

GeoBib



GEOREFERENZIERTE
ONLINE-BIBLIOGRAPHIE FRÜHER
HOLOCAUST- UND
LAGERLITERATUR

**QUALITÄTSSICHERUNGSKONZEPT (M1.3R)
UND BERICHT ZU DURCHGEFÜHRTEN QUA-
LITÄTSSICHERUNGSMAßNAHMEN IM HIN-
BLICK AUF DIE ANNOTATION (M4.5R)**

INHALT

1. Überblick.....	3
2. Der GeoBib-Redaktionsprozess.....	3
2.1. Identifizierung relevanter Werke.....	3
2.2. Materialbeschaffung.....	4
2.3. Redaktionsprozess.....	4
2.4. Überprüfung der Annotationen auf den Einsatz in unterschiedlichen Bereichen.....	4
3. Durchgeführte computergestützte Qualitätssicherungsmaßnahmen bezüglich der gesammelten bibliographischen und Biographischen Daten.....	6
3.1. XML.....	6
3.2. MediaWiki.....	8
4. Referenzen.....	8

1. ÜBERBLICK

Qualitätssicherung ist ein Querschnittsaufgabenbereich und kann vielfältig gesehen bzw. gestaltet werden. Im GeoBib-Projekt zielt die Qualitätssicherung vor allem darauf ab, 1) die Qualität der Informations- und Wissensbestände, die im Laufe des Projekts gesammelt, verwaltet, und in einem Informationssystem verfügbar gemacht werden, auf einem hohen Niveau sicherzustellen und 2) ein qualitativ hochwertiges Informationssystem zu diesen Daten bereitzustellen. Der Schlüssel zu diesen Qualitätsanforderungen liegt in gut organisierten und sauber durchgeführten Prozessen, die sich nach Erkenntnissen des Qualitätsmanagements richten und auf vorherigen Erfahrungen der Beteiligten aufbauen. Eine wesentliche Aufgabe des GeoBib-Projektes ist es, für die zentralen Arbeitsprozesse der einzelnen Teilprojekte jeweils eine geeignete Struktur bzw. eine geeignete Vorgehensweise zu finden, anzuwenden und zu dokumentieren. GeoBib orientiert sich dabei an etablierten Modellen und speist sich aus den Erfahrungen der Beteiligten, etwa aus Erfahrungen im Bereich der bibliographischen Arbeit, aus früheren Editionsprojekten, enzyklopädischer Arbeit, und aus den Bereichen der Software-Entwicklung sowie Software-Qualitätssicherung. Im Projekt wird demnach dynamische Qualitätssicherung betrieben. Der vorliegende Bericht zur Qualitätssicherung enthält dementsprechend vor allem eine Dokumentation der Prozesse im GeoBib-Projekt.

In den folgenden Abschnitten wird zunächst der Redaktionsprozess beschrieben einschließlich der Erarbeitung maschinenlesbarer Annotationsdokumente. Anschließend wird die computergestützte Weiterverarbeitung der so entstandenen Informationsbestände erläutert, insbesondere deren Einspeisung in die Datenbank des Informationssystems und dabei durchgeführte Qualitätssicherungsmaßnahmen im Zuge der Abbildung auf das speziell für das Projekt entworfene Datenbankschema. Zudem wird kurz auf die Überprüfung der Annotationen mit Blick auf den Einsatz in der historischen und politischen Bildungsarbeit eingegangen werden.

2. DER GEOBIB-REDAKTIONSPROZESS

2.1. IDENTIFIZIERUNG RELEVANTER WERKE

In einem ersten Schritt wurden die einschlägigen Spezialbibliographien zum Themenfeld im Hinblick auf frühe Texte der Holocaust- und Lagerliteratur ausgewertet, des Weiteren die Nachschlagewerke zum Thema. Ergänzt wurden diese Recherchen durch eine gezielte Durchsicht der Fachliteratur, durch umfassende Befragung der Bibliothekssysteme sowie der Kataloge der Anbieter antiquarischer Bücher. Während des laufenden Arbeitsprozesses, vor allem

der Auswertung zeitgenössischer Zeitungen und Zeitschriften, kamen durch die dort abgedruckten Rezensionen und Inserate weitere Hinweise auf relevante Werke hinzu.

2.2.MATERIALBESCHAFFUNG

Die als projektrelevant identifizierten Werke der Holocaust- und Lagerliteratur wurden erfasst, aus den örtlich vorhandenen Bibliotheksbeständen (Universitätsbibliothek Gießen, Herder-Institut Marburg sowie Deutsche Nationalbibliothek Frankfurt) oder über das Fernleihsystem beschafft. Überdies wurden Werke, die auf diesen Wegen nicht bestellt werden konnten, antiquarisch erworben. Eine Vielzahl polnischsprachiger Titel, die über eine nationale oder internationale Fernleihe nicht zu beschaffen waren, konnte die Bibliothek des Herder Instituts in direktem Kontakt zur Biblioteka Narodowa in Warschau von dort beschaffen.

2.3.REDAKTIONSPROZESS

Nach Lektüre der Werke und dem Verfassen einer Inhaltszusammenfassung und -analyse, einer Kurzbiographie der Verfasser sowie einer Werkgeschichte durchlaufen die so erstellten Annotationen einen ersten Redaktionsprozess. Die Annotationen werden jeweils von zwei Mitarbeiterinnen und Mitarbeitern gelesen und korrigiert und mit den Bearbeitern besprochen. Die Korrekturen und Ergänzungen fügen die Bearbeiter anschließend in die Annotation ein. Zu einem späteren Zeitpunkt wird ein Redaktionsteam alle bis dahin verfassten Annotationen erneut einer Prüfung unterziehen.

2.4.ÜBERPRÜFUNG DER ANNOTATIONEN AUF DEN EINSATZ IN UNTERSCHIEDLICHEN BEREICHEN

Der Einsatz der Annotationen auf ihre Verwendung in der historischen und politischen Bildungsarbeit wurde unter anderem in Kooperation mit der Didaktik der Geschichte der JLU Gießen durchgeführt. So wurde das Projekt sowie eine Betaversion des Portals durch Annalena Schmidt auf dem Fachtag Geschichte der JLU Gießen vorgestellt. Bei diesem Tag handelt es sich um eine von der Universität organisierte Fortbildungsveranstaltung für Lehrerinnen und Lehrer. Bei einem Vortrag mit anschließender Diskussion konnten das zukünftige Portal zum einen einer potentiellen Zielgruppe – Multiplikatoren der historischen und politischen Bildungsarbeit – bekannt gemacht werden und zum anderen auf Fragen und Anregungen eingegangen werden. So wurde von den Anwesenden das Projekt sehr positiv aufgenommen und bekundet, dass man dies zukünftig auch im Unterricht bzw. im Zuge der Unterrichtsvorbereitung einsetzen wolle. Allerdings gab es auch Lehrerinnen und Lehrer, die beklagten, dass es

sich lediglich um eine Bibliographie handelt. Für eine breitere Wirkung dieser frühen Texte der Holocaust- und Lagerliteratur wurde eine „Online-Bibliothek“ gewünscht, da die Texte nur in sehr wenigen Bibliotheken vorhanden sind und diese allenfalls über die Fernleihe zu erhalten sind, was zwar möglich, aber im Zuge einer Unterrichtsvorbereitung oftmals nicht möglich ist. Zwar leuchtete das Argument des Urheberrechts als Schranke für die Veröffentlichung weiterer Teile der Werke ein, allerdings haben wir gerade diese Anregung im Hinterkopf behalten und prüfen nun, ob auf Grundlage des „Gesetzes über die Nutzung von verwaisten und vergriffenen Werken“ eine Bereitstellung eines Teilkorpus evtl. noch im Zuge der Projektlaufzeit und der Implementierung des Portals möglich ist.

Zudem wurde durch Annalena Schmidt ein Proseminar für angehende Lehrer und Lehrerinnen ebenfalls in Kooperation mit der Didaktik der Geschichte an der JLU Gießen in Person von Monika Rox-Helmer mit dem Titel *„Frühe Holocaust- und Lagerliteratur aus dem Projekt GeoBib als Quelle: Vorbereitung von Schullésungen zum Holocaust Gedenktag“* durchgeführt. Der Ankündigungstext der Veranstaltung lautete: *„Am 27. Januar jährt sich der Tag der Befreiung des Konzentrationslagers Auschwitz zum 70. Mal. Das soll in diesem Seminar zum Anlass genommen werden, um Lesungen in Schulklassen zum Thema Holocaust durchzuführen. Gleichzeitig soll damit eine Möglichkeit erprobt werden, diesen Gedenktag im schulischen Kontext zu begehen. Im Rahmen dieses Seminars werden die Lesungen vorbereitet, durchgeführt und reflektiert. Grundlage für die Lesungstexte soll das Projekt GeoBib sein. In diesem Projekt entsteht eine annotierte und georeferenzierte Online-Bibliographie für die frühe deutsche und polnische Holocaust- und Lagerliteratur. Im Seminar sollen Kriterien der Textauswahl für unterschiedliche Unterrichtssituationen erarbeitet und angewendet werden sowie die bisherigen Annotationen auf ihre Brauchbarkeit überprüft werden.“* Wie in dem letzten Satz der Ankündigung der Lehrveranstaltung bereits geschrieben, war neben der Organisation der Lesungen für den „Tag des Gedenkens an die Opfer des Nationalsozialismus“ eines der mit dem Proseminar intendierten Ziele die Evaluation der im Projekt angefertigten Annotationen für den Einsatz bei der Auswahl von Texten für die historische und politische Bildungsarbeit durch die angehenden Lehrerinnen und Lehrer. Die Studierenden wurden bei der Auswahl der Texte und damit auch bei der Evaluation der Annotationen durch eine erfahrene Studienrätin im Hochschuldienst – Monika Rox-Helmer – begleitet, die für die didaktischen Bestandteile des Seminars verantwortlich war. Annalena Schmidt war für die geschichtswissenschaftliche Begleitung des Seminars verantwortlich. Auf Grundlage von gut 30 der im Projekt erstellten Annotationen wählten die Studierenden die Texte aus, die sie folgend im Ganzen lasen und nach der Gesamtlektüre die Texte für die Lesung auswählten. Dabei

zeigte sich, dass die Annotationen der GeoBib einen guten Eindruck über die Inhalte und Besonderheiten der Primärtexte vermitteln und auf Grundlage der Zusammenfassungen eine Auswahl von Texten für den Einsatz in der historischen und politischen Bildungsarbeit erfolgen kann.

3. DURCHGEFÜHRTE COMPUTERGESTÜTZTE QUALITÄTSSICHERUNGSMABNAMEN BEZÜGLICH DER GESAMMELTEN BIBLIOGRAPHISCHEN UND BIOGRAPHISCHEN DATEN

Computergestützte Verfahren zur Qualitätssicherung in der Phase der Datenerhebung spielen insbesondere bei der Verwendung von XML und MediaWiki eine wichtige Rolle. Im Folgenden sollen die regelmäßig durchgeführten Arbeitsschritte bzw. Maßnahmen, die es erlauben konzeptionell Einfluss auf die Qualität zu nehmen, erläutert werden. Eine genauere Übersicht über den Arbeitsauflauf sowie die Verbindung von XML, Autorenoberfläche und Wiki findet sich in Entrup et al 2013.

3.1.XML

XML bietet über die Definition von Dokumentgrammatiken die Möglichkeit, die Eingabe und Form von Daten direkt zu bestimmen. Da im Projekt keine simplen DTDs (Documenttype declarations) verwendet wurden, sondern aussagekräftigere und besser konfigurierbare RelaxNG-Schemata erstellt und verwendet wurden, ist diese Möglichkeit sogar noch etwas ausgeschöpft worden.

Neben der Definition welche Elemente in welcher Reihenfolge oder Hierarchie verwendet werden müssen, bietet eine Grammatik wie RelaxNG die Möglichkeit, Datentypen zu definieren. Neben bekannten und weit verbreiteten Datentypen (Strings, Integers, float, date etc.), besteht die Möglichkeit mit Hilfe von regulären Ausdrücken die Eingaben von Daten zu definieren und einzuschränken. Eine Element-Definition, die von solchen Datentypen Gebrauch macht, schränkt die möglichen Eingabewerte ein. Dadurch kann sichergestellt werden, dass z.B. fehlerhafte Datumseingaben frühzeitig, schon bei der Eingabe, erkannt und vor allem vom System selbst bemängelt werden. Selbst definierte Dateitypen wurden besonders für die Eingabe von Preisen und Sprachen verwendet, um sie an das notwendige Format anzupassen.

Die Verwendung des oXygen-Autorenmodus bietet darüber hinaus die Möglichkeit, wiederkehrende eingaben zu automatisieren und somit möglich Fehlerquellen wie z.B. Tippfehler

o.ä. zu verhindern. Dies wird besonders bei der Eingabe von Verlinkungen zur Orts- und Personenseiten im Wiki angewendet. Die Links werden vom System generiert. Der User wählt den korrekten Link aus und das System generiert den notwendigen XML-Code.

Die für Laien geläufigere Darstellung der XML-Datei in Form einer Eingabemaske hilft weiterhin, einige Probleme, die bei der Arbeit mit XML auftreten können, zu vermeiden und führte im Projekt zu einer breiten und bereitwilligen Annahme von XML, da wenig Umstellungen zu Arbeiten mit anderen WYSIWYG-Editoren von Nöten war.

Um den sicheren Umgang mit XML aber auch mit dem Autoren-Modus, der zeitversetzt eingeführt wurde, zu gewährleisten, wurden verschiedene Workshops und Treffen abgehalten, die im Meilenstein M4.3V beschrieben wurden (http://geb.uni-giessen.de/geb/frontdoor.php?source_opus=10951&la=de).

Zur Überprüfung der Eingaben wurden folgende Skripte entwickelt, die Fehler und Unstimmigkeiten in Form einer verständlichen Liste anzeigen und somit die Fehlerfindung und -behebung erleichtern:

- `checkValidity.sh`: Kontrolliert jede Datei auf ihre Wohlgeformtheit sowie ihre Validität (Übereinstimmung mit der RelaxNG-Datei) und gibt eine Liste mit Fehler per Datei (sowie deren Autor) aus.
- `check_before_del.xquery`: Script das vom Wiki aus aufgerufen wird, bevor eine Seite dort gelöscht wird: es überprüft, dass keine XML-Datei auf diese Seite verlinkt.

Weitere kleinere Xquery Skripte wurden im Laufe des Projektes erstellt um Änderungen am Schema automatisch auf alle XML-Dateien anzuwenden, oder um trotz allem noch fehlerhafte (bzw. in der Form nicht gewünschte Eingaben) zu korrigieren.¹

Häufige oder wiederkehrende Fehler wurden regelmäßig ermittelt und besprochen. Dieses Feedback wurde dann wiederum auf die neuen Dateien angewendet und Probleme in der Kodierung fortan vermieden.

Ebenso wurde aber aus den Annotationen auch oft das Bedürfnis nach anderen oder genaueren Strukturen deutlich, die dann in einer neuen Version des Schemas mündeten.

¹ Die Anwender haben im Laufe des Projektes einige Kreativität gezeigt, wenn es darum ging, bestehende Elemente oder Strukturen für andere als dafür vorgesehene Zwecke zu nutzen.

3.2. MEDIA WIKI

Im MediaWiki-System können Daten zwar auch semi-strukturiert eingegeben werden, z.B. in Form von Tabellen o.ä., jedoch existiert keine Typisierung und eine Überprüfung der Eingabe ist nur schwer möglich.

Zur Strukturierung der Daten werden Infoboxen genutzt, die eine Eingabe in Tabellenform ermöglichen. Weiterhin können Daten an Templates übergeben und dort formatiert werden.

Die Überprüfung der Eingaben selbst erfordert allerdings einen Parsing-Schritt, der das Parsing im Zuge des Importes der Daten in unsere Datenbank vorwegnimmt und die Eingaben überprüft.

Zu diesem Zwecke wurde eine Reihe von Python Skripte geschrieben, die z.B. die Datumsangaben auf das von uns verwendete ISO-konforme Schema überprüft. Des Weiteren werden fehlende Eingaben oder fehlerhafte Verknüpfungen auf diese Art und Weise überprüft. Liste mit Fehlern nach Seiten sortiert können dann als Feedback an die Autoren zurückgegeben und von ihnen behoben werden.

Erst danach erfolgt der eigentliche Import auf den dann sauberen Daten.

4. REFERENZEN

Binder, Frank (2014) "From Collaborative Data Editing to Library Catalogues – Towards a 'Sharable Data Strategy' for the GeoBib Project" – Vortrag im Rahmen des Leipzig eHumanities Seminars am 8. Januar 2014. Folien: http://www.staff.uni-giessen.de/~g91214/Folien-GeoBib_Talk_EHum-Seminar-2013-14-FBinder-Export.pdf.

Abstract: <http://www.e-humanities.net/assets/seminar/2013/revised/2013-eHumSem-GER-Binder-Frank-GeoBib.pdf>.

Entrup, Bastian, Frank Binder und Henning Lobin (2013) "Extending the Possibilities for Collaborative Work with TEI/XML Through the Usage of a Wiki System". In: Proceedings of the 1st International Workshop on Collaborative Annotations in Shared Environment: metadata, vocabularies and techniques in the Digital Humanities (DH-CASE 2013), Florenz, Italien, 10 September 2013. <http://doi.acm.org/10.1145/2517978.2517988>.