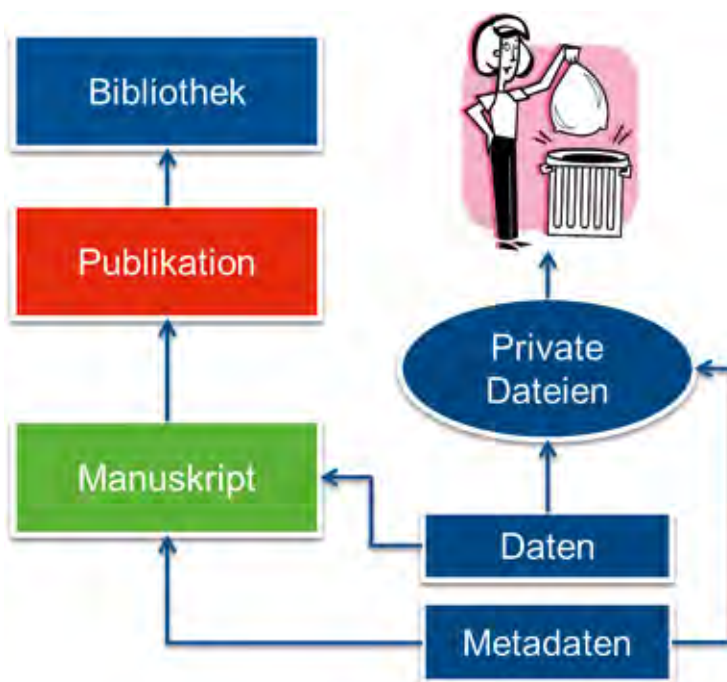
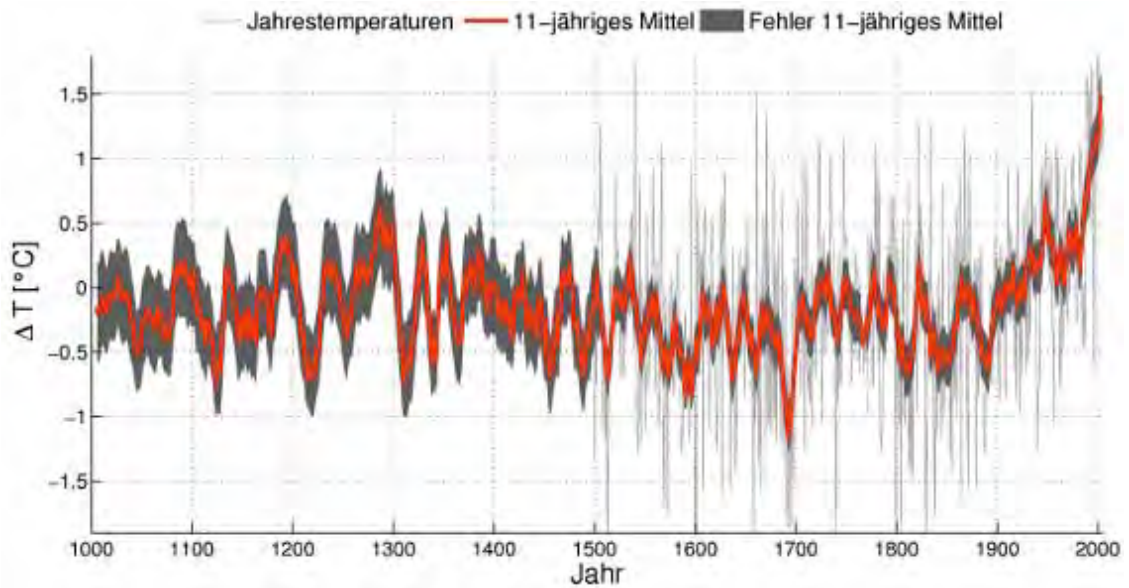


Abbildung 1: Das Papierkorbmodell. Schematische Darstellung des wissenschaftlichen Informationsflusses in der Forschung (Quelle: 2011, S. 26 nach Büttner, Hobohm & Müller (verändert))

¹ TIB HANNOVER/ FIZ CHEMIE BERLIN/ UNIVERSITÄT PADERBONN: Konzeptstudie Vernetzte Primärdaten-Infrastruktur für den Wissenschaftler-Arbeitsplatz in der Chemie, Paderborn 2010.

² URL: http://de.wikipedia.org/wiki/Historische_Klimatologie [27. Juni 2012].

Abbildung 2:
Rekonstruierte Temperaturentwicklung
Mittleuropas aus
historischen Quellen
(Quelle, RIEMANN
2010)



1. Recherche von Texten mit klimarelevanten Hinweisen
2. Katalogisieren der Texte
3. Auswertung der Angaben im Text und Parametrisierung
4. Verortung der Daten
5. Zeitliche Zuordnung

Dies soll hier anhand eines Beispiels näher erläutert werden: Am Institut für Physische Geographie Freiburg werten derzeit Wissenschaftler Reisetagebücher des Jesuitenpaters Gerbillon mit Wetternotizen aus. Eine solche Wetterinformation sieht wie folgt aus:

*20. Mai 1688 * Il fit grand chaud pendant tout le jour ... (Übersetzung: Es war den ganzen Tag über sehr heiß).*

Angaben zum Aufenthaltsort des Patres finden sich ebenfalls im Text:

... 50 Lys nördlich von Peking bis nach Tcha Ho, von dort aus etwa 30 Lys nach Nord-Nord-West, dann 19 bis 12 Lys nach Norden, dann 8 bis 10 Lys in etwas westliche Richtung bis zum Lager in Nan Keou ...

Diese Angaben werden klimatologisch kodiert. Dazu werden Worte und Phrasen mit klimatologischem Gehalt extrahiert und grammatikalisch normalisiert. Gegebenenfalls müssen die Ausdrücke vorher übersetzt werden. Jeder Ausdruck wird einem Katalog mit Klimaparametern gegenübergestellt. Der Parameter, der am besten passt, wird übernommen. Ein Parameter kann z.B. „Regen“ oder aber auch „starker Regen“ sein. Sobald die Kodierung abgeschlossen ist, wird das Ereignis, welches im Text beschrieben ist, verortet. Unter dem Begriff Verortung versteht man die

räumliche Zuordnung. Sie kann unterschiedliche geographische Objekte umfassen: neben Städten z.B. auch Flüsse oder Routenpunkte. Weil jedes dieser Objekte verschiedene Namen haben kann, benötigt der Forscher ein geographisches Namensverzeichnis. Dort findet er neben dem heutigen offiziellen Namen auch historische Namen und Übersetzungen in andere Sprachen. Neben der Verortung wird versucht, aus dem Text das Ereignis zeitlich einzuordnen. Hier muss der Historiker Datierung an Hand des historischen Kontextes in das heutige Zeit- und Kalendersystem übersetzen³.

Die Ergebnisse können dann z.B. statistisch ausgewertet werden oder in Form von Themenkarten dargestellt werden (Abb. 2).

Der Arbeitsablauf lässt sich auch so zusammenfassen: aus Textmaterial werden Zahlen. Der Aufwand ist beträchtlich: Neben der mühsamen Durchforstung nach geeignetem Material, ist die Auswertung und die Transkription der Texte Fleißarbeit und erfordert ein hohes Maß an Erfahrung und Wissen.

Bisher wurden die Daten entweder in Form von Excel-Sheets oder in lokalen Datenbanken gespeichert. Das Institut für Physische Geographie erkannte früh den Wert historischer Klimadaten und unternahm mit der Datenbank „Hisklid“ einen ersten Versuch, seine Daten auch anderen Forschern zur Verfügung zu stellen⁴.

Da das Interesse an diesen Daten in der Klimaforschergemeinde groß war, kam der Wunsch auf, für die Forschungsgemeinschaft eine virtuelle Forschungsumgebung aufzusetzen. Das Anforderungsprofil war ziemlich ambitiös: Die virtuelle Forschungsumgebung

³ BOREL, Franck: „HISCLIDCore“, in *Expressum* Nr. 6 (2010) S. 16-19.

⁴ URL: <http://www.tambora.org/index.php?r=site/page&view=hisklid> [12. Juli 2012].



Abbildung 3:
Vom Papier-
prototypen zur
fertigen Nutzer-
oberfläche

sollte nicht nur Datensammlungen archivieren können, sondern darüber hinaus die Aufnahme von Einzeldaten mit dem oben beschriebenen Arbeitsablauf erlauben. Die Wissenschaftler sollen sich in Projektgruppen zusammenstellen können mit verschiedenen Rechten und Rollen innerhalb des Projektes. Zusätzlich sollen Änderungen beim Editieren und Bearbeiten der Einträge mitverfolgt werden können.

Eine Fülle von Werkzeugen wie ein Thesaurus, ein Zeitkonverter oder ein geographisches Namensverzeichnis waren ebenfalls auf der Wunschliste. Für das Projekt standen nur anderthalb Entwickler für zwei Jahre zur Verfügung. Es versteht sich von selbst, dass nicht alles realisiert werden konnte. Einer der ersten Schritte nach der Projektgenehmigung war daher die Streichung einiger Arbeitspakete aus der Wunschliste.

Realisierung

Die Realisierung von Tambora erfolgte in sechs Schritten:

6. Untersuchung der Struktur von bestehenden umwelt- und klimahistorischen Datensätzen
7. Aufbau der Primärdaten-Datenbank
8. Untersuchung der bisherigen Arbeitsabläufe der Klimaforscher und Umsetzung dieser in benutzerfreundliches Werkzeug (Erfassungswerkzeug)
9. Import der Daten aus dem Hisklid- und dem Weikinn-Projekt
10. Aufbau des Repositoriums für Forschungsdaten an der UB Freiburg
11. Anbindung des Erfassungswerkzeugs an das Repositorium

Die Weboberflächen wurden gemeinsam mit den Forschern realisiert. Hier haben sich vor allem Papierprototypen bewährt (siehe Abb. 3).

Viel Wert wurde auf kurze Iterationszyklen bei der Konzipierung und Realisierung der Software gelegt.

Dadurch konnten früh Fehler oder falsche Ansätze erkannt werden. Das bedeutete auch, dass die Programmierer auf Änderungswünsche schnell reagieren mussten. Dies wiederum setzte eine flexible Softwarestruktur voraus und die Bereitschaft, auch mal fertige Programme umzuschreiben.

Die Software wurde größtenteils testgetrieben entwickelt. Testgetrieben bedeutet: Es werden zunächst die Software-Tests geschrieben, dann die zu testende Software. Dieses Verfahren ist aufwendiger, erhöht aber die Softwarequalität und erleichtert die Überarbeitung der Software.

Neben den rein technischen Aspekten spielte bei der Realisierung der intensive Austausch zwischen Entwicklern und Forschern eine wichtige Rolle. Interessant ist, dass immer wieder Begriffe verschieden interpretiert wurden. Dabei ging es nicht nur um fachspezifische Terminologie, sondern auch um die Interpretation von geläufigen Begriffen, wie „Text“, „Version“, „Ort“ und „Zeit“. Hierfür waren viele Sitzungen notwendig und es bedurfte von beiden Seiten einer Menge Geduld, bis schließlich Entwickler und Forscher (auch untereinander) die Begriffe in gleicher Art interpretierten. Hier hatten wir das Glück der unmittelbaren Nähe: Die UB Freiburg ist nur 100 m vom Institut für Physische Geographie entfernt.

Architektur

Kern von Tambora ist eine PostgreSQL-Datenbank mit der Erweiterung PostGIS. Mit PostGIS können geographische Objekte gespeichert werden (z.B. Geokoordinaten); dies ist notwendig, da die Arbeit mit klimahistorischen Daten verlangt, dass die Daten georeferenziert werden können.

Der Zugriff erfolgt über eine Webschnittstelle, welches mit dem PHP-Framework Yii realisiert wurde. Die Entscheidung für Yii wurde auf Grund der bisherigen Erfahrungen mit anderen Frameworks gefällt und auch deshalb, weil an der UB Freiburg noch andere

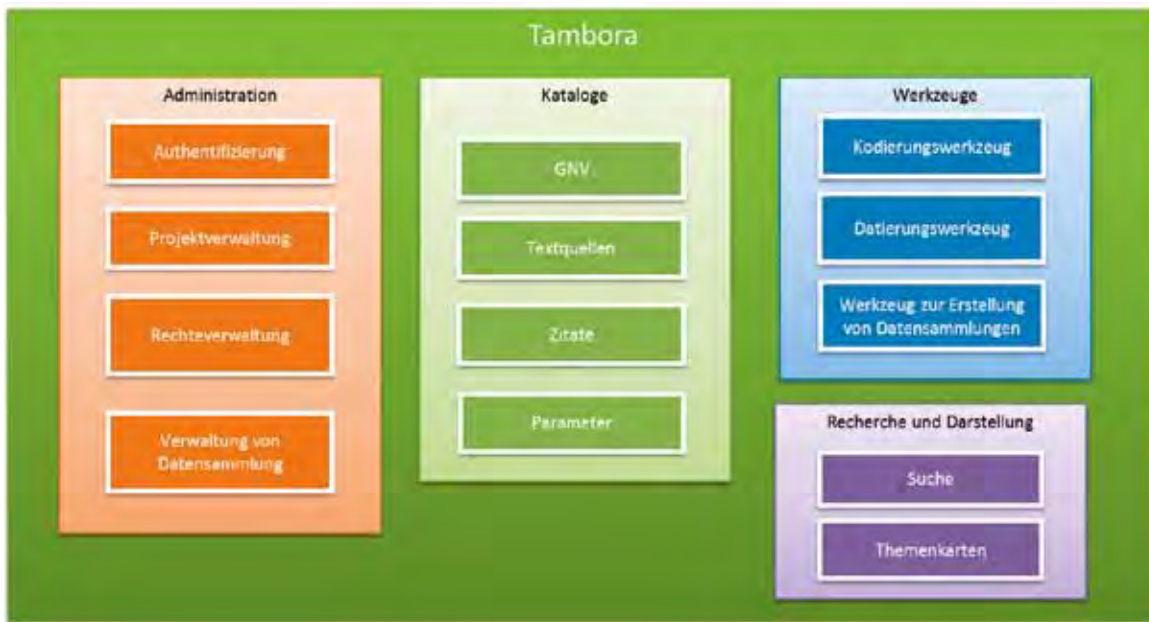


Abbildung 4.: Modularer Aufbau von Tambara

Projekte mit Yii realisiert werden sollten. Dadurch konnten Know-how und auch Codes ausgetauscht werden, was den Programmieraufwand reduzierte.

Die Datenbankstruktur musste im Laufe des Projektes mehrfach überarbeitet werden. Hier halfen uns so genannte Migration-Skripte: Statt direkt an der Datenbank Änderungen vorzunehmen, wurden Skripte geschrieben, die diese Änderung vornehmen. Der Clou an diesen Skripten: Die Skripte verfügen über eine Methode, mit welcher man die Änderung jederzeit rückgängig machen kann.

Tambora verfügt über verschiedene Module (siehe Abb. 4), welche die Datenbank nutzen. Eines dieser

Module ist die Projektverwaltung und die dazu gehörige rollenbasierte Benutzerverwaltung. Der Forscher kann Projekte anlegen, beliebige Nutzer mit verschiedenen Rollen hinzufügen und behält für die Dauer des Projektes die Kontrolle über seine Daten. So kann er verhindern, dass andere seine Daten einsehen können. Sobald er möchte, kann er die Daten veröffentlichen, einzeln, alle oder nur einen Ausschnitt davon. Dann stehen die Daten auch anderen Forschern zur Verfügung.

Wichtigstes Modul für den Forscher ist das Kodierungswerkzeug (Abb. 5). Hier werden die Textpassagen wie oben im schematisierten Arbeitsablauf

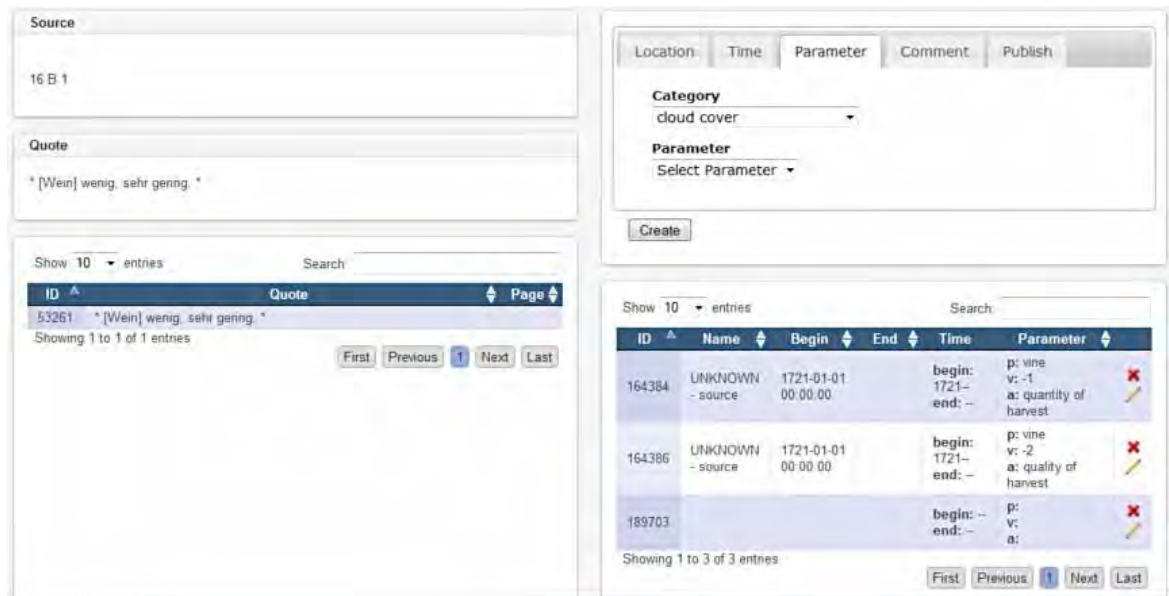
The screenshot shows a search interface for a parameter catalog. At the top, a search box contains the text 'regen'. Below it is a dropdown menu set to 'oder'. A grid of search results is displayed, with 'Wind' highlighted in purple. Below the grid is a 'zurücksetzen' button. At the bottom, a detailed list of search results is shown, with 'Regen' highlighted in red and 'Wind' highlighted in purple.

Search Result	Category
astronomische Beobachtung	Ernährungssicherung
hydrologisches Ereignis	Luftdruck
Messwert	Naturerscheinung
Plage	Temperatur
	Wirtschaft
	Himmelserscheinung
	Luftfeuchte
	Niederschlag
	Verkehr
	Wolkenbedeckung
	Hochwasser
	Massenbewegung
	Phänologie
	Wind

<input type="checkbox"/> Föhn	Wind
<input type="checkbox"/> Schneetreiben	Niederschlag Wind
<input checked="" type="checkbox"/> Windstärke	Wind
<input type="checkbox"/> Blutsregen	Niederschlag Naturerscheinung
<input checked="" type="checkbox"/> Wolkenbedeckungsgrad	Wolkenbedeckung
<input type="checkbox"/> abwechselnd Regen und Sonnensc...	Niederschlag
<input checked="" type="checkbox"/> Temperatur	Temperatur
<input checked="" type="checkbox"/> Regenintensität	Niederschlag
<input checked="" type="checkbox"/> Windrichtung	Wind
<input type="checkbox"/> vereinzelte Regentage, regnete...	Niederschlag
<input checked="" type="checkbox"/> Niederschlag	Niederschlag
<input checked="" type="checkbox"/> Wind	Wind
<input type="checkbox"/> Regen auf Schnee	Hochwasser Temperatur hydrologisches Ereignis Niederschlag
<input type="checkbox"/> Regen und Schnee, mit Regen ve...	Niederschlag
<input type="checkbox"/> Regen	Niederschlag
<input type="checkbox"/> Sprühregen (Staubregen), feine...	Niederschlag
<input type="checkbox"/> Regenbogen	Himmelserscheinung

Abbildung 5.: Suche im Parameterkatalog

Abbildung 6:
Kodierungs-
werkzeug



beschrieben parametrisiert. Der Parameterkatalog ist umfangreich und verfügt über eine komplizierte Struktur, weswegen hierfür ein eigenes Suchverfahren entwickelt wurde, welches eine Freitextsuche und mit einer Filtersuche kombiniert. Der Nutzer erhält das Suchergebnis in Form einer hierarchischen Liste, aus der er einen oder mehrere Parameter auswählen kann (Abb. 6).

Abbildung 7:
Werkzeuge, die
das Leben eines
Klimahistorikers
erleichtern:
Das geographi-
sche Namens-
verzeichnis

Werkzeuge wie die automatische Berechnung eines Zeitstempels, der Parameterkatalog und das geographische Namensverzeichnis helfen dem Nutzer bei der Übersetzung des Textes in quantifizierbare Werte.

Exemplarisch soll das geographische Namensverzeichnis stellvertretend für diese Werkzeuge näher beschrieben werden. Das geographische Namensverzeichnis (GNV) verfügt über eine ausgeklügelte Verwaltungsoberfläche und kann neben Orten auch andere Typen von geographischen Objekten wie Regionen, Flüsse oder Routenpunkte aufnehmen. Zu jedem geographischen Objekt können beliebig viele Namen und Namensvarianten gespeichert werden, wobei Zusatzinformationen wie Angaben zu Sprache, Schrift oder zeitlichem Vorkommen der Namensvariante über tags hinzugefügt werden können. Dieses

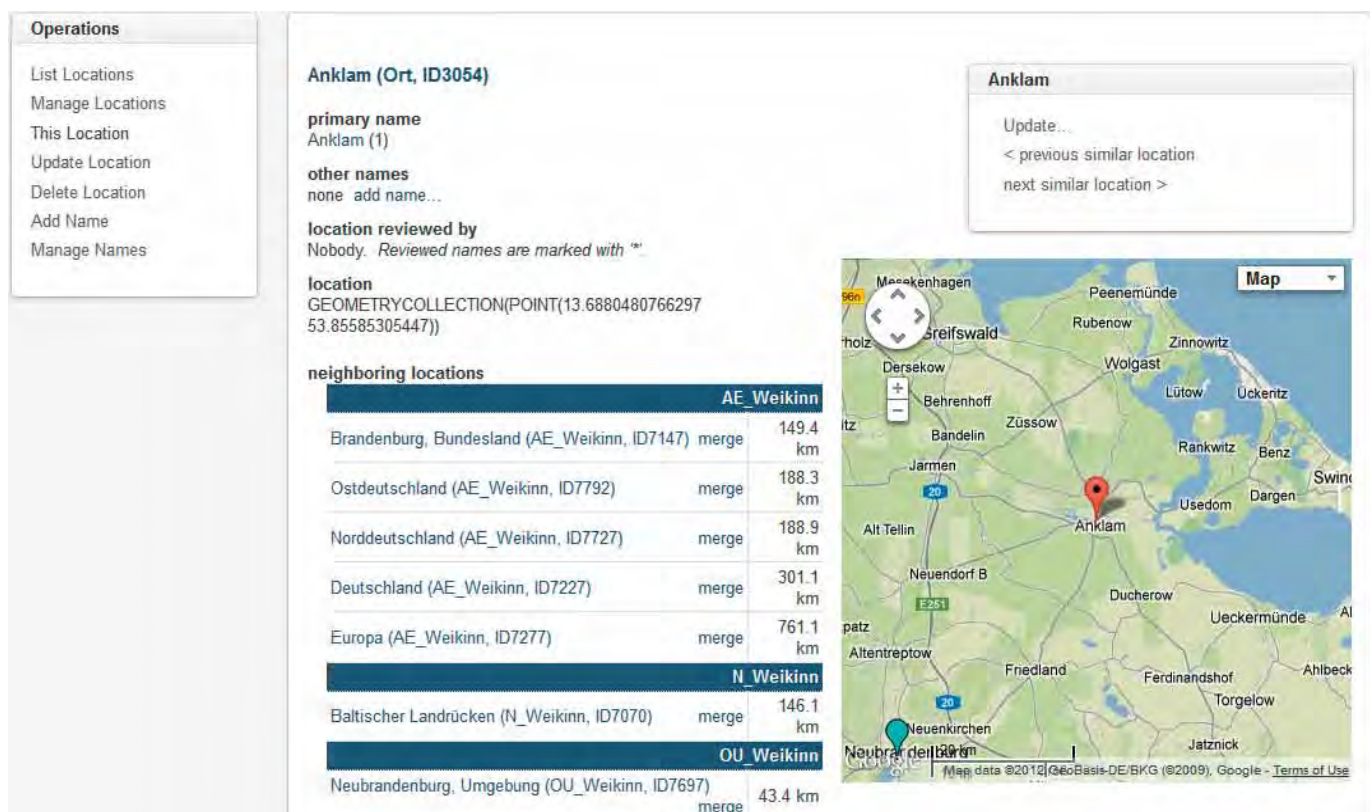




Abbildung 8:
Übertragen der
Datensammlung ins
Repositorium

Werkzeug ist sowohl für Geographen wie auch für Historiker interessant, weswegen auch darauf geachtet wurde, dass das geographische Namensverzeichnis auch ohne Tambora voll funktionsfähig ist. Das mühevoll durchforstete von historischen Ortsnamen in verschiedenen Veröffentlichungen und das anschließende Heraussuchen der Geokoordinaten entfällt mit dem GNV (Abb. 7).

Für die Publikation von Datensammlungen werden ausgewählte Datensätze in das Repositorium der UB Freiburg exportiert. Hierbei handelt es sich um einen Server zur Archivierung von Datensammlungen, welcher speziell für Forschungsdaten ausgelegt wurde. Die Übertragung geschieht wie in Abb. 8 dargestellt über FedoraProxy. FedoraProxy ist eine in PHP geschriebene Software, welche über eine REST-Schnittstelle mit dem Repositorium kommuniziert. Diese Schnittstelle ist notwendig, da Tambora und das

Repositorium auf getrennten Maschinen laufen und voneinander unabhängige Komponenten sind. FedoraProxy bietet Methoden an, mit der die REST-Schnittstellen (API-M, API-A) bequem angesprochen werden können. Weitere Methoden für das Exportieren von Dateien und zur Prüfung von vorhandenen Objekten und Datenströmen gehören ebenfalls dazu.

DOI-Vergabe

Damit die Rohdaten zitierfähig gemacht werden können, werden diese in Form einer Datensammlung gebündelt, mit Metadaten versehen und im Repositorium archiviert. Zu den Metadaten gehört auch der DOI (Digital Object Identifier). Ein DOI ist ein eindeutiger Bezeichner für digitale Objekte. In diesem Fall ist das Objekt eine Datensammlung. Die DOI werden zurzeit noch von Hand durch das Fachpersonal vergeben. Das Fachpersonal prüft zuvor die Metadaten

Präsentationsmöbel für Bibliotheken

Unsere Stufenpräsentier für AV-Medien passen für CD – DVD – Blu-Ray. Sie haben sich schon viele Jahre in vielen Bibliotheken bewährt. Bei der Verwendung von Weichhüllen sind die entsprechenden HADOS-Präsentier bestens geeignet.

Versetzbare Trenner ermöglichen die Präsentation von unterschiedlichen Produktbreiten.

Bei Wand- und Mittenpräsentationen ist eine durchge-

hende Stellmöglichkeit von Vorteil.

Die Warenträger können nach Ihren Wünschen und Bedürfnissen kombiniert werden.

Farbwünsche berücksichtigen wir gern. Die Montage ist einfach und problemlos.

Erfahren Sie weitere Details durch einen Kontakt zu uns.

Wir beraten Sie auch gern unverbindlich bei Ihnen vor Ort.



und das Format der Forschungsdaten und gibt dann die Daten zur Veröffentlichung frei.

Bislang ist noch offen, ob die Daten nachträglich verändert oder angereichert werden dürfen, sobald diese über eine DOI verfügen. Im Gegensatz zu anderen Objekten, wie z.B. Artikel, gibt es hierzu noch keine einheitlichen Richtlinien. Wir gehen aber davon aus, dass sich die Richtlinie durchsetzen wird, dass ein mit einem DOI versehenes Objekt nicht verändert werden darf und dass jede Änderung als neue Version gespeichert und mit einem neuen DOI versehen werden muss. In Tambora können Daten mit einem DOI daher nachträglich nicht mehr verändert werden. Stattdessen wird ein neues Datum mit Referenz auf das DOI-Datum angelegt.

Ausblick und Fazit

Die bisherige Resonanz ist durchweg positiv: Jede Woche gibt es Nutzer, die einen Zugang beantragen und es gibt auch Anfragen aus England, die gerne ihre Daten in Tambora ablegen wollen. Dies, obwohl Tambora noch lange nicht alle Funktionen zur Verfügung stellt, die sich die Wissenschaftler wünschen. Hierzu zählen zum Beispiel:

- Eine Möglichkeit, Laienforscher einzubinden („Citizen Science“)
- Automatische und halbautomatische Verfahren, welche manuelle Verfahren ergänzen
- Stärkere Verdrahtung mit den Repositorien der UB Freiburg

Tambora ist eine maßgeschneiderte Software: Versuche, vorgefertigte Software aus dem OpenSource-Bereich und anderen Projekten für die Bedürfnisse der Klimaforscher anzupassen, scheiterten am Anforderungsprofil. Der Ursprungsgedanke, Erfassungswerkzeug und Repository in einem System zu bündeln, erwies sich als Sackgasse. Dies wurde z.B. mit DSpace versucht, was letztendlich massive Änderungen am Softwarekern verlangte. Damit wurde es unmöglich, neue Versionen von DSpace nach den Änderungen einzuspielen.

Gleich von Beginn an wollten die Forscher, dass Tambora auch nach dem Projekt weiter besteht. Software, mit der die Rohdaten erzeugt wurden, ebenfalls nachhaltig zu sichern und zu pflegen, ist die natürliche Fortführung des Gedankens, neben der wissen-

schaftlichen Veröffentlichung auch die dazugehörigen Rohdaten zu archivieren. Dies ist für die UB Neuland und verlangt, dass hierfür auch Fachpersonal zur Verfügung gestellt wird, welches für die Nachhaltigkeit garantieren kann. Unklar ist, wer die Software weiterentwickeln soll und wie eine solche Weiterentwicklung aussehen soll. Eine Möglichkeit besteht darin, die technische Leitung langfristig bei einer zentralen Einrichtung zu etablieren. Bei Folgeprojekten gibt es dann einen oder mehrere Ansprechpartner, die dann die weitere Softwareentwicklung koordinieren. Dieser Weg wird an der UB Freiburg derzeit erprobt. ■

Literaturangaben und Referenzen

1. BOREL, Franck: „HISCLIDCore“, in Expressum Nr. 6 (2010) S. 16-19.
2. DEUTSCHE FORSCHUNGSGEMEINSCHAFT: Virtuelle Forschungsumgebung, in DFG-Vordruck 12-12-5/11, Bonn 2011.
3. MÜLLER, Bernd: „Stürmt den Elfenbeinturm“, in Technology Review 04.2012 (2012) S.48-51.
4. BÜTTNER, Stephan/ HOBÖHM, Hans-Christoph/ MÜLLER, Lars: Handbuch Forschungsdatenmanagement, Bad Honnef 2011.
5. TIB HANNOVER/ FIZ CHEMIE BERLIN/ UNIVERSITÄT PADERBORN: Konzeptstudie Vernetzte Primärdaten-Infrastruktur für den Wissenschaftler-Arbeitsplatz in der Chemie, Paderborn 2010.
6. LÖWER, Chris: „Auch du bist Astronom“, in Technology Review 12.2011 (2011) S.54-57
7. RIEMANN, Dirk: Methoden zur Klimarekonstruktion aus historischen Quellen am Beispiel Mitteleuropas, Diss. Freiburg 2010.
8. URL: <http://www.tambora.org> [12. Juli 2012].

Franck Borel

Systementwicklung
Universitätsbibliothek Freiburg
borel@ub.uni-freiburg.de

Heike Steller

Kartographie und Visuelle Kommunikation
Leibniz-Institut für Länderkunde Leipzig
H_Steller@ifl-leipzig.de