

FACHHOCHSCHULE POTSDAM

MASTERARBEIT

---

Die digitale Edition der „preußischen  
Zeitungsberichte“: Evaluation von  
Editionswerkzeugen zur nutzergenerierten  
Transkription handschriftlicher Quellen

---

Fachbereich Informationswissenschaften

Master Informationswissenschaften

Profil: Records Management und digitale Archivierung

Eingereicht von: Jens Brokfeld

Matrikelnummer: 8508

Erstgutachter: Prof. Dr. Mario Glauert

Zweitgutachter: Prof. Dr. Felix Sasaki

Eingereicht am: 6. August 2012

Toujours je rêve d'une plume qui soit une seringue.

---

Jacques Derrida über das Schreiben

## **Zusammenfassung**

Die Erstellung digitaler Editionen auf Grundlage von Manuskripten aus den Beständen von Archiven, Bibliotheken und Museen erfordert Software mit Funktionen zur Transkription, Verwaltung und Präsentation der Textdaten. Die nutzergenerierte Transkription (Crowdsourcing) der historischen Quellen stellt wiederum spezifische Anforderungen an die Transkriptionswerkzeuge. In einer Nutzerumfrage werden diese Ansprüche sowie bestehende Erfahrungen in Bezug auf die Art der handschriftlichen Quellen, die eingesetzten Software-Werkzeuge, die Einbeziehung von Freiwilligen in die Codierung der Transkripte, die Schaffung von Anreizen zur Mitarbeit, die Qualitätssicherung, Funktionen zur Kommentierung der behandelten Texte, die Erstellung von Indizes, die Speicherung der Editionstexte in Formaten wie TEI, den Verweis auf Normdaten und Erschließungsinformationen sowie die Beteiligung an Web-Portalen erörtert. Die Auswahl der zu analysierenden Werkzeuge beruht auf den Kriterien Entwicklungsstand, Online-Werkzeuge und Crowdsourcing-Funktionen. Den Kriterien entsprechend wird die Software FromThePage, Refine!, Wikisource, Scriptor, T-PEN sowie Bentham Transcription Desk behandelt. Der Funktionsumfang wird anhand von Prüfkriterien der Zugänglichkeit, Auffindbarkeit, intellektuellen Redlichkeit und Gebrauchstauglichkeit betrachtet und verglichen. Bei der Evaluation gilt die Prämisse, dass Anpassungen der Software bei gleichzeitiger Erfüllung projektspezifischer Anforderungen möglichst gering gehalten werden sollten. Dementsprechend wird der Funktionsumfang der Werkzeuge in Bezug auf Anwenderkreise mit spezifischen Anforderungsprofilen evaluiert.

**Schlagworte:** Digitale Editionen, Transkription, Crowdsourcing, Transkriptionswerkzeuge

## **Abstract**

To create digital editions on the basis of manuscripts from archives, libraries and museums it is necessary for the software to come with functions for the transcription, management and presentation of textual data. Crowdsourcing the transcriptions moreover imposes specific requirements on the software. The requirements as well as experiences from other projects in relation to the type of manuscripts, the software, the participation of volunteers in text encoding, the motivation of transcribers, quality control, the integration of text commentary, generating indices, storing the text in formats like TEI, referencing authority files as well as participation in web portals are analysed on the basis of a survey. The choice of tools that are examined rests upon the following selection criteria: development stage, online tools and crowdsourcing functionality. Corresponding to the criteria, FromThePage, Refine!, Wikisource, Scripto and T-PEN are included in the analysis. Their functionality is analysed and compared in reference to test criteria of accessibility, findability, intellectual integrity and usability. The evaluation rests on the assumption that there should be as little software customization as possible while the project specific requirements are met. To that effect, software functionality is evaluated in relation to target groups with specific requirement profiles.

**Keywords:** Digital Editions, Transcription, Crowdsourcing, Transcription Tools

# Inhaltsverzeichnis

|          |   |           |
|----------|---|-----------|
| <b>1</b> | <b>Einleitung</b>   | <b>1</b>  |
| <b>2</b> | <b>Crowdsourcing und Editionswerkzeuge</b>                  | <b>5</b>  |
| <b>3</b> | <b>Umfrage zur Funktions- und Bedarfsanalyse</b>            | <b>9</b>  |
| 3.1      | Durchführung . . . . .                                      | 9         |
| 3.2      | Auswertung . . . . .  | 10        |
| <b>4</b> | <b>Auswahlkriterien</b>                                     | <b>24</b> |
| 4.1      | Checkliste für potentielle Transkriptionsprojekte . . . . . | 24        |
| 4.2      | Auswahl der Software für die weitere Analyse . . . . .      | 29        |
| <b>5</b> | <b>Prüfkriterien</b>  | <b>32</b> |
| 5.1      | Zugänglichkeit . . . . .                                    | 33        |
| 5.1.1    | Browser-Unterstützung . . . . .                             | 33        |
| 5.2      | Auffindbarkeit . . . . .                                    | 34        |
| 5.2.1    | Schnittstellen . . . . .                                    | 34        |
| 5.2.2    | Verweise auf Normdaten . . . . .                            | 34        |
| 5.3      | Langlebigkeit . . . . .                                     | 35        |
| 5.3.1    | Lizenz . . . . .  | 35        |
| 5.3.2    | Codierungsformat . . . . .                                  | 35        |
| 5.3.3    | Hosting . . . . .   | 36        |
| 5.4      | Intellektuelle Redlichkeit . . . . .                        | 36        |
| 5.4.1    | Textauszeichnung . . . . .                                  | 37        |
| 5.5      | Gebrauchstauglichkeit . . . . .                             | 37        |
| 5.5.1    | Transkriptionsmodus . . . . .                               | 38        |
| 5.5.2    | Textpräsentation und -Navigation . . . . .                  | 38        |
| 5.5.3    | Bearbeitungsstatistiken . . . . .                           | 38        |
| 5.5.4    | Benutzerverwaltung . . . . .                                | 38        |

|           |  |            |
|-----------|--|------------|
| <b>6</b>  | <b>Analyse und Evaluation</b>                                      | <b>40</b>  |
| 6.1       | Editionswerkzeuge . . . . .  | 40         |
| 6.1.1     | FromThePage . . . . .  | 40         |
| 6.1.2     | Refine! . . . . .  | 48         |
| 6.1.3     | Wikisource . . . . .   | 55         |
| 6.1.4     | Scripto . . . . .  | 62         |
| 6.1.5     | T-PEN . . . . .  | 69         |
| 6.1.6     | Bentham Transcription Desk . . . . .                               | 74         |
| <b>7</b>  | <b>Programmvergleich</b>   | <b>83</b>  |
| 7.1       | Programmvergleich anhand der Prüfkriterien . . . . .               | 83         |
| 7.2       | Programmvergleich anhand der Zielgruppe . . . . .                  | 88         |
| <b>8</b>  | <b>Fazit</b>   | <b>101</b> |
| <b>9</b>  | <b>Literaturverzeichnis</b>  | <b>104</b> |
| <b>10</b> | <b>Anhang</b>  | <b>116</b> |
| 10.1      | Umfrage: Textantworten . . . . .                                   | 116        |
| 10.1.1    | Art der Quellen . . . . .  | 116        |
| 10.1.2    | Eingesetzte Editoren . . . . .                                     | 119        |
| 10.1.3    | Konfrontation des Bearbeiters mit Quelltext . . . . .              | 121        |
| 10.1.4    | Anreize zur Mitarbeit an Transkriptionsprojekten . . . . .         | 123        |
| 10.1.5    | Technische Hilfsmittel zur Qualitätssicherung . . . . .            | 124        |
| 10.1.6    | Funktion zur text- oder sachkritischen Kommentierung . . . . .     | 126        |
| 10.1.7    | Speicherung in TEI . . . . .                                       | 128        |
| 10.1.8    | Erstellung von Indizes . . . . .                                   | 131        |
| 10.1.9    | Weitere Funktionen zur Textnavigation . . . . .                    | 132        |
| 10.1.10   | Verweise auf Normdaten oder Erschließungsinformationen . . . . .   | 133        |
| 10.1.11   | Datenlieferung an größere Portale . . . . .                        | 135        |
| 10.1.12   | Weitere Kommentare zu Anforderungen an Editionswerkzeuge . . . . . | 136        |
| 10.2      | Umfrage: Multiple-Choice-Antworten . . . . .                       | 138        |
| 10.3      | Liste der Transkriptionswerkzeuge . . . . .                        | 141        |

# Abbildungsverzeichnis

|      |  |    |
|------|--|----|
| 6.1  | FromThePage: Transkriptionsmodus . . . . .                               | 42 |
| 6.2  | FromThePage: Graph . . . . .   | 43 |
| 6.3  | FromThePage: Bearbeitungsstatistik . . . . .                             | 44 |
| 6.4  | Refine!: Transkriptionsmodus . . . . .                                   | 50 |
| 6.5  | Wikisource: Transkriptionsmodus . . . . .                                | 57 |
| 6.6  | Wikisource: Textdarstellung mit Marginalien . . . . .                    | 57 |
| 6.7  | Scripto: Komponenten . . . . .   | 64 |
| 6.8  | Scripto: Transkriptionsmodus . . . . .                                   | 64 |
| 6.9  | Scripto: Versionsgeschichte . . . . .                                    | 65 |
| 6.10 | T-PEN: Transkriptionsmodus . . . . .                                     | 70 |
| 6.11 | T-PEN: Werkzeugleiste . . . . .  | 71 |
| 6.12 | T-PEN: Projektmanagement . . . . .                                       | 72 |
| 6.13 | Bentham Transcription Desk: Transkriptionsmodus . . . . .                | 76 |
| 6.14 | Bentham Transcription Desk: Digitales Bentham Repositorium . . . . .     | 77 |
| 6.15 | Bentham Transcription Desk: Moderations-Workflow . . . . .               | 79 |
| 6.16 | Bentham Transcription Desk: Rangliste der aktivsten Bearbeiter . . . . . | 80 |

# Verzeichnis der Codebeispiele

|     |   |    |
|-----|---|----|
| 6.1 | Refine!: TEI-header . . . . .                   | 52 |
| 6.2 | Wikisource: Metadaten . . . . .                 | 59 |
| 6.3 | Bentham Transcription Desk: Metadaten . . . . . | 77 |



# Tabellenverzeichnis

|       |   |     |
|-------|---|-----|
| 2.1   | Rose Holley's checklist for crowdsourcing . . . . .                   | 6   |
| 7.1   | Programmvergleich: Teil 1 . . . . .                                   | 85  |
| 7.2   | Programmvergleich: Teil 2 . . . . .                                   | 87  |
| 7.3   | Zielgruppenspezifische Verteilung der Programmeigenschaften . . . . . | 91  |
| 10.24 | Mutiple Choice: Ergebnisse Deutsch und Englisch . . . . .             | 141 |
| 10.25 | Liste der Transkriptionswerkzeuge . . . . .                           | 147 |

# Diagrammverzeichnis

|  |     |
|--|-----|
| 7.1 FromThePage: Zielgruppenspezifische Eigenschaften . . . . .                | 92  |
| 7.2 Refine!: Zielgruppenspezifische Eigenschaften . . . . .                    | 93  |
| 7.3 Wikisource: Zielgruppenspezifische Eigenschaften . . . . .                 | 95  |
| 7.4 Scripto: Zielgruppenspezifische Eigenschaften . . . . .                    | 97  |
| 7.5 T-PEN: Zielgruppenspezifische Eigenschaften . . . . .                      | 99  |
| 7.6 Bentham Transcription Desk: Zielgruppenspezifische Eigenschaften . . . . . | 100 |

# Abkürzungsverzeichnis

|               |   |
|---------------|---|
| AGPL .....    | Affero General Public License                             |
| API .....     | Application Programming Interface                         |
| CMS .....     | Content Management System                                 |
| CSS .....     | Cascading Stylesheets                                     |
| DCB .....     | Dictionary of Canadian Biography                          |
| DFG .....     | Deutsche Forschungsgemeinschaft                           |
| DNB .....     | Dictionary of National Biography                          |
| EAD .....     | Encoded Archival Description                              |
| ECL .....     | Educational Community License                             |
| EPL .....     | Eclipse Public License                                    |
| ePub .....    | Electronic Publication                                    |
| ESTC .....    | English Short Title Catalogue                             |
| FTP .....     | File Transfer Protocol                                    |
| GKD .....     | Gemeinsame Körperschaftsdatei                             |
| GND .....     | Gemeinsame Normdatei                                      |
| GPL .....     | General Public License                                    |
| GUI .....     | Graphical User Interface                                  |
| HTML .....    | Hypertext Markup Language                                 |
| HTTP .....    | Hypertext Transfer Protocol                               |
| JSON .....    | JavaScript Object Notation                                |
| LAMP .....    | Linux Apache HTTP Server MySQL PHP                        |
| LC .....      | Library of Congress                                       |
| LGPN .....    | Lexicon of Greek Personal Names                           |
| MAB .....     | Maschinelles Austauschformat für Bibliotheken             |
| MARC .....    | Machine-Readable Cataloging                               |
| MIT .....     | Massachusetts Institute of Technology                     |
| NDB .....     | Neue Deutsche Biographie                                  |
| OAI-PMH ..... | Open Archives Initiative Protocol for Metadata Harvesting |

|              |   |
|--------------|---|
| OCR .....    | Optical Character Recognition   |
| OPAC .....   | Online Public Access Catalogue  |
| PDF .....    | Portable Document Format  |
| PND .....    | Personennamendatei  |
| RDF .....    | Resource Description Framework  |
| RTF .....    | Rich Text Format  |
| TGN .....    | Thesaurus of Geographic Names   |
| URL .....    | Uniform Resource Locator  |
| VD17 .....   | Verzeichnis der im deutschen Sprachraum erschienenen Drucke des<br>17. Jahrhunderts |
| VIAF .....   | Virtual International Authority File  |
| W3C .....    | World Wide Web Consortium   |
| WYSIWYG .... | What You See Is What You Get  |
| XHTML .....  | eXtensible HyperText Markup Language  |
| XQuery ..... | XML Query Language  |
| XSLT .....   | Extensible Stylesheet Language Transformations                                      |

# 1 Einleitung

Die Anregung für das Thema der vorliegenden Arbeit gaben Planungen zur Erstellung einer digitalen Edition für den Archivbestand der Immediats-Zeitungsberichte des Potsdamer Regierungspräsidenten von 1855/1867 bis 1890/1914 im Brandenburgischen Landeshauptarchiv<sup>1</sup>, den sogenannten preußischen Zeitungsberichten. Die Transkription der handschriftlichen Berichte sollte mit Mitteln des Crowdsourcing vorgenommen werden und als Ergebnis wurde eine digitale Edition mit Indizes zu Ortsnamen, Personennamen etc. sowie Anmerkungen und Verweisen auf korrespondierende archivalische Quellen angesetzt. Die Evaluation geeigneter Software zur Realisierung dieser Zielstellung war das Thema dieser Arbeit bei der Einreichung des Titels zu Beginn der Bearbeitungszeit. Die Projektfinanzierung für die digitale Edition der preußischen Zeitungsberichte konnte allerdings nicht gesichert werden und deshalb wurde die Zielgruppe der hier vorgestellten Evaluation von der genannten Edition auf vergleichbare Projekte ausgeweitet. Der ursprüngliche Haupttitel „Die digitale Edition der preußischen Zeitungsberichte“ wird aus formellen Gründen beibehalten.

Das zentrale Thema der Arbeit besteht in einer Evaluation der Möglichkeiten zur Erstellung digitaler Editionen für Gedächtnisorganisationen mit Software-Werkzeugen zur nutzergenerierten Transkription handschriftlicher Quellen. Zu der anvisierten Zielgruppe zählen Archive, Bibliotheken, Museen und vergleichbare Einrichtungen, die ein Transkriptionsprojekt mit Hilfe von Crowdsourcing initiieren wollen. Die Problemstellung ist durch die Anforderungen an die Funktionen der digitalen Edition auf der einen- und das durch vorhandene Werkzeuge Realisierbare auf der anderen Seite gekennzeichnet. Dabei soll der Aufwand für Anpassungen der Software bei gleichzeitiger Erfüllung der projektspezifischen Anforderungen möglichst gering gehalten werden.

Die Betrachtung der Programme erfolgt vor dem Hintergrund, dass sich Gedächtnisorganisationen zur Online-Publikation ihrer Bestände vermehrt mit dem Thema der digitalen Editionen beschäftigen. Zur Transkription der handschriftlichen Quellen in diesen

---

<sup>1</sup>Vgl. Brandenburgisches Landeshauptarchiv. *Brandenburgisches Landeshauptarchiv*. URL: <http://www.landeshauptarchiv-brandenburg.de/> (besucht am 16.07.2012).

Beständen bietet sich die Einbeziehung Freiwilliger an. Eine grundlegende Anforderung an die Software besteht demnach in Funktionen zur Präsentation der Inhalte im Rahmen einer digitalen Edition sowie zur nutzergenerierten Transkription der Manuskripte. Die in Frage kommenden Werkzeuge decken unterschiedliche Facetten beider Bereiche ab. Unter dem allgemeineren Begriff der Editionswerkzeuge wird hier diejenige Software verstanden, die zur Erstellung und Publikation digitaler Editionen eingesetzt werden kann. Letztere umfasst auch die Transkriptionswerkzeuge, wobei in diesem Fall das Augenmerk auf den konkreten Funktionen zur Transkription wie dem Editor zur Eingabe der Texte etc. liegt.

Zur Analyse der Anforderungen an Editionswerkzeuge sowie an ihre bereits vorhandenen oder für die Zukunft wünschenswerten Funktionen werden die Antworten aus einer Online-Umfrage ausgewertet. Aus der Umfrage sowie Recherchen zu aktuellen Transkriptionsprojekten wird der aktuelle Stand der Bemühungen um Crowdsourcing-Projekte deutlich. Die behandelten Aspekte umfassen die Art der handschriftlichen Quellen, die eingesetzten Software-Werkzeuge, Fragen nach der Einbeziehung von Freiwilligen in die Codierung der Transkriptionstexte, die Schaffung von Anreizen zur Mitarbeit, die Qualitätssicherung, Funktionen zur Kommentierung der behandelten Texte, die Speicherung der Editionstexte in Formaten wie TEI, die Erstellung von Indizes, den Verweis auf Normdaten und Erschließungsinformationen sowie die Beteiligung an Web-Portalen durch Datenlieferung bzw. Datenaustausch.

Die Möglichkeit zur Transkription der handschriftlichen Quellen durch verschiedene (freiwillige) Bearbeiter nimmt bei der Betrachtung eine herausgehobene Stellung ein. Erfahrungen aus aktuellen Transkriptionsprojekten zeigen, dass die Einarbeitungszeit in die Bedienung des Systems möglichst kurz sein und der Benutzer sich ohne weitergehende technische Kenntnisse auf die Transkription der Inhalte konzentrieren können sollte. Eine tiefgehende semantische Erschließung des Editionstextes ist, basierend auf den Erfahrungen aus aktuellen Projekten, durch den anvisierten heterogenen Nutzerkreis nicht zu leisten. Dennoch existieren verschiedene Hilfsmittel zur Erleichterung der Textcodierung für den Bearbeiter etwa durch den Einsatz eines WYSIWYG-Editors, die automatisierte Generierung von flachen XML-Dokumenten z. B. im TEI-Standard oder den Einsatz einer vereinfachten Auszeichnungssprache bzw. einer stark reduzierten Anzahl möglicher Auszeichnungselemente.

Die Auswahlkriterien für die näher zu betrachtenden Programme orientieren sich an den Bedürfnissen der Zielgruppe von Gedächtnisorganisationen. Aus der Umfrage und der Recherche zu aktuellen Transkriptionsprojekten wird deutlich, dass es sich um web-

basierte Werkzeuge auf einem möglichst hohen Entwicklungsstand in Bezug auf Benutzeroberfläche und Dokumentation handeln sollte. Weitere Merkmale wie die Verwendung bestimmter Standards zur Textauszeichnung oder der Unterstützung bestimmter Kommentierungs- bzw. Indexierungsfunktionen werden bei der Auswahl nicht betrachtet. Diese spielen vielmehr bei der Entscheidung eines konkreten Projekts für ein bestimmtes Werkzeug eine Rolle und hängen von dessen spezifischen Anforderungen ab. Dementsprechend werden diese Aspekte in die Evaluation der einzelnen Programme einbezogen und können an dieser Stelle den Ausschlag für eine Empfehlung für Projekte mit einem bestimmten Anforderungsprofil geben. In den wiederum aus der Literaturrecherche und der Umfrage abgeleiteten Prüfkriterien werden einige grundlegende Funktionen und Konzeptionsmerkmale hervorgehoben. Diese bilden den Hintergrund für die Analyse und die Evaluation der ausgewählten Programme.

In die Evaluation aufgenommen wird das Transkriptionswerkzeug FromThePage<sup>2</sup>, der Editor Refine!<sup>3</sup>, das Projekt Wikisource<sup>4</sup>, das Transkriptionsmodul Scripto<sup>5</sup> im Zusammenhang mit dem System Omeka<sup>6</sup>, welches zur Verwaltung digitaler Sammlungen eingesetzt wird, das Transkriptionswerkzeug T-PEN<sup>7</sup> sowie das Programm Bentham Transcription Desk<sup>8</sup>. FromThePage ist eine freie Software zur Transkription und Präsentation von Manuskripten im Internet. Refine! bietet eine Arbeitsumgebung zur Transkription und Indexierung von Quellenmaterial an und wurde schon für vergleichbare Projekte durch die Firma 3pc angepasst. Wikisource bietet Möglichkeiten zur Transkription und Veröffentlichung von Quellenmaterial, das entweder urheberrechtsfrei ist oder unter einer freien Lizenz steht. Scripto bietet Funktionen zur Transkription digitaler Objekte als Modul eines bestehenden CMS. T-PEN erlaubt die zeilenbasierte Transkription von Handschriften in Kombination mit einer Funktion zum Projektmanagement und dem Export der in TEI codierten Editionstexte. Der Bentham Transcription Desk schließlich verfügt in seinem Editor über eine Werkzeugleiste zur Codierung ausgewählter TEI-Tags.

---

<sup>2</sup>Vgl. Ben W. Brumfield. *FromThePage*. URL: <http://beta.fromthepage.com/> (besucht am 16.06.2012).

<sup>3</sup>Vgl. Staatsbibliothek zu Berlin, Humboldt-Universität zu Berlin und 3pc GmbH Neue Kommunikation. *Refine!*

<sup>4</sup>Vgl. Wikimedia Foundation. *Wikisource*. URL: [http://en.wikisource.org/wiki/Main\\_Page](http://en.wikisource.org/wiki/Main_Page) (besucht am 16.06.2012).

<sup>5</sup>Vgl. Center for History and New Media. *Scripto*. URL: <http://scripto.org/> (besucht am 16.06.2012).

<sup>6</sup>Vgl. Roy Rosenzweig Center for History and New Media. *Omeka*. URL: <http://omeka.org/> (besucht am 16.06.2012).

<sup>7</sup>Vgl. Saint Louis University, Center for Digital Theology. *T-PEN*. 2012. URL: <http://digital-editor.blogspot.de/> (besucht am 16.06.2012).

<sup>8</sup>Vgl. University of London Computer Centre, UCL Bentham Project. *Bentham Transcription Desk*. 2012. URL: <http://www.ucl.ac.uk/transcribe-bentham/> (besucht am 16.06.2012).

In einem abschließenden Vergleich werden die Programme anhand der Prüfkriterien in kurzgefasster Form gegenübergestellt. Ferner gibt eine Zielgruppenanalyse Aufschluss über die Schwerpunktsetzung der Werkzeuge und zeigt ihre Eignung für die Gruppen Wissenschaft, Familiengeschichtsforschung, Archive, Bibliotheken und Museen auf.



## 2 Crowdsourcing und Editionswerkzeuge

Die Beteiligung Freiwilliger an der Transkription handschriftlicher Quellen fällt unter den Begriff Crowdsourcing. Die folgende Definition unterstreicht die beiderseitigen Vorteile solcher Projekte für Nutzer und Crowdsourcer:

Crowdsourcing is a type of participative online activity in which an individual, an institution, a non-profit organization, or company proposes to a group of individuals of varying knowledge, heterogeneity, and number, via a flexible open call, the voluntary undertaking of a task. The undertaking of the task, of variable complexity and modularity, and in which the crowd should participate bringing their work, money, knowledge and/or experience, always entails mutual benefit. The user will receive the satisfaction of a given type of need, be it economic, social recognition, self-esteem, or the development of individual skills, while the crowdsourcer will obtain and utilize to their advantage that what the user has brought to the venture, whose form will depend on the type of activity undertaken.<sup>9</sup>

Man könnte hinzufügen, dass der Crowdsourcer nicht nur von den Früchten der Arbeit der Nutzer profitiert, sondern auch etwas für die positive öffentliche Wahrnehmung der eigenen Institution tut sowie das Bewusstsein der Öffentlichkeit für den Wert der kulturellen Überlieferung stärkt<sup>10</sup>. Bekannte Beispiele für erfolgreiche Crowdsourcing-Projekte sind z. B. Wikipedia<sup>11</sup> und das Australian Newspaper Digitisation Program<sup>12</sup>. Im Rah-

---

<sup>9</sup>Estellés-Arolas und Guevara, „Towards an integrated crowdsourcing definition“, S. 9-10.

<sup>10</sup>Vgl. Rose Holley. „Crowdsourcing: How and Why Should Libraries Do It?“ In: *D-Lib Magazine* 16.3/4 (2012). URL: <http://www.dlib.org/dlib/march10/holley/03holley.html> (besucht am 05.07.2012).

<sup>11</sup>Vgl. Wikipedia Community. *Homepage von Wikipedia*. URL: <http://www.wikipedia.org/> (besucht am 05.07.2012).

<sup>12</sup>Vgl. National Library of Australia. *Australian Newspaper Digitisation Program*. URL: <http://www.nla.gov.au/ndp/> (besucht am 05.07.2012).

| The Thing     | The System         | The People   | The Content     |
|---------------|--------------------|--------------|-----------------|
| Clear Goal    | Easy and Fun       | Acknowledge  | Interesting     |
| Big Challenge | Reliable and Quick | Reward       | New             |
| Progress      | Intuitive          | Team Support | Lots            |
| Results       | Options            | Trust        | History/Science |

Tabelle 2.1: Rose Holley’s checklist for crowdsourcing

men von Transkriptionsprojekten setzen etwa FamilySearch Indexing<sup>13</sup> und Transcribe Bentham<sup>14</sup> Crowdsourcing ein. Es ist nicht das Ziel dieser Arbeit auf verschiedene Projekte vertieft einzugehen, aber im folgenden werden einige Projekte im Zusammenhang mit der verwendeten Software kurz vorgestellt.

Die folgenden Ratschläge<sup>15</sup> sowie die Checkliste (siehe Tabelle 2.1 Seite 6) für Crowdsourcing-Projekte stellen konzeptionelle Ratschläge dar, welche sich auch auf die Anforderungen an die zu verwendende Software auswirken. Sie werden in den Kriterien zur Auswahl und Prüfung der hier behandelten Programme wieder aufgenommen.

1. Have a transparent and clear goal on your home page (which goal MUST be a BIG challenge).
2. Have a transparent and visible chart of progress towards your goal.
3. Make the overall environment easy to use, intuitive, quick and reliable.
4. Make the activity easy and fun.
5. It must be interesting.
6. Take advantage of transitory and topical events if they help you.
7. Keep the site active by addition of new content/work.
8. Give volunteers options and choices
9. Make the results/outcome of your work transparent and visible.
10. Let volunteers identify and make themselves visible if they want acknowledgement.

---

<sup>13</sup>Vgl. FamilySearch. *FamilySearch Indexing*. URL: <https://familysearch.org/volunteer/indexing> (besucht am 05.07.2012).

<sup>14</sup>Vgl. University College London. *Transcribe Bentham*. URL: <http://www.ucl.ac.uk/Bentham-Project/> (besucht am 05.07.2012).

<sup>15</sup>Vgl. Holley, „Crowdsourcing: How and Why Should Libraries Do It?“

11. Reward high achievers by having ranking tables and encourage competition.
12. Give the volunteers an online team/communication environment to build a dynamic, supportive team environment.
13. Treat your 'super' volunteers with respect and listen to them carefully.
14. Assume volunteers will do it right rather than wrong.

Neben den aufgezählten allgemeinen Hinweisen, die sich auf die Funktionalität der verwendeten Software wie Usability, Auswertung von Statistiken zu den von Freiwilligen transkribierten Seiten und Möglichkeiten zur Kommunikation innerhalb des Online-Projekts beziehen, kommen für Editionsprojekte noch spezifische Funktionen zur Präsentation der Texte als Ergebnis des Projekts und zum Transkriptionsmodus hinzu. Dabei zeigt die Realität der vorhandenen Software-Lösungen, dass die technischen Entwicklungen in vielen Fällen noch nicht mit der Menge der Anforderungen Schritt halten können. Am zweiten Februar 2011 prognostizierte Ben W. Brumfield in seinem Blog "Collaborative Manuscript Transcription"<sup>16</sup> die Entwicklung auf dem Markt für Transkriptions-Software: "We will not see the development of a single tool that supports transcription of both structured and free-form manuscripts, nor both paleographic and semantic annotation in 2011"<sup>17</sup>. Brumfield begründet seine skeptische Haltung im folgenden so: "The field is too young and fragmented – most toolmakers have enough work providing the basic functionality required by their own manuscripts"<sup>18</sup>. Die meisten Software-Werkzeuge werden für einen bestimmten Anwendungsfall entwickelt. Ihre Funktionen sind dann gezielt auf die Anforderungen eines konkreten Projekts und die Eigenschaften der betreffenden Manuskripte ausgelegt. Das Ziel der Interoperabilität sollte bei der Weiterentwicklung dieser Systeme nicht aus den Augen verloren werden, auch wenn beim jetzigen Stand die projektspezifische Anpassung noch einen großen – und für die Realisierung spezifischer Anforderungen auch unvermeidbaren – Aufwand bedeutet.

Die automatische Schrifterkennung (OCR) stellt derzeit für die hier betrachteten Anwendungsfälle komplexer Dokumentstrukturen und idiosynkratischer Handschriften wie sie sich in den Beständen vieler Gedächtnisorganisationen finden keine Konkurrenz dar.

---

<sup>16</sup>Vgl. Ben W. Brumfield. *Collaborative Manuscript Transcription*. Blog. URL: <http://manuscripttranscription.blogspot.de/> (besucht am 09.05.2012).

<sup>17</sup>Vgl. Ben W. Brumfield. *2010: The Year of Crowdsourcing Transcription*. Blog-Eintrag. 2. Feb. 2011. URL: <http://manuscripttranscription.blogspot.de/2011/02/2010-year-of-crowdsourcing.html> (besucht am 09.05.2012).

<sup>18</sup>Vgl. ebd.

Eine Ablösung der manuellen Transkription historischer Manuskripte durch automatisierte Verfahren erscheint für die nähere Zukunft als unrealistisch. So kommentiert Richard Entlich den Stand der Technik folgendermaßen: “Unconstrained machine translation of handwriting appears particularly far off, and may be unachievable. Even a less ambitious goal, such as software to reliably create partial indexes from good quality single author material, is unlikely to be met within the next several years.”<sup>19</sup> Bianca Feldmann identifiziert stark strukturierte und syntaktisch einfache Dokumente als mögliche Kandidaten für die Handschriftenerkennung: „Es gibt auch schon erste kommerzielle Produkte auf dem Markt, z. B. Programme für handgeschriebene Buchstaben in Formularen. Die letzten Forschungen haben gezeigt, dass Handschriftenerkennung machbar ist für kleine Lexika und/oder wenn die Sätze syntaktisch sehr einfach sind.“<sup>20</sup> Da es sich bei den meisten Manuskripten in Archiven und Bibliotheken um wesentlich komplexere Dokumente handelt, ist die manuelle Transkription zur Erstellung digitaler Editionen momentan unumgänglich.

---

<sup>19</sup>Vgl. Richard Entlich. „Handwriting Recognition for Historical Documents“. In: *RLG DigiNews* 8.1 (2004). URL: <http://chnm.gmu.edu/digitalhistory/links/cached/chapter3/link3.33b.RLGhandwriting.html> (besucht am 14.05.2012).

<sup>20</sup>Vgl. Bianca Feldmann. „OCR von Handschriften. Ein Forschungsüberblick“. In: *Fundus – Forum für Geschichte und ihre Quellen* (1), S. 107–143. URL: <http://webdoc.gwdg.de/edoc/p/fundus/1/feldmann.pdf> (besucht am 14.05.2012), S. 130.

# 3 Umfrage zur Funktions- und Bedarfsanalyse

## 3.1 Durchführung

Ziel der Umfrage ist die Analyse des Bedarfs nach einem Transkriptionswerkzeug in Archiven, Bibliotheken, Museen und Wissenschaftseinrichtungen sowie die Eingrenzung der für diese Arbeit anvisierten Anwendungsfälle. Übereinstimmungen in den Anforderungen an die Software sollen verglichen und wünschenswerte Entwicklungstendenzen aufgezeigt werden. Die Umfrage wurde mit SurveyMonkey<sup>21</sup> erstellt.

In der Mehrzahl der Fragen geht es um qualitative Aussagen zu Stand und Perspektiven aktueller Transkriptionsprojekte. Mit Ja oder Nein zu beantwortende Fragen dienen in der Hauptsache dem Zweck dem Befragten in Abhängigkeit von seinen Antworten bestimmte Fragen zu stellen, die genauer auf sein Projekt zugeschnitten sind („Fragenlogik“). So werden bei einer positiven Beantwortung der Frage „Arbeiten Sie in Ihrer Institution an Projekten zur Erstellung digitaler Editionen?“ weitere Fragen zu der konkreten Edition und dem verwendeten Transkriptionsmodul vorgelegt. Bei einer negativen Beantwortung werden ähnliche Fragen in Bezug auf geplante oder wünschenswerte Transkriptionsprojekte in der betreffenden Einrichtung gestellt.

Der Fragebogen wurde aktiv an einzelne Adressaten versandt, die sich mit Transkriptionsprojekten beschäftigen. Darüber hinaus wurde ein Aufruf zur Teilnahme an der Umfrage an E-Mail-Listen wie die Archivliste<sup>22</sup>, die Digital Humanities Liste für den Berliner Raum<sup>23</sup>, die öffentliche Diskussionsliste der Text Encoding Initiative<sup>24</sup> sowie

---

<sup>21</sup>Vgl. SurveyMonkey. *Survey Monkey*. URL: <http://de.surveymonkey.com/> (besucht am 24.06.2012).

<sup>22</sup>Vgl. Archivschule Marburg. *Archivliste*. [archivliste@lists.uni-marburg.de](mailto:archivliste@lists.uni-marburg.de). E-Mail-Liste.

<sup>23</sup>Vgl. Berlin-Brandenburgische Akademie der Wissenschaften. *Digital Humanities Liste Berlin*. [dhberlin@bbaw.de](mailto:dhberlin@bbaw.de). E-Mail-Liste.

<sup>24</sup>Vgl. Text Encoding Initiative. *Text Encoding Initiative Public Discussion List*. [tei-l@listserv.brown.edu](mailto:tei-l@listserv.brown.edu). E-Mail-Liste.

die Liste europäischer Bibliothekare mit dem Arbeitsschwerpunkt Handschriften<sup>25</sup> versandt. Zudem wurde der Aufruf in Web-Portalen wie Archivalia<sup>26</sup> und H-Soz-u-Kult<sup>27</sup> veröffentlicht. Ebenso verbreitete er sich ohne aktive Bemühungen des Autors etwa über Twitter, auf Facebook oder über den Blog Collaborative Manuscript Transcription<sup>28</sup>. Die Umfrage wurde in einer deutschen- und einer englischen Sprachversion bereitgestellt.

## 3.2 Auswertung

Insgesamt sind 70 abgeschlossene Fragebögen bei 210 gestarteten Umfragen eingegangen (33 % abgeschlossene Fragebögen). Die ersten Einladungen zur Teilnahme an der Umfrage wurden am 24.05.2012 verschickt und bis zum 08.06.2012 konnten die Fragebögen ausgefüllt werden. Die englische Sprachversion wurde von 31 % der Befragten angewählt und die restlichen 69 % entfallen auf die deutsche Sprachversion. Die gesammelten Textantworten werden im Anhang (Absatz 10.1 Seite 116) wiedergegeben. Die Ergebnisse der Multiple-Choice-Fragen sind ebenfalls im Anhang (Absatz 10.2 Seite 138) dargestellt.

### 1. *Arbeiten Sie in Ihrer Institution an Projekten zur Erstellung digitaler Editionen?*

Diese Frage zielt, wie oben beschrieben, auf eine Einteilung in bereits vorhandene und geplante oder wünschenswerte Editionsprojekte ab. Zusammen in der englischen und deutschen Sprachversion gaben 67 % der Befragten an, dass sie in ihrer Institution an Projekten zur Erstellung digitaler Editionen arbeiten. 54 % der Befragten, welche die deutsche Sprachversion genutzt haben, beschäftigen sich schon mit einer konkreten digitalen Edition, während der Rest sich noch im Planungsstadium befindet. Bei der englischen Sprachversion dagegen arbeiten 92 % schon an digitalen Editionen und lediglich 8 % der Befragten verneinten diese Frage. Dieses Ergebnis ist vermutlich auf die Verteilung der Umfrage über die englischsprachige E-Mail-Liste der Text Encoding Initiative auf der einen- und die deutschsprachige Archivliste auf der anderen Seite zurückzuführen. Die Abonnenten der E-Mail-Liste

---

<sup>25</sup>Vgl. Association of European Research Libraries. *Discussion list of European Manuscript Librarians*. E-Mail-Liste.

<sup>26</sup>Vgl. Klaus Graf. *Archivalia*. Blog. URL: <http://archiv.twoday.net/> (besucht am 05.06.2012).

<sup>27</sup>Vgl. Humboldt-Universität zu Berlin. *H-Soz-u-Kult*. URL: <http://hsozkult.geschichte.hu-berlin.de/> (besucht am 05.06.2012).

<sup>28</sup>Vgl. Brumfield, *Collaborative Manuscript Transcription*.

der TEI sind zu einem großen Teil mit Editionsprojekten im Rahmen geisteswissenschaftlicher Fakultäten und Wissenschaftseinrichtungen beschäftigt und setzen sich intensiv mit diesem Thema auseinander, während die Empfänger der Archivliste sich zum größten Teil aus deutschsprachigen Archivaren zusammensetzt, für die Editionsprojekte eher in der Peripherie ihres Aufgabenspektrums angesiedelt sind. Nicht alle Archive besitzen die Ressourcen, um digitale Editionen erstellen zu können und sie befinden sich häufig noch in der Planungsphase bzw. besitzen wertvolle Bestände für die ein Editionsprojekt als wünschenswert erscheint. Die o. g. Ergebnisse könnten diesen Tatbestand widerspiegeln. Auch die folgenden Ergebnisse der Umfrage sprechen (mit einigen Ausnahmen) für eine solche Schlussfolgerung. Die aus diesem Tatbestand erwachsenden unterschiedlichen Bedürfnisse der Befragten in Bezug auf Editionswerkzeuge werden daraufhin genauer analysiert.

#### 2. *Um was für eine Art von Quellen handelt es sich bei Ihrer digitalen Edition?*

Die genannten Quellen stellen ein große Bandbreite an Dokumenten dar. Vereinzelt wurden auch Druckerzeugnisse bzw. maschinenschriftliche Unterlagen genannt, aber die Mehrzahl der Befragten beschäftigt sich mit Transkriptionsprojekten zu handschriftlichen Quellen. Bei Druckerzeugnissen wie Zeitungen oder Büchern müssten die Möglichkeiten von Schrifterkennungssoftware geprüft werden, bevor man auf die Transkription zurückgreift. Das Thema dieser Arbeit bezieht sich jedoch auf handschriftliche Quellen, deren Transkription zum jetzigen Zeitpunkt und auch in absehbarer Zukunft unumgänglich sein wird (siehe Kapitel 2 Seite 5).

Die genannten Manuskripte reichen von recht allgemeinen Beschreibungen wie Korrespondenz, Archivgut, Akten und Urkunden bis hin zu Beschreibungen mit Dokumenttyp, Autor und Entstehungszeitraum wie der erweiterte Nachlass des Medizinnobelpreisträgers Emil von Behring (1854–1917) und der Behringwerkkorrespondenz (1904–1917) (Tabelle 10.1 Seite 118).

Häufig genannt wurden mittelalterliche Urkunden. Für diesen Dokumenttyp existieren spezialisierte Projekte und Werkzeuge wie das Urkundenportal Monasterium<sup>29</sup> oder das Transkriptionswerkzeug T-PEN<sup>30</sup>, das für mittelalterliche Hand-

---

<sup>29</sup>Vgl. Monasterium-Konsortium. *Monasterium.net*. URL: <http://monasterium.net> (besucht am 13.06.2012).

<sup>30</sup>Vgl. Center for Digital Theology. *T-PEN*. URL: <http://t-pen.org/TPEN/> (besucht am 13.06.2012).

schriften konzipiert wurde. Die Digitale Westfälische Urkunden-Datenbank<sup>31</sup> bietet Möglichkeiten zur Online-Publikation von Urkundenregesten.

Einen weiteren häufig genannten Dokumenttyp stellen Briefe und Tagebücher dar, die in der Regel weniger stark strukturiert sind als mittelalterliche Urkunden. Je nach dem für das einzelne Projekt geforderten Detailgrad bei der Erfassung von Textmerkmalen für die Aufbereitung in der Edition kann hier ein Werkzeug mit grundlegenden Funktionen etwa zur Erstellung von Namensindizes ausreichen oder es müssen spezielle Anpassungen zur Erfassung und späteren Präsentation der Textmerkmale vorgenommen werden.

Für die genealogische Forschung sind Zivilstandsregister, Kirchenbücher und vergleichbare Quellen wichtig. Diese stellen auch für Archive häufig nachgefragte Bestände dar und lassen digitale Editionen aus bestandserhalterischen Gründen und zur besseren Durchsuchbarkeit als wünschenswert erscheinen.

#### 3. *Mit welchen Editionswerkzeugen / Editoren arbeiten Sie?*

Bei den genutzten Editionswerkzeugen und Editoren wurde sowohl Software zur Codierung der Texte als auch zur Erstellung der Web-Präsentation und Verwaltung der Inhalte genannt. Werkzeuge, die alle drei Bereiche in einer Applikation abdecken, also eine Benutzerschnittstelle zur Transkription mit Funktionen zur Erstellung der Präsentationsebene und Verwaltung der Inhalte kombinieren, stellen hier die Minderheit dar.

Das mit Abstand am meisten eingesetzte Werkzeug ist der oXygen XML-Editor. In der deutschen Sprachversion wurde oXygen 11 mal genannt und in der englischen sogar 20 mal. Der Editor bietet Unterstützung für XSLT, XQuery und TEI. Er lässt sich darüber hinaus mit XML-Datenbanken verbinden und mit Hilfe von CSS können visuelle Eingabemasken erstellt werden (Author-Modus)<sup>32</sup>. Damit wird die Bearbeitung der Inhalte auch für Nutzer ohne XML-Kenntnisse ermöglicht. Für Transkriptionsprojekte kleinerer Einrichtungen mit dem Wunsch nach einer Software zur nutzergenerierten Transkription im Sinne des Crowdsourcing, wie sie in dieser Arbeit im Vordergrund stehen, ist der Editor dennoch nur bedingt einsetzbar. Er bietet die ganze Bandbreite an Konfigurationsmöglichkeiten, die für

---

<sup>31</sup>Vgl. LWL-Institut für westfälische Regionalgeschichte und LWL-Archivamt für Westfalen. *Digitale Westfälische Urkunden-Datenbank*. URL: [http://www.lwl.org/westfaelische-geschichte/portal/Internet/urkunden\\_datenbank/haupt.php?urlNeu=Ja](http://www.lwl.org/westfaelische-geschichte/portal/Internet/urkunden_datenbank/haupt.php?urlNeu=Ja) (besucht am 13.06.2012).

<sup>32</sup>Vgl. Syncro Soft. *oXygen XML-Editor*. URL: <http://www.oxygenxml.com/> (besucht am 16.06.2012).



maßgeschneiderte digitale Editionen notwendig sind, aber für die hier anvisierten Projekte ist die damit einhergehende Komplexität bei der Erstellung nicht gerechtfertigt. Es existieren für diesen Bereich spezialisierte Systeme, die eine begrenzte Funktionalität auch ohne die Komplexität einer vollwertigen XML-Entwicklung bereitstellen. Die begrenzteren Möglichkeiten zur projektspezifischen Anpassung dieser Systeme stellt die Kehrseite dieses Ansatzes dar. Was dennoch auf diesem Weg möglich ist, welche Anpassungen realistisch sind und welche Entwicklungen für die Zukunft angestrebt werden sollten, wird in dieser Arbeit in einer Analyse ausgewählter Systeme demonstriert.

Für Projekte, die sich für die Arbeit mit einem Editor wie oXygen entscheiden, ist dennoch eine Einbeziehung wenig technisch geschulter Bearbeiter bei der Transkription z. B. mit Hilfe des o. g. Author-Modus möglich. Eine weitere Möglichkeit ist die Nutzung der TEI-Unterstützung von OpenOffice<sup>33</sup>. Hier ist die Dateneingabe und der Export nach TEI-XML möglich, allerdings ist die Aufbereitung der Daten für die Online-Präsentation damit noch nicht abgedeckt. Diese Herangehensweise erfordert einen Workflow, bei dem die transkribierten Texte im XML-Format in einem zweiten Schritt für die digitale Edition mit Hilfe von XSLT in das Zielformat HTML transformiert werden müssen.

Zu den verwendeten CMS zählen Drupal<sup>34</sup>, Omeka<sup>35</sup> und XTF<sup>36</sup>. Einige Transkriptionsmodule besitzen eine CMS-Integration wie z. B. Scripto<sup>37</sup> (z. B. für Omeka, Wordpress<sup>38</sup>, Drupal) oder TEICHI<sup>39</sup> (für Drupal). In dieser Arbeit wird das Transkriptionsmodul Scripto in Verbindung mit Omeka genauer untersucht.

Im Rahmen von Online-Werkzeugen, wie sie in dieser Arbeit im Mittelpunkt stehen, wurden bei der Umfrage T-PEN<sup>40</sup>, das o. g. Transkriptionswerkzeug Scripto,

---

<sup>33</sup>Vgl. Apache Software Foundation. *Apache OpenOffice*. URL: <http://www.openoffice.org/de/> (besucht am 16.06.2012).

<sup>34</sup>Vgl. Drupal Community. *Drupal*. URL: <http://www.drupal.de/> (besucht am 16.06.2012).

<sup>35</sup>Vgl. Roy Rosenzweig Center for History and New Media, *Omeka*.

<sup>36</sup>Vgl. California Digital Library. *XTF*. URL: <http://xtf.cdlib.org/> (besucht am 16.06.2012).

<sup>37</sup>Vgl. Center for History and New Media, *Scripto*.

<sup>38</sup>Vgl. inpsyde. *Wordpress*. URL: <http://wpde.org/> (besucht am 16.06.2012).

<sup>39</sup>Vgl. Sebastian Pape und Stefan Achler. *TEICHI*. URL: <http://www.teichi.org/> (besucht am 16.06.2012).

<sup>40</sup>Vgl. Saint Louis University, Center for Digital Theology, *T-PEN*.

Refine!<sup>41</sup>, TextGrid<sup>42</sup>, MOM-CA<sup>43</sup>, FromThePage<sup>44</sup> sowie eLaborate<sup>45</sup> genannt. Diese Werkzeuge bieten den Vorteil einer web-basierten Transkription und erlauben mit der Ausnahme von T-PEN (hier können lediglich die XML-Dateien in ein CMS exportiert werden) auch eine Aufbereitung der Texte für die Online-Präsentation.

4. *Unterstützt Ihr Editionswerkzeug die Erstellung nutzergenerierter Transkriptionen?*

Bei der Frage, ob das betreffende Editionswerkzeug die Erstellung nutzergenerierter Transkriptionen unterstützt, antworteten die Befragten in etwa zu gleichen Teilen mit Ja und Nein, d. h. 51 % der aktuellen Projekte in der Umfrage beinhalten diese Funktion und 49 % verzichten darauf. Im Rahmen der geplanten Transkriptionsprojekte wollen 16 von 18 Befragten eine solche Funktion realisieren.

5. *Ist die Transkription auf einen beschränkten Nutzerkreis (z. B. von Wissenschaftlern) oder eine breitere Community (Crowdsourcing) ausgelegt?*

Mit dieser Frage sollte der anvisierte Nutzerkreis für die Transkriptionsprojekte ermittelt werden. Insgesamt sind 65 % der geplanten und aktuellen Projekte in beiden Sprachversionen auf einen beschränkten Nutzerkreis ausgelegt. Noch deutlicher fällt die Dominanz von Projekten mit beschränktem Nutzerkreis bei isolierter Betrachtung der Projekte in der englischen Sprachversion auf: Hier geht es bei 15 Projekten um einen beschränkten Nutzerkreis, während lediglich 3 Projekte mit Mitteln des Crowdsourcing arbeiten. Bei den Befragten in der deutschen Sprachversion ergibt sich ein ausgewogenes Bild: 8 der aktuellen Projekte arbeiten mit einem beschränkten Nutzerkreis und 9 zielen auf eine breitere Community ab.

6. *Halten Sie eine Transkriptionsfunktion ohne Konfrontation des Bearbeiters mit dem Quelltext (z. B. TEI, HTML) für sinnvoll?*

---

<sup>41</sup>Vgl. Staatsbibliothek zu Berlin, Humboldt-Universität zu Berlin und 3pc GmbH Neue Kommunikation, *Refine!*

<sup>42</sup>Vgl. TextGrid Forschungsverbund. *Textgrid*. URL: <http://www.textgrid.de/> (besucht am 17.06.2012).

<sup>43</sup>Vgl. Monasterium.net und Historisch-Kulturwissenschaftliche Informationsverarbeitung der Universität Köln. *MOM-CA*. URL: <http://www.mom-wiki.uni-koeln.de/> (besucht am 17.06.2012).

<sup>44</sup>Vgl. Brumfield, *FromThePage*.

<sup>45</sup>Vgl. Huygens Institute for the History of the Netherlands. *eLaborate*. URL: <https://www.elaborate.huygens.knaw.nl/> (besucht am 17.06.2012).

Viele der Befragten wiesen darauf hin, dass die Konfrontation des Benutzers mit einer Auszeichnungssprache bei der Erstellung von Editionen von dem Bearbeiterkreis abhängig zu machen ist. Wenn die Bearbeiter mit der Codierung in TEI vertraut sind, spricht natürlich nichts gegen eine solche Funktion und die Auszeichnung der entsprechenden Textmerkmale kann so in einem Arbeitsschritt mit der Transkription auf sehr effiziente Weise gestaltet werden. Bei Crowdsourcing-Projekten trifft das aber in den meisten Fällen nicht zu und hier muss gut abgeschätzt werden, in welchem Maß der Benutzer an eine solche Codierungsfunktion herangeführt werden sollte.

Eine Möglichkeit besteht darin, den Nutzer allein mit der Transkription zu betrauen und die Textauszeichnung in einem zweiten Schritt von einem Mitarbeiter ausführen zu lassen. Eine weitere Option ist die wahlweise Sichtbarmachung der XML-Tags. So kann der Bearbeiter selber auswählen, ob er auf der Markup-Ebene arbeiten möchte. Die Arbeit mit einer Werkzeugleiste zur Vergabe bestimmter Auszeichnungselemente oder die Verwendung einer vereinfachten Version der Auszeichnungssprache (siehe Tabelle 10.5 Seite 122) stellen eine weitere Möglichkeit dar.

Das Projekt „Transcribe Bentham“ arbeitet mit einer Werkzeugleiste zur Vergabe von TEI-Tags. Weil die Transkription in eine Klartextbox eingegeben wird, kann der Code nicht durch farbliche Kennzeichnung oder andere Formatierungen vom Transkriptionstext unterschieden werden. Als wünschenswerten Entwicklungsschritt für die Zukunft nennen Tim Causer et al. die Einführung eines (optionalen) WYSIWYG-Editors zur Ausblendung des Code und Darstellung des gerenderten Textes<sup>46</sup>. Die Antworten aus einer Freiwilligenbefragung zum Bentham-Projekt weisen ebenfalls in diese Richtung: „In total, 55 % of respondents to the volunteer survey agreed that a WYSIWYG interface would be beneficial to transcribing, particularly for new participants. Although a handful of respondents expressed satisfaction with the TEI toolbar, one volunteer – perhaps speaking for many others who abandoned the project without transcribing – noted that he had ,given up because of the encoding“<sup>47</sup>.

---

<sup>46</sup>Vgl. Tim Causer, Justin Tonra und Valerie Wallace. „Transcription maximized; expense minimized? crowdsourcing and editing The Collected Works of Jeremy Bentham“. In: *Literary and Linguistic Computing* 27.2 (2012), S. 119–137. DOI: 10.1093/llc/fqs004. eprint: <http://llc.oxfordjournals.org/content/early/2012/03/28/llc.fqs004.full.pdf+html>. URL: <http://llc.oxfordjournals.org/content/early/2012/03/28/llc.fqs004.abstract> (besucht am 21.06.2012), S. 131.

<sup>47</sup>Ebd., S. 131.

7. *Werden in Ihrem Transkriptionsprojekt Anreize zur Mitarbeit gegeben (z. B. Veröffentlichung der aktivsten Bearbeiter)?*

Bei den aktuellen Projekten gaben 71 % der Befragten an, dass keine Anreize zur Mitarbeit an den Transkriptionen gegeben werden. Dabei sollte allerdings beachtet werden, dass 66 % der Projekte in dieser Kategorie auf einen beschränkten Bearbeiterkreis abzielen (s. o.) und bei Editionen, die durch einen kleinen Kreis bezahlter Wissenschaftler erstellt werden, der Anreiz zur Mitarbeit ohnehin durch das Voranbringen der eigenen Forschungsarbeit und die berufliche Beschäftigung mit der Edition gegeben sind. In Crowdsourcing-Projekten dagegen, die durch die Beteiligung Freiwilliger vorangetrieben werden, spielt die Schaffung von Anreizen zur Mitarbeit eine wesentliche Rolle.

8. *Um welche Anreize zur Mitarbeit handelt es sich?*

An dieser Stelle wurde die Nennung der Bearbeiter auf der Projekt-Website und auch in der gedruckten Edition angegeben. Zusätzliche Details zur Anzahl der verfassten Transkriptionen ergänzen die Aussagekraft der Nennungen und bieten einen weiteren Anreiz, um einen möglichst hohen Platz in der „Rangliste“ der Bearbeiter einzunehmen. Auch ein Fortschrittsbalken kann die Motivation zur Mitarbeit positiv beeinflussen.

Ein weiteres wichtiges Mittel besteht im Feedback der Moderatoren. Dieses kann sowohl zur Verbesserung der Qualität der Transkriptionen beitragen als auch die Motivation der Bearbeiter steigern, weil ihr Einsatz damit gewürdigt wird und sie mehr in die Arbeit des Projektes einbezogen werden. Zum Feedback könnten auch E-Mails und virtuelle Urkunden zum Dank für die Erstellung einer bestimmten Anzahl von Transkriptionen gehören.

Darüber hinaus ist das inhaltliche Interesse an den handschriftlichen Quellen häufig ein Anreiz zur Beteiligung. Auch das Interesse an der technischen Umsetzung des Projektes und der freiwillige Beitrag zu einem Projekt von öffentlichem Interesse spielen häufig eine Rolle<sup>48</sup>. Ebenso kann die Einbeziehung von Studentengruppen

---

<sup>48</sup>Vgl. Causer, Tonra und Wallace, „Transcription maximized; expense minimized? crowdsourcing and editing The Collected Works of Jeremy Bentham“, S. 127.

in die Transkription einen Mehrwert für beide Seiten generieren. Dieser Ansatz wurde bei der digitalen Edition des Nachlasses Franz Brümmer<sup>49</sup> verfolgt.

9. *Wird in Ihrem Editionsprojekt eine Qualitätssicherung (Redaktion der Transkriptionen) vorgenommen?*

Die Beantwortung dieser Frage fällt recht eindeutig aus: In 84 % der aktuellen Transkriptionsprojekte wird eine Qualitätssicherung vorgenommen. Bei den geplanten Projekten wollen alle Befragten eine Qualitätssicherung vornehmen.

10. *Welche technischen Hilfsmittel werden zur Qualitätssicherung eingesetzt?*

In den meisten Projekten wird die Qualitätssicherung durch Korrekturlesen vorgenommen. Dieser Arbeitsschritt kann von den Moderatoren durchgeführt werden, aber bei einem entsprechend großen und aktiven Bearbeiterkreis könnten sich diese auch gegenseitig kontrollieren (so wird es im Projekt Wikisource<sup>50</sup> bereits gehandhabt). Darüber hinaus wurde von einem der Befragten eine Kommentarfunktion zur Diskussion schwer lesbarer Textpassagen durch die Bearbeiter beschrieben (siehe Tabelle 10.8 Seite 124). Die meisten Transkriptionsprogramme bieten auch eine Versionskontrolle für die Transkripte an, mit deren Hilfe auch auf ältere Versionen zurückgegriffen werden kann. Ein weiteres Mittel zur Qualitätssicherung besteht in der Validierung von XML-Dateien mit Hilfe einer Schemasprache.

Das Deutsche Textarchiv arbeitet mit einem webbasierten Werkzeug zum Finden und Korrigieren von Fehlern. Es ermöglicht verschiedene Sichten auf Digitalisate, Transkriptionen und Präsentationsformen. Die Qualitätssicherung wird über ein Ticket-System organisiert, bei dem Tickets klassifiziert und einzelnen Benutzern zugewiesen werden können<sup>51</sup>.

11. *Besitzt Ihr Editionswerkzeug eine Funktion für die textkritische oder sachkritische Kommentierung der Texte?*

---

<sup>49</sup>Vgl. Staatsbibliothek zu Berlin und Humboldt-Universität zu Berlin. *Nachlass Franz Brümmer*. URL: <http://bruemmer.staatsbibliothek-berlin.de/nlbruemmer/> (besucht am 21.06.2012).

<sup>50</sup>Vgl. Wikimedia Foundation. *Wikisource*. URL: [http://en.wikisource.org/wiki/Main\\_Page](http://en.wikisource.org/wiki/Main_Page) (besucht am 21.06.2012).

<sup>51</sup>Vgl. Deutsches Textarchiv. *DTAQ: Kollaborative Qualitätssicherung im Deutschen Textarchiv*. URL: <http://kaskade.dwds.de/dtaq/about> (besucht am 23.06.2012).

Bei den aktuellen Projekten besitzen 56 % der Werkzeuge eine Funktion zur Erstellung von Kommentaren. Bei den geplanten Projekten soll in 94 % der Fälle eine solche Funktion unterstützt werden. Dabei ist zu beachten, dass Befragte, die mit TEI und einem Editor wie oXygen arbeiten, natürlich im Rahmen der Verwendung von TEI vielfältige Möglichkeiten zur Kommentierung besitzen. Wie auch in den Antworten auf diese Frage angemerkt wurde (siehe Tabelle 10.10 Seite 126), sind diese Eigenschaften von TEI allerdings kaum als Funktion zu bezeichnen. Die Frage zielte eher auf eine Funktion im Rahmen von Online-Werkzeugen ab, die es erlaubt, Kommentare (in TEI oder anderen Formaten) in den Text einzufügen. Dieses ist etwa mit T-PEN<sup>52</sup> möglich.

Die Kommentierung von Texten etwa mit Verweisen auf korrespondierende archivalische Quellen oder mit Worterläuterungen ist bei mehr als der Hälfte der Befragten ein Teil der Arbeit mit Editionen. Im Rahmen von Crowdsourcing-Projekten wird diese Aufgabe allerdings häufig in einem zweiten Arbeitsschritt durch die Projektmoderatoren vorgenommen werden. Die Schaffung von Konsistenz z. B. bei Verweisen auf korrespondierende Akten in Archiven oder die Qualität und formale Einheitlichkeit von Worterläuterungen ist bei komplexen Fragestellungen durch einen heterogenen Bearbeiterkreis häufig nicht zu leisten. Dennoch gibt es Ansätze zur kollaborativen Arbeit auch im Rahmen der Kommentierung, wie etwa das Projekt Wikisource<sup>53</sup> zeigt. Zudem ist eine Kommentierung besonders im Fall von Crowdsourcing-Projekten oftmals nicht Teil der Bemühungen, weil der Schwerpunkt dieser Projekte i. d. R. mehr auf der Transkription der Inhalte der Handschriften liegt.

12. *Halten Sie eine Speicherung der Editionstexte in einem XML-basierten Format wie TEI für unumgänglich?*

65 % der Befragten, die sich mit aktuellen Editionsprojekten beschäftigen, gaben an, dass sie Alternativen zur Speicherung von Editionstexten in TEI sehen. In den zahlreichen Textkommentaren zu dieser Frage wurde auf die Stärken von TEI verwiesen und gleichzeitig die Existenz von Alternativen hervorgehoben. Viele der Befragten sahen sich außer Stande, die Frage mit Ja oder Nein zu beantworten und vermerkten ihren Standpunkt im Kommentarfeld.

---

<sup>52</sup>Vgl. Saint Louis University, Center for Digital Theology, *T-PEN*.

<sup>53</sup>Vgl. Wikimedia Foundation, *Wikisource*.

Die für TEI genannten Vorteile sind Interoperabilität, eine große NutzerCommunity, weite Verbreitung sowie Reichhaltigkeit (siehe Tabelle 10.12 Seite 129). Verschiedentlich wurde auch auf den Nutzen von TEI als Master-Archivformat hingewiesen und betont, dass es sich in vielen Fällen bewährt, die Transkriptionen zuerst in einer Datenbank zu speichern und die XML-Dateien später zu exportieren oder die Daten z.B. in DOCX zu erfassen und später nach TEI zu konvertieren.

In Bezug auf die Alternativen zu TEI wurden Datenbanken, XHTML, ePub, JSON und Wikilinks (in Verbindung mit dem Transkriptionswerkzeug FromThePage<sup>54</sup>) genannt. Bei der Antwort „Datenbanken“ beziehen sich die Befragten vermutlich auf die Speicherung der Textdateien in Datenbanken zur Generierung einer dynamischen Darstellungsform. Die Frage zielte allerdings auf die Wahl des Formats zur Textauszeichnung ab und in dieser Hinsicht stellen Datenbanken lediglich die Systeme zur Verwaltung der Textdateien dar. HTML bietet als weit verbreiteter Standard eine bedenkenwerte Option zur Textauszeichnung, aber sie wird nicht in allen Fällen den Ansprüchen zur Codierung etwa von Personen- oder Ortsnamen ausreichen. Das Format ePub als Textformat für E-Book Reader spielt im Rahmen von digitalen Editionen, die zur Benutzung am Personalcomputer konzipiert werden, lediglich als optionales Exportformat zur Benutzung mit den genannten Endgeräten eine Rolle. Wikilinks als Komponente der Auszeichnungssprache Wikitext stellen vielfältige Möglichkeiten zur Verknüpfung von Ressourcen dar und werden neben FromThePage auch im Projekt Wikisource<sup>55</sup> eingesetzt.

Insgesamt bleibt festzuhalten, dass TEI eine gute Möglichkeit zur Speicherung von Editionstexten bietet, besonders wenn es auf eine detaillierte Erfassung von Textmerkmalen ankommt. Alternativen existieren aber und es hängt von den Anforderungen des einzelnen Editionsprojektes ab, mit welchem Standard gearbeitet werden sollte. Allerdings sollte es sich aus Gründen der Langlebigkeit und Interoperabilität nicht um proprietäre Formate handeln.

13. *Ist die Erstellung von Indizes (z. B. für Orts- oder Personennamen) Teil Ihres Editionsprojektes?*

Mit Indizes wird in 74 % der aktuellen Projekte gearbeitet und ihre Erstellung ist sogar bei 100 % der geplanten Projekte vorgesehen. Dabei sollte unterschieden

---

<sup>54</sup>Vgl. Brumfield, *FromThePage*.

<sup>55</sup>Vgl. Wikimedia Foundation, *Wikisource*.

werden, ob ein Online-Werkzeug die automatisierte Erstellung von Indizes mit der Generierung der Einträge auf der Präsentationsebene unterstützt (wie z.B. bei FromThePage und Refine!), oder ob die Indizes im Rahmen eines weiteren Arbeitsschrittes halbautomatisch mit Hilfe von XSLT-Stylesheets generiert werden müssen.

14. *Bietet Ihre Edition weitere Funktionen zur Textnavigation?*

Über die Erstellung von Indizes hinaus werden bei 74 % der aktuellen Projekte weitere Funktionen zur Textnavigation angeboten. In den geplanten Projekten wollen 83 % der Befragten solche Funktionen realisieren.

15. *Um welche Funktionen handelt es sich?*

An Navigationsmöglichkeiten wurden zeitliche Navigation, Inhaltsverzeichnisse, Navigation nach Auftrittshäufigkeit, Strukturelementen sowie die Volltextsuche genannt. Hinzugefügt werden kann, dass bei einer semantischen Auszeichnung von Textmerkmalen die Suche auch gezielt auf Inhalte bestimmter Elemente beschränkt werden kann, etwa für die Recherche in den Marginalien.

16. *Verweisen Sie in Ihrer digitalen Edition auf Normdaten oder Erschließungsinformationen?*

74 % der aktuellen Projekte verweisen von ihren digitalen Editionen auf Normdaten oder Erschließungsinformationen. Bei allen geplanten Projekten in der Umfrage ist dieses geplant.

17. *Um welche Daten handelt es sich (z. B. eigene Kataloge, OPAC, Kalliope, GND o.ä.)?*

Die Arbeit mit Normdaten ermöglicht den konsistenten Umgang mit Bezeichnungen für Personen-, Orts-, Körperschaftsnamen usw. Damit ist bei der Suche die Eingabe verschiedener Schreibweisen eines Namens möglich, die dann auf eine maßgebliche Ansetzungsform zurückgeführt wird. Zudem können unterschiedliche Schreibweisen in der Vorlage eindeutig zugeordnet werden, sodass eine entsprechende Suche jedes Auftauchen z. B. eines Namens Kürzels auf die gemeinte Person mit



ihrem vollständigen Namen bezieht. Neben der Vernetzung von Normdaten durch ihre Referenzierung in digitalen Editionen kann die Editionsarbeit auch zur Anreicherung der Datenbestände beitragen. So stellt die Datenlieferung etwa von bisher nicht in die Normdatei aufgenommenen Schriftstellernamen aus einem literarischen Nachlass sowohl für das Editionsprojekt als auch für die mit den Normdaten arbeitenden Institutionen einen Mehrwert dar.

In Bezug auf Normdaten wurden für den deutschsprachigen Raum GND, PND, GKD sowie VD17 genannt. An Verweisen auf internationale Normdaten bzw. Normdaten für den englischen Sprachraum sind VIAF, DNB, DCB, mythindex.com, mythologia.fr, LC Authorities, TGN, DBpedia, LGPN und ESTC aufgeführt worden. Darüber hinaus verweisen viele Projekte auf Kataloge der eigenen Institution.

18. *Liefern Sie Daten an größere Portale (z. B. Europeana, Deutsche Digitale Bibliothek o.ä.)?*

64 % der aktuellen Projekte beteiligen sich nicht an größeren Portalen. Bei den geplanten Projekten wollen 56 % Daten an größere Portale liefern.

19. *Um welche Portale handelt es sich?*

Die für den Datenimport von Web-Portalen geforderten Metadatenstandards sind heterogen. Grundsätzlich ist eine Exportfunktion in einem XML-Format für ein Editionswerkzeug vorteilhaft. Auf der Grundlage dieser Dateien können dann die entsprechenden Importformate erzeugt und die vorhandenen Elemente auf das geforderte Kerndatenset gemappt werden.

Genannt wurden Portale mit einem Schwerpunkt auf deutschsprachigen Inhalten wie Deutsche Digitale Bibliothek<sup>56</sup>, TextGrid Repository<sup>57</sup>, Archive NRW<sup>58</sup>, Kalliope<sup>59</sup> und Manuscripta Mediaevalia<sup>60</sup>. In Bezug auf internationale Portale wurden

---

<sup>56</sup>Vgl. Deutsche Digitale Bibliothek. *Deutsche Digitale Bibliothek*. URL: <http://www.deutsche-digitale-bibliothek.de/> (besucht am 25.06.2012).

<sup>57</sup>Vgl. TextGrid. *TextGrid Repository*. URL: <http://www.textgridrep.de/> (besucht am 25.06.2012).

<sup>58</sup>Vgl. Landesarchiv Nordrhein-Westfalen. *Archive in NRW*. URL: <http://www.archive.nrw.de/> (besucht am 25.06.2012).

<sup>59</sup>Vgl. Staatsbibliothek zu Berlin. *Kalliope*. URL: <http://kalliope.staatsbibliothek-berlin.de> (besucht am 25.06.2012).

<sup>60</sup>Vgl. Staatsbibliothek zu Berlin, Deutsches Dokumentationszentrum für Kunstgeschichte - Bildarchiv Foto Marburg und Bayerische Staatsbibliothek München. *Manuscripta Mediaevalia*. URL: <http://www.manuscripta-mediaevalia.de/> (besucht am 25.06.2012).

Europeana<sup>61</sup>, The European Library<sup>62</sup>, Wikimedia<sup>63</sup>, ProQuest<sup>64</sup>, JISC Collections<sup>65</sup>, ArchiveGrid<sup>66</sup>, Gallica<sup>67</sup> sowie Isidore<sup>68</sup> aufgezählt.

20. *Haben Sie weitere Kommentare zu Ihrem Editionsprojekt und Ihren Anforderungen an Editionswerkzeuge?*

Bei der abschließenden offenen Frage drückten zwei der Befragten ihre Wünsche nach einem Werkzeug mit einer überschaubaren Nutzeroberfläche bzw. einem Überblick zu Editionswerkzeugen aus (siehe Tabelle 10.22 Seite 137) aus. Ein weiterer Kommentar unterstrich, dass bisher nur wenig zur Entwicklung nutzerfreundlicher Werkzeuge für technisch wenig geschulte Anwender getan worden sei, während viel Arbeit in die Beschreibung textueller Daten (TEI) investiert wurde. In der vorliegenden Arbeit soll sowohl ein Überblick zu Werkzeugen gegeben werden, die einen vergleichsweise ausgereiften Entwicklungsstand erreicht haben und sich für Crowdsourcing-Projekte eignen, als auch wünschenswerte Entwicklungen sowie Defizite der existierenden Lösungsansätze aufgezeigt werden.

Ein weiterer Befragter wies auf seine Skepsis in Bezug auf web-basierte kollaborative Werkzeuge zur Erstellung digitaler Editionen hin. Dies gelte auch für projektspezifisch angepasste Software. Des Weiteren wurde in diesem Kommentar eine Empfehlung zur Schulung der Bearbeiter in XML und der Arbeit mit oXygen ausgesprochen. Im Rahmen von Crowdsourcing-Projekten bzw. kollaborativen Ansätzen, die technisch wenig geschulte Bearbeiter einbeziehen, ist diesen Forderungen aus Sicht des Autors dieser Arbeit nicht zuzustimmen. Die Erfahrungen etwa aus dem Projekt Transcribe Bentham zeigen, dass die Komplexität der Codierungsaufgaben

---

<sup>61</sup>Vgl. Europeana Foundation. *Europeana*. URL: <http://www.europeana.eu/portal/aboutus.html> (besucht am 25.06.2012).

<sup>62</sup>Vgl. The European Library. *The European Library*. URL: <http://www.theeuropeanlibrary.org/> (besucht am 25.06.2012).

<sup>63</sup>Vgl. Wikimedia Foundation. *Wikimedia*. URL: <http://www.wikimedia.org/> (besucht am 25.06.2012).

<sup>64</sup>Vgl. ProQuest. *ProQuest*. URL: <http://www.proquest.co.uk/en-UK/> (besucht am 25.06.2012).

<sup>65</sup>Vgl. JISC Collections. *JISC Collections*. URL: <http://www.jisc-collections.ac.uk/> (besucht am 25.06.2012).

<sup>66</sup>Vgl. Online Computer Library Center. *ArchiveGrid*. URL: <http://www.archivegrid.org/> (besucht am 25.06.2012).

<sup>67</sup>Vgl. Bibliothèque Nationale de France. *gallica – Bibliothèque Numérique*. URL: <http://gallica.bnf.fr/> (besucht am 25.06.2012).

<sup>68</sup>Vgl. Centre national de la recherche scientifique. *Isidore*. URL: <http://www.rechercheisidore.fr/> (besucht am 25.06.2012).

für die Arbeit in Crowdsourcing-Projekten wohl überlegt werden muss<sup>69</sup>. Darüber hinaus existieren ausgereifte web-basierte kollaborative Werkzeuge wie T-PEN, die auch den Einsatz von anspruchsvollem Markup für entsprechend geschulte Bearbeiter ermöglichen. Ein weiterer Passus aus diesem Kommentar erklärte, dass die Präsentationsebene projektspezifisch von entsprechend ausgebildeten Kräften erstellt werden sollte. In diesem Punkt sollte darauf hingewiesen werden, dass Ansätze wie Wikisource, FromThePage und Refine! schon jetzt zeigen, was mit Hilfe von automatisierten Verfahren zur Generierung der Web-Präsentation möglich ist. Die genannten Werkzeuge werden im folgenden ebenfalls besprochen.

---

<sup>69</sup>Vgl. Causer, Tonra und Wallace, „Transcription maximized; expense minimized? crowdsourcing and editing The Collected Works of Jeremy Bentham“, S. 131.

## 4 Auswahlkriterien

Bei der Auswahl eines geeigneten Programms für ein Transkriptionsprojekt sind vorab einige Fragen zum Quellenmaterial, dem Ziel des Projekts, dem organisatorischen Umfeld sowie zu technischen und finanziellen Ressourcen zu klären. Die im Anhang abgebildete Liste mit Transkriptionswerkzeugen gibt erste Anhaltspunkte zu den Eigenschaften der vorhandenen Software (siehe Tabelle 10.25 Seite 147). Die Auswahlkriterien für die in dieser Arbeit näher behandelte Software werden ebenso in diesem Kapitel behandelt (siehe Abschnitt 4.2 Seite 29).

### 4.1 Checkliste für potentielle Transkriptionsprojekte

Die hier vorgestellte Checkliste orientiert sich an den von Ben W. Brumfield in seinen Vortragsfolien für die Texas Conference on Digital Libraries formulierten Fragen zur Auswahl von Software für Transkriptionsprojekte mit Crowdsourcing-Funktionen<sup>70</sup>. In der vorliegenden Arbeit wird von einem Anwenderkreis ausgegangen, der als Ziel des Projekts eine digitale Edition definiert hat. Darüber hinaus wird angenommen, dass die zu transkribierenden Dokumente in einer prinzipiell freien Textform verfasst sind, d. h. es handelt sich nicht um Datenfelder, die zur zweckmäßigen Abbildung der Beziehungen zwischen Feldtitel und -Inhalt eine speziell angepasste Software erfordern. Letzteres ist z. B. bei der Transkription von Karteikarten mit Informationen zur Vogelmigration des North American Bird Phenology Program der Fall<sup>71</sup>. Die Abbildung einfacher Tabellen ist dagegen mit den meisten auf eine Transkription von Dokumenten in freier Textform ausgerichteten Programmen möglich.

---

<sup>70</sup>Vgl. Ben W. Brumfield. *Transcription Tools at TCDL 2012*. URL: <http://manuscripttranscription.blogspot.de/2012/05/transcription-tools-at-tcdl2012.html> (besucht am 06.07.2012).

<sup>71</sup>Vgl. United States Geological Survey. *North American Bird Phenology Program*. URL: <http://www.pwrc.usgs.gov/bpp/> (besucht am 16.06.2012).

## Quellenmaterial

- Für wen ist es von Interesse?

Die Inhalte des Quellenmaterials spielen für die zukünftigen Bearbeiter eine wichtige Rolle. Ist es lediglich für einen spezialisierten (kleinen) Kreis interessant oder könnten sich größere Gruppen dafür gewinnen lassen? Bestehen schon Verbindungen zu potentiellen Bearbeitern der Texte? Wäre eine gezielte Werbung für das Projekt sinnvoll? Auch Überlegungen zur Motivation der Freiwilligen z. B. durch die Veröffentlichung der aktivsten Bearbeiter und regelmäßiges Feedback durch die Moderatoren sollten einbezogen werden.

- Ist es urheberrechtlich geschützt?

Die Wahl des Lizenzmodells sollte im Vorhinein geklärt werden. Die Forderung nach der Vergabe einer Open-Access-Lizenz spielt z. B. auch bei der Förderung durch die DFG eine Rolle. Als Mitunterzeichner der „Berliner Erklärung über den offenen Zugang zu wissenschaftlichem Wissen“<sup>72</sup> fördert die DFG schwerpunktmäßig Projekte, die ihre Inhalte unter einer freien Lizenz zugänglich machen. Darüber hinaus sind Projekte wie Wikisource<sup>73</sup> auf urheberrechtsfreie Dokumente beschränkt.

- Wie komplex und wie wichtig ist das Layout?

Diese Frage hat Auswirkungen auf den zu verwendenden Standard zur Textauszeichnung und die Programmfunktionen. Es sollte geklärt werden, welche Textmerkmale in der digitalen Edition abgebildet werden müssen. Für Crowdsourcing-Projekte bietet sich grundsätzlich eine flache Auszeichnung mit Layout-Informationen an. Letzteres verringert die Hürde für den Bearbeiter, sich neben der Transkription selbst auch mit den Editionsrichtlinien zur Textauszeichnung zu beschäftigen. Eine weitere Möglichkeit besteht in einer Arbeitsteilung zwischen Transkription und Codierung, wobei die Transkription von den Freiwilligen und die Textauszeichnung von Projektmitarbeitern durchgeführt wird.

---

<sup>72</sup>Vgl. *Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities*. 22. Okt. 2003. URL: [http://www.zim.mpg.de/openaccess-berlin/berlin\\_declaration.pdf](http://www.zim.mpg.de/openaccess-berlin/berlin_declaration.pdf) (besucht am 08.07.2012).

<sup>73</sup>Vgl. Wikimedia Foundation, *Wikisource*.

## Zielstellung

- Wie sollen die transkribierten Texte genutzt werden?

Hierbei handelt es sich um die Präsentation der Texte in der fertigen Edition und die Funktionen zur Textnavigation. Auch Funktionen zum Download von Texten als PDF oder TEI-Dateien sind für viele Anwendungsfälle sinnvoll. Die Generierung einer für den Druck optimierten Datei kann das Funktionsangebot ergänzen. Projektspezifische Anpassungen auf der Präsentationsebene erfordern in den meisten Fällen Programmierkenntnisse.

- Soll der Text mit der Software auch analysiert werden?

Eine Funktion zur Analyse der Vorkommenshäufigkeit verschiedener Themen mit einem Suchwort wird z.B. von FromThePage angeboten (siehe Abschnitt 6.1.1 Seite 40). Der Einsatz projektspezifischer Analysefunktionen ist vom Bedarf der späteren Nutzer der fertigen digitalen Edition abhängig zu machen.

## Organisatorisches Umfeld

- Welche Rolle spielt der editorische Workflow?

Optionen zum editorischen Workflow bestehen in der Redaktion der Transkripte durch ein Team von Editoren, die Qualitätssicherung durch die Freiwilligen selbst oder eine Kombination aus beiden Möglichkeiten.

- Sollen Freiwillige in die Arbeit einbezogen werden?

Einige Projekte arbeiten z.B. mit Studentengruppen im Rahmen von Seminaren an der Transkription. Andere beziehen einen heterogenen Kreis Freiwilliger in das Projekt ein.

- Auf welchen Zeitraum soll sich das Projekt erstrecken?

Sind die finanziellen und personellen Ressourcen für den Projektzeitraum gesichert?

Das Ausfindigmachen von Freiwilligen und die Aufrechterhaltung ihrer Motivation ist für die meisten Crowdsourcing-Projekte eng mit einem großzügigen Zeitrahmen verbunden. Erfahrungen aus dem Projekt Transcribe Bentham<sup>74</sup> zeigen, dass ein Projektzeitraum von sechs Monaten für derartige Vorhaben nicht ausreichend ist. Causer et al. unterstreichen diese Anforderung folgendermaßen: „In an ideal world,

---

<sup>74</sup>Vgl. University College London, *Transcribe Bentham*.

large scale crowdsourcing projects such as *Transcribe Bentham* would have sufficient funding for at least twelve months– and ideally fifteen to eighteen months with full staff support“<sup>75</sup>. In dem genannten Projekt hatten die meisten regelmäßig aktiven Freiwilligen nach der Ankündigung der Beendigung des vollständigen Support nach einer sechsmonatigen Projektphase die Mitarbeit eingestellt. Eine Ausnahme könnten Projekte mit Quellenmaterial von größerem öffentlichen Interesse und bedeutendem medialen Echo bilden, weil in diesen Fällen die erfolgreiche Akquirierung von Freiwilligen bedeutend schneller umzusetzen ist<sup>76</sup>. In allen Fällen ist aber eine offene Informationspolitik in Bezug auf den Projektzeitraum nötig, um die freiwilligen Bearbeiter nicht mit einer plötzlichen Ankündigung des Projektendes vor den Kopf zu stoßen.

- Gibt es eine abschließende Version der digitalen Edition?  
Sollen Veränderungen an der Edition auf lange Zeit hin möglich sein oder gibt es einen definierten Projektabschluss mit einer finalen Version des Editionstextes?
- Ist die Benutzung von TEI unumgänglich?  
TEI bietet viele Vorteile und hat sich als Standard für wissenschaftliche digitale Editionsprojekte etabliert. Als möglicher Nachteil ist die hohe Komplexität bei der Auszeichnung der Texte zu nennen. Als Lösung dieser Schwierigkeit bietet sich die automatische Generierung der TEI-Dateien, die Benutzung einer reduzierten Anzahl von Auszeichnungselementen oder eine Codierung der Texte durch Projektmitarbeiter an.

Laurent Romary kleidet die Überlegungen zur Auswahl einer sinnvollen Textcodierung in zwei Fragen. Das Ziel des Projekts soll in Bezug auf die Textauszeichnung im Rahmen der Frage „what do you need encoding for?“<sup>77</sup> eingegrenzt werden. Letzteres hängt von der Zielgruppe der digitalen Edition ab, d.h. welche Textmerkmale sind für die (wissenschaftliche) Arbeit mit dem Endprodukt von Wichtigkeit. Weiterhin ist die Frage nach den für die Verteilung der Daten etwa über Web-Portale notwendigen Informationen zu stellen. Auch die für die gewünschte Textpräsentation notwendigen Auszeichnungselemente müssen im Vorhinein erwogen werden. Nach der Festlegung des Ziels des Editionsprojekts folgt die konkretere

---

<sup>75</sup>Causer, Tonra und Wallace, „Transcription maximized; expense minimized? crowdsourcing and editing The Collected Works of Jeremy Bentham“, S. 132.

<sup>76</sup>Vgl. ebd., S. 132.

<sup>77</sup>Laurent Romary. „Questions & Answers for TEI Newcomers“. In: *Jahrbuch für Computerphilologie* 10 (2008). URL: <http://computerphilologie.de/jg08/romary.pdf> (besucht am 29.04.2012), S. 2.

Frage nach den tatsächlich zu codierenden Textmerkmalen: „What do I really (really) need to encode?“<sup>78</sup>. Diese Frage bezieht sich auf die Tiefe der semantischen Auszeichnung für die Codierung der strukturellen Texteigenschaften auf der Makroebene (Absätze, Abbildungen etc.) und von Oberflächenmerkmalen (Personennamen, Ortsnamen etc.). Zudem muss die Textauszeichnung für dokumentarische Zwecke (Referenzierung des Originals, Angabe des Bearbeiters etc.) festgelegt werden<sup>79</sup>.

Namentlich in Bezug auf Crowdsourcing-Projekte ist jedoch die von Laurent Romary ausgesprochene Warnung vor (unverhältnismäßig) detaillierter Textcodierung zu unterstreichen: „One of the main risks here is to tend towards an encoding overkill, whereby a lot of effort is put into encoding detailed phenomena in a text, which no user other than the encoder himself will ever need to use“<sup>80</sup>. Dieses spielt bei Crowdsourcing-Projekten eine besondere Rolle, weil die Textauszeichnung auf Grundlage umfangreicher Codierungsrichtlinien eine Hürde für die Mitarbeit Freiwilliger darstellen kann.

### Technische und finanzielle Ressourcen

- Stehen Systemadministratoren zur Installation nicht gehosteter Software zur Verfügung?

Die Installation der Software auf eigenen Servern erfordert in vielen Fällen tiefergehende technische Kenntnisse. Die Dokumentation ist häufig recht kurzgefasst und geht nicht auf die Einzelheiten spezieller Systemkonfigurationen ein. Kann auf entsprechende Mitarbeiter zur Installation und Konfiguration des Systems zurückgegriffen werden?

- Existieren finanzielle Mittel zur Bezahlung eines Hosting?

Das Hosting der digitalen Edition zieht in den meisten Fällen Kosten nach sich. Diese müssen in der Projektplanung einkalkuliert werden.

- Sind Programmierkenntnisse zur Anpassung des Programms vorhanden?

Die projektspezifischen Anpassungen müssen durch eigene Mitarbeiter mit entsprechenden Kenntnissen oder die Vergabe eines Auftrags vorgenommen werden.

---

<sup>78</sup>Romary, „Questions & Answers for TEI Newcomers“, S. 3.

<sup>79</sup>Vgl. ebd., S. 3.

<sup>80</sup>Ebd., S. 2.



- Sind finanzielle Mittel für laufende Kosten (Aufrechterhaltung des Online-Angebots) vorhanden?  
Die Langzeitverfügbarkeit des Angebots hängt nicht allein von den technischen Rahmenbedingungen ab. Die Finanzierung muss ebenso eingeplant werden.

## 4.2 Auswahl der Software für die weitere Analyse

Für die hier näher zu analysierenden Programme werden die folgenden Auswahlkriterien festgelegt:

### Entwicklungsstand

Für den anvisierten Benutzerkreis kommen nur Systeme in Frage, deren Entwicklungsstand und Dokumentation so ausgereift sind, dass ihre Konfiguration, Anpassung und Benutzung mit knappen Zeit- und Personalressourcen realisierbar sind. Eine Anpassung der Systeme durch eigene Mitarbeiter oder externe Aufträge wird dennoch in den meisten Fällen nicht zu vermeiden sein. Indes wird bereits vorhandenen und unter freier Lizenz stehenden Programmen der Vorrang gegenüber kostenpflichtigen oder mit Programmierarbeiten verbundenen Lösungen gegeben.

### Online-Werkzeuge

Die Programme sollten eine web-basierte Benutzeroberfläche anbieten, die eine vom Aufenthaltsort des Bearbeiters unabhängige Mitarbeit an der Transkription ermöglichen. Der Transkriptionsmodus und die digitale Edition als Endprodukt der Bemühungen sollten mit gängigen Browsern abrufbar sein. Die Möglichkeit für ein Hosting durch den Software-Anbieter ist besonders für kleinere Einrichtungen ohne eigene Server-Infrastruktur vorteilhaft. Sie bildet allerdings kein Ausschlusskriterium für die hier näher zu behandelnden Programme.

### Crowdsourcing-Funktionen

Die Einbeziehung größerer und heterogener Gruppen von Freiwilligen sollte Teil der angebotenen Funktionalität sein. Die Arbeit mit stärker definierten Gruppen wie etwa die Beteiligung einer studentischen Seminargruppe an dem Transkriptionsprojekt fällt ebenfalls unter den hier verwendeten Begriff des Crowdsourcing, weil in diesem Fall

ähnliche Voraussetzungen wie eine kurze Einarbeitungszeit und unterschiedliche technische Vorkenntnisse vorliegen. Die Ausweitung des Bearbeiterkreises etwa von kleineren Studentengruppen auf die breite Öffentlichkeit ist in den meisten Fällen ohne größere Anpassungen möglich. In der Tat eignet sich die Arbeit mit kleineren Gruppen häufig als „test case“ zur Überprüfung der Abläufe und dem reibungslosen Zusammenspiel der Programmfunktionen.

### Programmauswahl

Die folgenden Programme werden auf Grund ihrer Übereinstimmung mit den Auswahlkriterien in dieser Arbeit näher analysiert (siehe Abschnitt 6.1 Seite 40) und hier kurz mit ihren individuellen Schwerpunkten vorgestellt:

FromThePage<sup>81</sup> wurde Ben W. Brumfield entwickelt. Neben der Transkriptionsfunktion können die Texte semantisch ausgezeichnet und mit Hilfe hierarchischer Indizes durchsucht werden. Ein Hosting wird ebenfalls angeboten.

Der Refine!Editor<sup>82</sup> bietet eine einfache Benutzerführung und unterstützt die Indexerstellung sowie Transkription durch Bearbeiter ohne tiefgehende technische Vorkenntnisse. Das Programm wurde in Zusammenarbeit der Deutschen Staatsbibliothek, der Humboldt-Universität und der Firma 3pc<sup>83</sup> entwickelt. Es kann von dem genannten Unternehmen für Editionsprojekte angepasst werden.

Wikisource<sup>84</sup> bietet die Möglichkeit zur Publikation von Quellen, die entweder urheberrechtsfrei sind oder unter einer freien Lizenz stehen. Es beinhaltet Funktionen zum Korrekturlesen der Transkriptionen sowie ein Hosting der Texte. Bei der Arbeit an neuen Transkriptionsprojekten kann auf der Website nach Freiwilligen aus der bereits vorhandenen Wikisource-Community gesucht werden.

Das Plugin Scripto<sup>85</sup> kann in ein CMS integriert werden. Dort stellt es eine Transkriptionsfunktion für digitale Objekte mit der Möglichkeit zur Textauszeichnung in HTML zur Verfügung. In dieser Arbeit wird es in Zusammenhang mit Omeka, einem System zur Veröffentlichung digitaler Sammlungen, untersucht. Omeka bietet ein Hosting der Editionstexte an.

---

<sup>81</sup>Vgl. Brumfield, *FromThePage*.

<sup>82</sup>Vgl. Staatsbibliothek zu Berlin, Humboldt-Universität zu Berlin und 3pc GmbH Neue Kommunikation, *Refine!*

<sup>83</sup>Vgl. 3pc GmbH Neue Kommunikation. *3pc GmbH Neue Kommunikation*. URL: <http://3pc.de/> (besucht am 30.06.2012).

<sup>84</sup>Vgl. Wikimedia Foundation, *Wikisource*.

<sup>85</sup>Vgl. Center for History and New Media, *Scripto*.

T-PEN<sup>86</sup> ermöglicht die zeilenbasierte Transkription von Manuskripten. Bearbeiter mit unterschiedlichen technischen Kenntnissen können sich auf die Transkription selbst beschränken oder sich für eine Codierung des Textes in TEI entscheiden. Funktionen zum Projektmanagement ergänzen das Angebot. Die fertigen Transkripte können als TEI-XML exportiert werden. Eine Möglichkeit zur Erstellung des Front-End, d.h. die Präsentation der Texte im Internet, ist nicht Teil des Projekts.

Das Programm Bentham Transcription Desk<sup>87</sup> verfügt im Transkriptionsmodus über eine Werkzeugleiste zur Vergabe einer Auswahl von TEI-Auszeichnungselementen. Es ist in seiner Funktionalität wie T-PEN auf das Back-End beschränkt.

---

<sup>86</sup>Vgl. Saint Louis University, Center for Digital Theology, *T-PEN*.

<sup>87</sup>Vgl. University of London Computer Centre, UCL Bentham Project, *Bentham Transcription Desk*.

## 5 Prüfkriterien

Jedes Transkriptionsprojekt besitzt individuelle Anforderungen an das Editionswerkzeug. Einige Fragen zur Eingrenzung der geeigneten Werkzeuge wurden in einer Checkliste formuliert (siehe Abschnitt 4.1 Seite 24). Die in diesem Kapitel aufgeführten Prüfkriterien bilden den Hintergrund für die Analyse der näher zu behandelnden Programme. Mit ihrer Hilfe soll eine Einschätzung des Profils und der Schwerpunktsetzung der Werkzeuge möglich werden. Dabei zeigen die Kriterien den Rahmen des für bestimmte Anwendungsfälle möglichen und wünschenswerten auf. Sie bilden keine Ausschlusskriterien, sondern zeigen sinnvolle Kombinationen von Programmeigenschaften auf und verzeichnen eine Gewichtung dieser Merkmale. Die Evaluation bezieht sich dann auf die Erfüllung der genannten Kriterien. Die Definition der Prüfkriterien wird auf Grundlage der Umfrage und Literatur zum Thema vorgenommen.

C. M. Sperberg-McQueen hat bereits 1994 allgemeine Anforderungen an digitale Editionen auf folgende Weise zusammengefasst: “I believe electronic scholarly editions must meet three fundamental requirements: accessibility without needless technical barriers to use; longevity; and intellectual integrity<sup>88</sup>”. Obgleich es sich bei dem Artikel um eine ältere Publikation handelt, sind die grundsätzlichen Anforderungen ähnlich geblieben. Die technische Entwicklung hat den konzeptionellen Anspruch an digitale Editionen nicht obsolet werden lassen. Die von Sperberg-McQueen formulierten Anforderungen beziehen sich allerdings auf die digitale Edition als Endprodukt eines Editionsprojektes. Die in dieser Arbeit behandelten Werkzeuge sollten die Funktionen zur Erstellung einer Edition mitbringen, die diesen Anforderungen genügt. Hinzu kommen noch zwei Kriterien, die die Programme als solche betreffen. Die Anforderung an die Benutzeroberflächen von Editionswerkzeugen werden hier unter dem Titel Gebrauchstauglichkeit (Usability) behandelt. Ebenso spielt die Auffindbarkeit (findability) von Informationen im Internet angesichts einer Vervielfachung der vorhandenen Angebote eine zunehmend wichtige Rolle und wird als Kriterium ergänzt.

---

<sup>88</sup>Vgl. Sperberg-McQueen, „Textual Criticism and the Text Encoding Initiative“.

Die behandelten Werkzeuge bieten jeweils einen unterschiedlichen Funktionsumfang und decken damit nicht alle genannten Kriterien ab. Dem kann teilweise durch die Einbeziehung ergänzender Programme abgeholfen werden. Ein offensichtliches Beispiel ist das Plugin Scripto, denn es erfordert die Integration in ein bestehendes CMS und beschränkt sich auf Funktionen zum Crowdsourcing der Transkription. Es wird allerdings nicht für jedes Kriterium ein adäquates Programm zur Ergänzung der vorhandenen Funktionen geben. An dieser Stelle müssen Programmanpassungen oder Neuentwicklungen vorgenommen werden oder das Projekt muss sich entsprechend der individuellen Zielsetzung auf die Realisierung einer bestimmten Auswahl von Kriterien beschränken und andere vernachlässigen.

Grundsätzlich sollte beachtet werden, dass es bei den Anforderungen an eine digitale Edition nicht darum gehen kann, möglichst viele multimediale Elemente einzubinden. Fotis Jannidis warnt davor „die neuen Freiheiten zu einem neuen Zwang zu machen, d. h. eben deshalb, weil viele Medien im neuen Medium verbunden werden können, muß dieses Potential auch ausgenutzt werden“<sup>89</sup>. Vielmehr gehe es darum, die Eigenschaften der digitalen Edition vom Verwendungszweck abhängig zu machen<sup>90</sup>.

## 5.1 Zugänglichkeit

Der Aspekt der Zugänglichkeit (Accessibility) bezieht sich auf die Reproduzierbarkeit der digitalen Inhalte mit verschiedener Software und auf unterschiedlichen Endgeräten. Die hier betrachteten Editionen werden für die Benutzung am Personalcomputer konzipiert. Deshalb wird auf eine Darstellung auf Smartphones oder E-Books kein besonderer Wert gelegt. Letzteres kann für andere Editionen ein wichtiges Kriterium sein, z. B. wenn die Darstellung eines Lesetextes für das E-Book im Vordergrund steht.

### 5.1.1 Browser-Unterstützung

Derzeit sollten die Inhalte mit den am meisten verbreiteten aktuellen Browserversionen ohne maßgebliche Unterschiede in Darstellung und Funktionalität wiedergegeben werden können. Die in dieser Arbeit behandelten Werkzeuge sind ausnahmslos web-basiert (Auswahlkriterium „Online-Werkzeuge“) und erfüllen das Prüfkriterium.

---

<sup>89</sup>Vgl. Fotis Jannidis, Hrsg. *Bewertungskriterien für elektronische Editionen*. IASL Diskussionsforum online. 1999. URL: <http://iasl.uni-muenchen.de/discuss/lisforen/jannidis.htm> (besucht am 29.04.2012).

<sup>90</sup>Vgl. ebd.

## 5.2 Auffindbarkeit

Um die Auffindbarkeit (findability) einer Ressource sicherzustellen, sollten die Inhalte nicht lediglich über Google gefunden werden können, sondern auch über Fachportale wie Europeana oder Deutsche Digitale Bibliothek erreichbar sein. Weitere Portale können als Beispiele für die Kooperationspartner der entsprechenden Projekte den Textantworten entnommen werden (siehe Abschnitt 19 Seite 21). Die Beteiligung an solchen Gemeinschaftsprojekten erhöht auch die Sichtbarkeit des Projekts im Internet und bindet es in den Kontext vorhandener Datenbestände ein. Ebenso wirkt sich die Referenzierung von Normdaten auf die Auffindbarkeit der betroffenen Inhalte aus.

### 5.2.1 Schnittstellen

Nicht nur die Benutzung verbreiteter Standardformate ist bei der Datenlieferung an Fachportale von Bedeutung (siehe Abschnitt 5.3.2 Seite 35), sondern auch das Vorhandensein entsprechender Schnittstellen. Dabei sind grundsätzlich zwei Ansätze zu unterscheiden: die einmalige Datenlieferung durch das Hochladen der Daten auf einen FTP-Server oder der wiederholte Datenabgleich, das Metadata Harvesting, mit Hilfe einer OAI-PMH-Schnittstelle. Bei abgeschlossenen Editionsprojekten mit einer finalen Version des Textes bietet sich der erste Ansatz an. Die Aktualisierung der Datenbestände in Portalen ist eine Option für Projekte, die auf einen langen Zeitraum ausgelegt sind und ihre Daten bereits während des laufenden Transkriptionsprojekts veröffentlichen wollen. Grundsätzlich ist die Möglichkeit zum Export der Daten in einem XML-basierten Format sehr vorteilhaft. Dieses kann dann entsprechend der Vorgaben des Portals in ein Importformat, z. B. mit einem bestimmten Kerndatensatz, transformiert werden.

### 5.2.2 Verweise auf Normdaten

Verschiedene Schreibweisen z. B. eines Personennamens werden mit Hilfe von Normdateien auf eine maßgebende Ansetzungsform zurückgeführt und so wird eine gezielte Suche unter Einbeziehung von Abkürzungen und abweichenden Schreibweisen im Editionstext möglich. Durch die gemeinsame Nutzung verbreiteter Normdateien wie VIAF und GND wird auch ein Beitrag zur Vernetzung der verschiedenen Ressourcen geleistet. Die in der Umfrage genannten Normdaten und Erschließungsinformationen, auf die in den Projekten der Befragten verwiesen wird, bilden ein weites Spektrum ab (siehe Abschnitt 17 Seite 20). Die aktuellen Editionswerkzeuge sind auf Grund des Aufwands

zur Programmanpassung nicht mit der Möglichkeit zur Referenzierung einer so großen Menge verschiedener Normdaten ausgestattet. Die Schwerpunktsetzung auf verbreitete und internationale Normdateien wie VIAF und GND könnte hier für die zukünftige Entwicklung aber einen realistischen Ansatz bieten.

## 5.3 Langlebigkeit

Die Langlebigkeit (longevity) der Edition ist von verschiedenen Parametern abhängig: Textcodierung, Software und Aufrechterhaltung des elektronischen Angebots durch eine Institution oder einen bestimmten Personenkreis<sup>91</sup>.

### 5.3.1 Lizenz

Im Abschnitt “Software is not the Answer”<sup>92</sup> legt C. M. Sperberg-McQueen dar, dass die verwendete Software nicht das wichtigste Kriterium in Bezug auf die Langlebigkeit der digitalen Edition ist. Ein Programm, das von der technischen Entwicklung überholt und nicht weiterentwickelt wird, kann innerhalb eines kurzen Zeitraums für den Nutzer unbrauchbar werden und nur mit großem technischen Aufwand (Emulation der alten Softwareumgebung oder Aufbau eines „Technikmuseums“) in seiner alten Funktionalität wieder hergestellt werden. Um die Langlebigkeit des Angebots sicherzustellen, wäre die Software also nicht der ideale Ansatzpunkt. Dennoch können die Bedingungen für eine lange Lebenszeit der Software durch die Benutzung von Open-Source-Produkten verbessert werden.

### 5.3.2 Codierungsformat

Der wichtigere Ansatzpunkt besteht in der Codierung der Inhalte in einem Format wie TEI, das unabhängig von spezifischen Hard- oder Software-Umgebungen auch nach einem langen Zeitraum mit einer anderen Software wieder dargestellt werden kann. Allerdings werden in dieser Arbeit auch Ansätze evaluiert, die bei der Codierung nicht auf TEI setzen. Die Rahmenbedingungen vieler Crowdsourcing-Projekte machen diesen Schritt nachvollziehbar, denn hier kommt es weniger auf eine tiefe semantische Auszeichnung

---

<sup>91</sup>Die Unterhaltung der digitalen Edition selbst ist in den meisten Fällen von öffentlichen Geldgebern und den Prioritäten der betroffenen Institution abhängig. An dieser Stelle können keine technischen Lösungen helfen, sondern es muss die öffentliche Wahrnehmung für die Wichtigkeit einer Überlieferung des kulturellen Erbes im digitalen Zeitalter gestärkt werden.

<sup>92</sup>Vgl. Sperberg-McQueen, „Textual Criticism and the Text Encoding Initiative“.

der Texte an, sondern es geht in der Hauptsache um die Transkription eines umfangreichen Archivbestands mit einer geringen Dichte von semantischen Auszeichnungen. Deshalb werden ebenso Ansätze evaluiert, die auf eine TEI-Codierung verzichten und stattdessen lediglich eine Speicherung etwa in HTML bzw. eine Aufbereitung von Metadaten z. B. nach dem Dublin-Core-Standard vorsehen. Wenn ein Projekt mit den in HTML verfügbaren Anzeichnungselementen auskommt, spricht im Sinne der Langlebigkeit auch nichts gegen dieses Format, denn es ist ebenso interoperabel wie TEI und wird von einem Standardisierungsgremium (W3C<sup>93</sup>) weiterentwickelt. Dabei sollte allerdings bedacht werden, dass im Sinne der Langlebigkeit auch die Aufnahme von Textmerkmalen in die Codierung erwogen werden sollte, die über den Projektzeitraum hinaus für zukünftige Benutzungsarten von Bedeutung sein könnten<sup>94</sup>.

Auch die Umfrage ergab in Bezug auf Alternativen zu TEI, dass die Befragten zum Teil selbst mit anderen Formaten arbeiten (HTML, Wikitext, RTF, Klartextformate) oder zum überwiegenden Teil von 65 % zumindest eine Speicherung in TEI nicht für unumgänglich halten (siehe Abschnitt 12 Seite 18). Wichtig ist dabei die Interoperabilität der verwendeten Textformate, d. h. die Benutzung eines proprietären Formats mit einer Abhängigkeit von der Software eines bestimmten Anbieters wäre der Forderung nach Langlebigkeit abträglich.

### 5.3.3 Hosting

Ebenso fällt die Frage des Hosting im weiteren Sinne unter das Kriterium der Langlebigkeit. Es muss für den Zeitraum des Transkriptionsprojekts sowie für die Aufrechterhaltung des Online-Angebots der finalen Textversion sichergestellt werden. Grundsätzlich bieten sich zwei Möglichkeiten des Hosting an: der Einsatz eigener Servertechnik oder der Einkauf entsprechender Leistungen bei einem Dienstleister. Einige der hier behandelten Software-Projekte bieten auch selbst ein Hosting der Inhalte an.

## 5.4 Intellektuelle Redlichkeit

Unter der intellektuellen Redlichkeit (intellectual integrity) ist die Aufnahme von Annotationen, Textvarianten, unsicheren oder spekulativen Lesarten einer Textstelle u. ä. zu verstehen. Als Prüfkriterium für die behandelten Programme stellt sich die Frage,

---

<sup>93</sup>Vgl. World Wide Web Consortium (W3C). *World Wide Web Consortium*. URL: <http://www.w3.org/> (besucht am 10.07.2012).

<sup>94</sup>Vgl. Romary, „Questions & Answers for TEI Newcomers“, S. 2.



inwiefern die für das konkrete Projekt angesetzten textkritischen Instrumente mit dem gegebenen Funktionsumfang umgesetzt werden können. Sperberg-McQueen unterscheidet dabei die Anforderungen für Editionen, die einen textkritischen Anspruch haben (text-critically aware) und solche, die auf diesen Anspruch verzichten<sup>95</sup>. Für die vorliegende Arbeit kommen beide Varianten in Frage und Editionen mit einer weniger wissenschaftlich orientierten Zielgruppe spielen besonders im Rahmen des Crowdsourcing eine wichtige Rolle.

### 5.4.1 Textauszeichnung

Als Prüfkriterium für die Programme muss untersucht werden, was für eine Textauszeichnung (Format und Tiefe der Codierung) über die Programmfunktionen realisiert werden kann. Es ist offensichtlich, dass projektspezifisch nicht weiter angepasste Programme nicht alle erdenklichen Funktionen zur Textkritik integrieren können. Dies trifft in besonderem Maße zu, wenn der Transkriptionsmodus mit der Präsentationsebene (etwa zur automatisierten Erstellung von Indizes) gekoppelt ist. Die Anreicherung der Texte mit semantischem Markup kann auf unterschiedliche Weise erfolgen: Durch direkte Eingabe der entsprechenden Auszeichnungselemente im Text, durch die Verwendung einer vereinfachten Auszeichnungssprache oder mit einem WYSIWYG-Editor.

## 5.5 Gebrauchstauglichkeit

In Teil 11 der ISO-Norm 9241 wird die Gebrauchstauglichkeit (usability) folgendermaßen definiert: „Extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use“<sup>96</sup>. Im Rahmen von Editionswerkzeugen mit Transkriptionsfunktion spielen die genannten Anforderungen sowohl im Back-End des Systems, d.h. hier für den Transkriptionsmodus und die Benutzerverwaltung, als auch im Front-End, respektive der Präsentationsebene mit ihren Navigations- und Benutzungssfunktionen eine Rolle. Da Bearbeiter und Ziele der für diese Arbeit in Frage kommenden Transkriptionsprojekte eine gewisse Heterogenität aufweisen können, gehen die Prüfkriterien auf verschiedene Lösungsansätze ein.

---

<sup>95</sup>Vgl. Sperberg-McQueen, „Textual Criticism and the Text Encoding Initiative“.

<sup>96</sup>International Organization for Standardization (ISO). *ISO 9241-11: Ergonomic requirements for office work with visual display terminals (VDTs) - Guidance on usability*. URL: <http://www.it.uu.se/edu/course/homepage/acsd/vt09/ISO9241part11.pdf> (besucht am 11.07.2012), Abschnitt 3.1.

### **5.5.1 Transkriptionsmodus**

Dieser Modus stellt die zentrale Benutzerschnittstelle zur Eingabe und Bearbeitung der Transkriptionen dar. Eine Box zur Texteingabe und ein zoombarer Viewer zur Betrachtung des Scans sind die grundlegenden Elemente in dieser Bearbeitungsform. Die Bandbreite der Lösungsansätze zur Textcodierung umfassen WYSIWYG-Editoren, optionale Einblendung der Codierung, vereinfachte Auszeichnungssprachen und die direkte Codierung mit HTML, Wikitext oder TEI. Eine Funktion zur Eingabe und Bearbeitung von Metadaten kann ebenfalls in den Transkriptionsmodus integriert sein. Auch die Angabe des Status der Transkription (unbearbeitet, in Bearbeitung, abgeschlossen) ist zweckmäßig.

### **5.5.2 Textpräsentation und -Navigation**

Bei einer Verknüpfung von Transkriptionsmodus und Präsentationsebene ist eine Vorschaufunktion zur Überprüfung des Layout sinnvoll. Ebenso sollte die Darstellung von grundlegenden Formatierungselementen wie Überschriften, Absätze, Unterstreichungen etc. zum Funktionsumfang gehören.

Bei der Erstellung von Navigationselementen ist eine Funktion zur Vergabe von Indexkategorien (z. B. für Personennamen, Ortsnamen etc.) für viele Anwendungsfälle angebracht. Im Sinne der Einheitlichkeit kann es aber auch zweckmäßig sein, diese Kategorien im Vorhinein durch die Moderatoren festzulegen.

Eine Diskussionsfunktion über die Transkription unsicherer Textstellen kann die kollaborative Arbeitsweise sinnvoll ergänzen.

### **5.5.3 Bearbeitungsstatistiken**

Zu den Bearbeitungsstatistiken zählen Informationen über die Anzahl der von den Freiwilligen vorgenommenen Transkriptionen und Indexierungen. Auch eine „Rangliste“ der aktivsten Bearbeiter ist zur Motivationssteigerung nützlich. Der Status der Bearbeitung kann graphisch z. B. mit einem Fortschrittsbalken angezeigt werden.

### **5.5.4 Benutzerverwaltung**

Die Anmeldung der Bearbeiter kann zur Vermeidung von Missbrauch mit Benutzername und Passwort erfolgen. Da eine Registrierung für viele Benutzer eine Hürde darstellt, ist die Möglichkeit zur anonymen Mitarbeit allerdings auch überlegenswert. In diesem Fall

kann die Transkription ohne Anmeldung z. B. auf bestimmte Quellen begrenzt werden. Nachdem ein Benutzer auf diese Weise erste Erfahrungen mit dem Transkriptionsprojekt gesammelt hat, wird er möglicherweise später in größerem Umfang dazu beitragen und mit den Moderatoren oder anderen Nutzern in Kontakt treten wollen. In dieser Situation wird der Nutzer einer Registrierung offener gegenüberstehen.

Darüber hinaus muss die Zuordnung der Editionstexte zu verschiedenen Bearbeitern für das entsprechende Projekt geklärt werden. Wenn die Anmeldung eines Bearbeiters für einen bestimmten Text erfolgt, kann die Sicherung dieser Transkription gegen Veränderung durch andere Bearbeiter sinnvoll sein. In diesem Fall sollten die anderen Bearbeiter das Transkript lediglich lesen und kommentieren können. Allerdings bietet eine freiere Handhabung der Textzuordnung ebenso Vorteile. So können Bearbeiter ohne die Bestätigung durch einen Moderator mit der Transkription eines Textes anfangen, der für sie von besonderem Interesse ist. Mit Hilfe einer Versionskontrolle sollte es in allen Fällen möglich sein auf ältere Textversionen zurückzugreifen.

# 6 Analyse und Evaluation

## 6.1 Editionswerkzeuge

Die Untersuchung der einzelnen Programme gliedert sich in eine Kurzbeschreibung, eine Erläuterung des Projekthintergrunds, eine Darstellung der angebotenen Funktionalität, eine Wiedergabe der verwendeten Standards und Schnittstellen sowie die Evaluation des Programms in Bezug auf die Prüfkriterien und seine Eignung für bestimmte Anwendungsfälle. Im folgenden Kaptiel werden die Tools gegenübergestellt und die Erfüllung oder Nichterfüllung der Prüfkriterien mit Hilfe einer Übersichtstabelle in kurzgefasster Form wiedergegeben (siehe Kapitel 7 Seite 83). Zudem werden zielgruppenspezifische Eigenschaften der Programme verglichen und mit Hilfe von Säulendiagrammen visualisiert (siehe Abschnitt 7.2 Seite 88).

### 6.1.1 FromThePage

Die freie Software erlaubt es Freiwilligen handschriftliche Quellen online zu transkribieren. Zu den grundlegenden Funktionen gehören die Indexierung und Annotierung von Themen mit Hilfe einer mit Wikitext vergleichbaren Auszeichnungssprache. Bearbeiter können schwer lesbare Wörter zur Verbesserung der Transkription diskutieren. Der entstandene Text wird im Internet gehostet und kann dort gelesen sowie durchsucht werden<sup>97</sup>.

#### Projekthintergrund

Das Programm FromThePage<sup>98</sup> wurde von Ben W. Brumfield ursprünglich zur Transkription der Tagebücher seiner Urgroßmutter entwickelt. Mit der Zeit wurden weitere Tagebücher aus Familienbesitz aufgenommen. Zur Erfüllung des Ziels einer Transkription und Präsentation der Tagebücher steht im Rahmen der Software-Entwicklung ledig-

---

<sup>97</sup>Brumfield, *FromThePage*, Vgl.

<sup>98</sup>Vgl. ebd.

lich noch die Integration eines Publish-on-Demand-Service für Familienmitglieder ohne Internetanschluss aus<sup>99</sup>.

Darüber hinaus entstand bei der Entwicklung die Idee das Programm auch anderen Interessierten, namentlich im Rahmen des Einsatzes von FromThePage für institutionelle Transkriptionsprojekte, zugänglich zu machen und in ein Mehrbenutzersystem umzuwandeln. Im Jahr 2010 wurde FromThePage unter einer AGPL-Lizenz veröffentlicht und von verschiedenen Institutionen für Transkriptionsprojekte eingesetzt<sup>100</sup> (siehe Abschnitt 6.1.1 Seite 45).

Das übergreifende Ziel besteht in einer Nutzbarmachung der privaten Anstrengungen zur Digitalisierung von historischen Quellen zur Familiengeschichte für eine größere Community. Die privaten „Archive“ im Besitz von Familien und Genealogen sollen einem breiteren Kreis zugänglich gemacht und als historische Primärquellen ediert werden<sup>101</sup>.

## Funktionen

Ein Benutzerhandbuch<sup>102</sup> aus einem Transkriptionsprojekt des San Diego Natural History Museum gibt Aufschluss über den Einsatz der Programmfunktionen. Nach einer Registrierung auf der Website können sich Benutzer an Transkriptionsprojekten beteiligen und Scans aus dem Internet Archive<sup>103</sup> importieren. Darüber hinaus bietet Ben W. Brumfield das Hosting von Scans zur Initiierung eines eigenen Projekts an. Als freie Software kann FromThePage auch heruntergeladen und auf dem eigenen Server installiert werden.

Das Dashboard der Internetseite verzeichnet importierte Inhalte und die letzten Veränderungen an Transkriptionstexten. Auf diese Weise können die neuesten Aktivitäten nachvollzogen werden und Bearbeiter haben die Möglichkeit auf die zuletzt bearbeiteten Seiten zurückzukehren.

Zur gemeinsamen Darstellung zusammenhängender Quellen auf der Website können Sammlungen (Collections) angelegt werden. In den Einstellungen können Titel, Beschreibung, FromThePage-Besitzer, Transkriptionskonventionen und weitere Metadaten hinterlegt werden.

---

<sup>99</sup>Vgl. Ben W. Brumfield. *My Goals for FromThePage*. Blog-Eintrag. 15. Feb. 2011. URL: <http://manuscripttranscription.blogspot.de/2011/02/my-goals-for-fromthepage.html> (besucht am 26.06.2012).

<sup>100</sup>Vgl. ebd.

<sup>101</sup>Vgl. ebd.

<sup>102</sup>Vgl. San Diego Natural History Museum. *FromThePage: User Guide*. URL: <http://files.balboaparkonline.org/psully/FromThePageUserGuide.pdf> (besucht am 26.06.2012).

<sup>103</sup>Vgl. Internet Archive. *Internet Archive*. URL: <http://archive.org/> (besucht am 27.06.2012).

## FromThePage

## Julia Brumfield Diaries — 1922

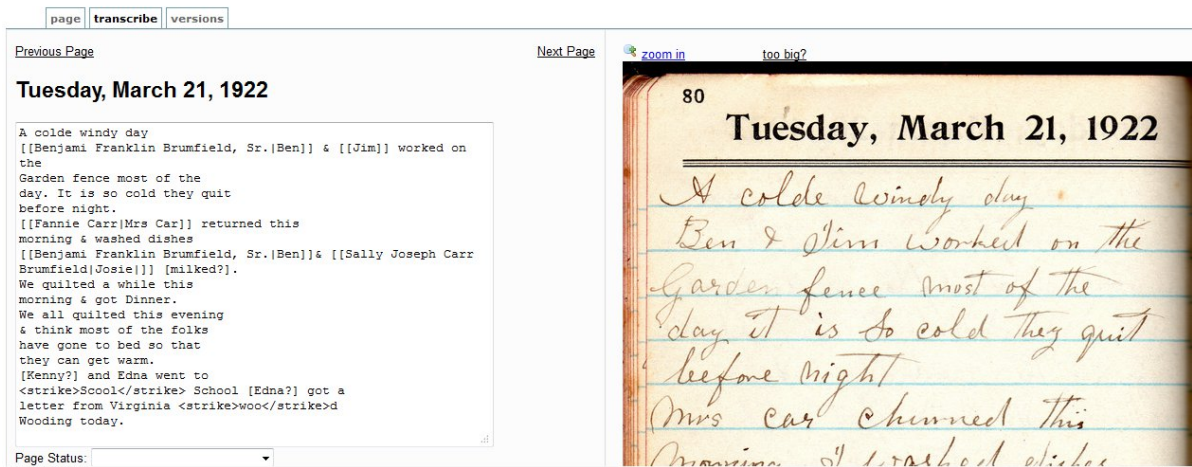


Abbildung 6.1: FromThePage: Transkriptionsmodus

Im Transkriptionsmodus (siehe Abbildung 6.1 Seite 42) wird die Bilddatei auf der rechten Seite angezeigt und die Transkription wird in eine Klartextbox auf der linken Seite eingegeben. Das Bild kann für eine bessere Lesbarkeit verkleinert und vergrößert werden. Eine Anmerkungsfunktion erlaubt die Eingabe von Kommentaren etwa zu schwer lesbaren Wörtern. Mit Hilfe des Versionskontrolle kann auf ältere Bearbeitungen des Transkriptionstextes zurückgegriffen werden. Zur Repräsentation des Layout im Transkriptionstext können grundlegende HTML-Elemente wie Kursivschrift, Unterstreichungen, Durchstreichungen, Fettdruck, Tabellen und Hyperlinks eingesetzt werden. Ein Zeilenumbruch wird durch die Enter-Taste eingefügt und Absätze mit Hilfe einer Leerzeile voneinander abgetrennt. Eine Vorschauansicht ermöglicht die Kontrolle der Formatierungen und des Textes in der gerenderten Form vor dem Abspeichern.

Im Präsentations- oder Lesemodus können mehrere Textseiten zusammen mit Thumbnails der Scans untereinander aufgelistet werden. Durch Scrollen gewinnt der Nutzer so einen schnellen Überblick z. B. über die inhaltlich zusammenhängenden Seiten eines Tagebuchs. In einer weiteren Darstellungsoption ist der Scan bei stärkerer Vergrößerung neben dem Transkriptionstext angeordnet. Hierbei wird je Bildschirmseite eine Seite des Tagebuchs gezeigt. Darüber hinaus kann der Leser alle Seiten aufrufen, in denen ein bestimmtes Thema vorkommt, wobei wie in der ersten Präsentationsform mehrere Tagebuchseiten auf einer Bildschirmseite dargestellt werden. Zur Visualisierung häufig zusammen auftretender Themen kann ein Graph erzeugt werden. Dabei rücken auf einer

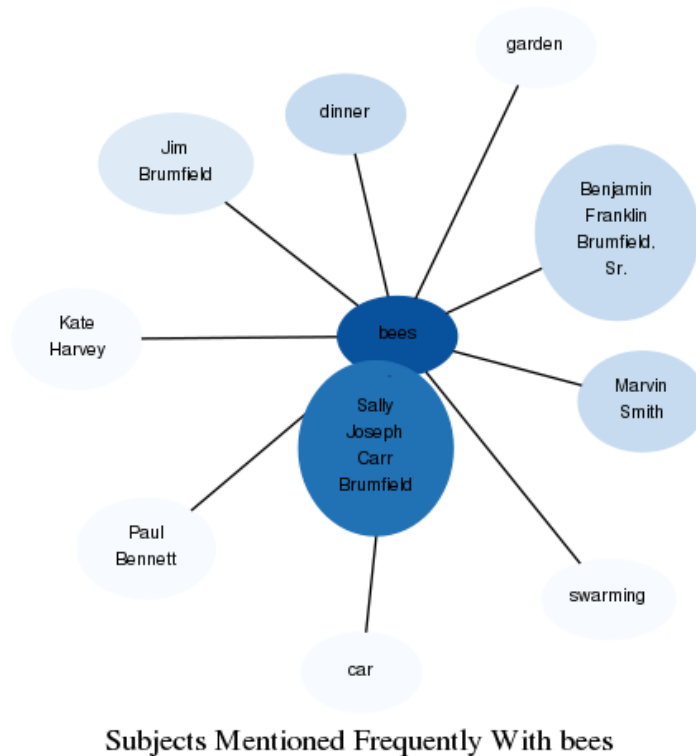


Abbildung 6.2: FromThePage: Graph

Tagebuchseite zusammen mit dem Suchwort auftretende Themen je näher an das Suchwort im Zentrum des Graphen heran, desto öfter sie zusammen mit ihm auftauchen. Zudem werden Themen in einem dunkleren Blau dargestellt, je näher sie zum Zentrum konvergieren (siehe Abbildung 6.2 Seite 43).

Für die zu indexierenden Themen und Namen (Subjects) wird im Vorhinein eine Hierarchie von Kategorien angelegt (Categories). Nach dem Abspeichern einer Transkriptionsseite können die Themen den entsprechenden Kategorien zugeordnet werden. Zur Indexierung im Transkriptionstext werden die zu indexierenden Wörter mit Hilfe der Wikilinks Syntax in doppelte eckige Klammern gesetzt: `[[Ben]]`. Um einen Verweis auf die Ansetzungsform eines Wortes zu setzen, wird die Ansetzungsform innerhalb der eckigen Klammern auf der linken Seite eingeben und die abweichende Schreibweise, getrennt durch einen senkrechten Strich, auf der rechten Seite: `[[Benjamin Franklin Brumfield, Sr.|Ben]]`. Die Funktion „Autolink“ vergleicht den Transkriptionstext mit bereits indexierten Texten und schlägt dem Bearbeiter Indexierungsmöglichkeiten vor (Textmining). Letzteres vereinfacht die Bearbeitung des Markup und sichert die Konsistenz der Ansetzungsformen.

FromThePage

## Julia Brumfield Diaries

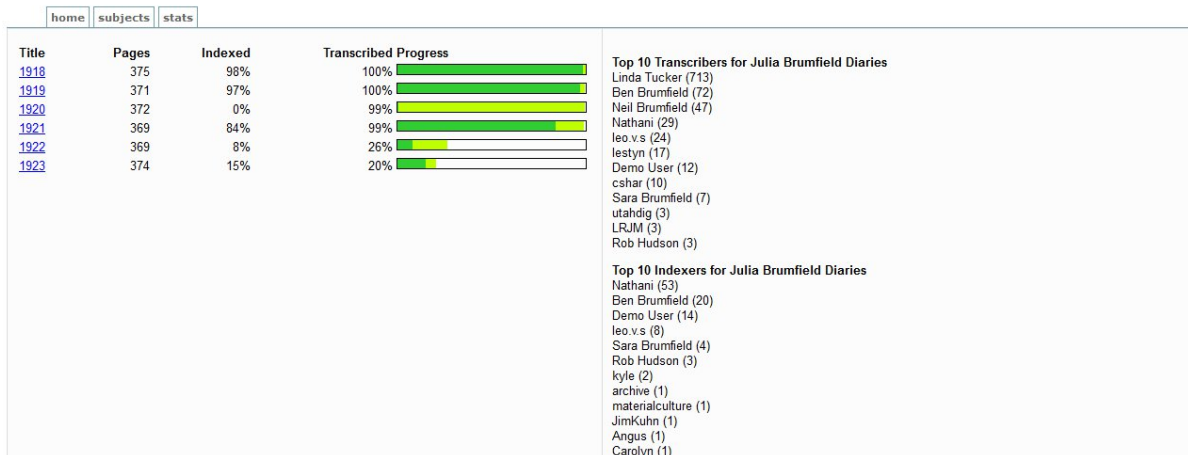


Abbildung 6.3: FromThePage: Bearbeitungsstatistik

Zu den einzelnen Sammlungen können Statistiken aufgerufen werden, die den jeweiligen Bearbeitungsstand darstellen (siehe Abbildung 6.3 Seite 44). Darüber hinaus werden die aktivsten zehn Bearbeiter der Transkriptionstexte (transcribers), Indexierer (indexers) und Editoren (editors) aufgelistet. Eine Gesamtansicht aller Bearbeiter kann ebenfalls angezeigt werden.

Im Bereich der Benutzerverwaltung bietet FromThePage momentan zwei Möglichkeiten zur Durchführung „privater“ Transkriptionsprojekte: Erstens kann der Kreis der Freiwilligen auf die vom Projekteigentümer zur Mitarbeit eingeladenen Personen beschränkt werden. In der zweiten Option wird das Projekt vollständig in einem privaten Rahmen organisiert, d. h. auch im reinen Lesemodus können die Inhalte lediglich von den zum Projekt eingeladenen Personen eingesehen werden<sup>104</sup>.

## Standards und Schnittstellen

Die Indexierung von Themen wird mit einem für das Projekt FromThePage spezifischen Markup ausgeführt. Darüber hinaus können grundlegende HTML-Elemente eingesetzt werden.

Mit Hilfe einer OAI-PMH-Schnittstelle können Metadaten aus dem Internet Archive<sup>105</sup> importiert werden. Die Metadaten werden in das native FromThePage-Format

<sup>104</sup>Vgl. Ben W. Brumfield. *Re: FromThePage Questions*. E-Mail. 18. Juli 2012.

<sup>105</sup>Vgl. Internet Archive, *Internet Archive*.



konvertiert und mit den vom Internet Archive bereitgestellten Scans assoziiert. Der Datenimport umfasst darüber hinaus weitere Bearbeitungsmöglichkeiten: (1) Vom Internet Archive als `type="Delete"` markierte Seiten wie Karten zur Farbkalibrierung können gelöscht werden. (2) (Gedruckte) Seitentitel<sup>106</sup> können auf die Ergebnisse des OCR-Laufs hin geparkt und die Metadaten entsprechend angepasst werden. (3) Die abschließende Konvertierung in das FromThePage-Format fasst das importierte Buch und seine Seiten zu einem Werk (work) in FromThePage zusammen. Dieses kann wiederum einer Sammlung (Collection) zugeordnet werden<sup>107</sup>.

### Systemvoraussetzungen

Die Installation von FromThePage auf einem eigenen Server kann unter Linux oder Windows als Plattform durchgeführt werden. Es benötigt Version 2.3.5 von Ruby on Rails sowie RMagick, hpricot, will\_paginate und OAI gems.

### Anwendungsbeispiele

Die Tagebücher von Julia Ann Craddock Brumfield<sup>108</sup> (geführt von 1915–1936) stellen den ersten Anwendungsfall für FromThePage dar und ihre Digitalisierung gab den Anstoß zur Entwicklung des Programms.

Das Balboa Park Online Collaborative am San Diego Natural History Museum setzt FromThePage momentan zur Transkription der Aufzeichnungen des Herpetologen Lawrence M. Klauber<sup>109</sup> ein. In dem Projekt sollen Notizen von Naturforschern transkribiert werden, um Vorkommen und Merkmale verschiedener Spezies in einer bestimmten Gegend mit aktuellen Daten zu vergleichen und Aufschluss über Veränderungen in den letzten 50 bis 100 Jahren zu geben. Die Dokumente wurden vom Internet Archive digitalisiert.

Ein weiteres Beispiel für den Einsatz der Software durch eine Institution sind die Kriegstagebücher von Zenas Matthews<sup>110</sup>, die von der Handschriftenabteilung der Bi-

---

<sup>106</sup>Im 20. Jahrhundert wurden häufig Tagebücher mit vorgedrucktem Seitentitel (Datumsangabe) benutzt. Diese sind für das Projekt FromThePage von besonderer Bedeutung.

<sup>107</sup>Vgl. San Diego Natural History Museum, *FromThePage: User Guide*.

<sup>108</sup>Vgl. Julia Ann Craddock Brumfield. *Julia Brumfield Diaries*. Hrsg. von Ben W. Brumfield. URL: [http://beta.fromthepage.com/JuliaBrumfield?ol=s\\_sp\\_diaries](http://beta.fromthepage.com/JuliaBrumfield?ol=s_sp_diaries) (besucht am 30.06.2012).

<sup>109</sup>Vgl. Laurence M. Klauber. *Field notes of herpetologist Lawrence Klauber*. Hrsg. von San Diego Natural History Museum. URL: <http://fromthepage.bpoc.org/> (besucht am 30.06.2012).

<sup>110</sup>Vgl. Zenas Matthews. *Zenas Matthews' 1846 U.S.-Mexico War Diary and Service Papers*. Hrsg. von Southwestern University Special Collections. URL: <http://beta.fromthepage.com/ZenasMatthews> (besucht am 30.06.2012).

bibliothek der Southwestern University zur Transkription bereitgestellt werden. Die Aufzeichnungen wurden während des Mexikanisch-Amerikanischen Kriegs (1846–1848) geführt.

## **Evaluation**

FromThePage stellt neben seiner ursprünglichen Orientierung an familiengeschichtlichen Transkriptionsprojekten auch für den Einsatz im institutionellen Rahmen eine bedenkenswerte Option dar.

Durch die Verknüpfung von Back- und Front-End kann der Benutzer seine Transkriptionen sofort im Internet publizieren, d. h. nach dem Abpeichern sind sie für jederman online erreichbar. Gleiches gilt für die Indexierung: die Erstellung von hierarchischen Kategorien und die Zuordnung von Begriffen zu diesen Kategorien wird umgehend auf der Präsentationsebene umgesetzt. Dieser Vorteil in Bezug auf die Geschwindigkeit des Workflow in der Standardkonfiguration wird allerdings nicht von allen Projekten als solcher eingeschätzt werden. Wissenschaftlich orientierte Editionen vertrauen häufig auf eine gründliche Redaktion der Texte und wollen diese vor der Online-Publikation durchführen. Die gegenwärtige Funktion zur Realisierung „privater“ Projekte (s. o.) zielt eher auf familiengeschichtlich orientierte Vorhaben ab, bei denen der Umgang mit sensiblen Personendaten die vollständige Publikation der Edition im Web u. U. verbietet. Ein Workflow mit der Möglichkeit zur Freischaltung einzelner Texte durch die Moderatoren eines Projektes wird derzeit nicht angeboten.

Mit Hilfe der Autolink-Funktion wird die Indexierung von Themen erheblich erleichtert. Sie macht eine Schaltfläche zum Einfügen der entsprechenden Textauszeichnung überflüssig. Im Fall der grundlegenden HTML-Elemente, die von dem Editor unterstützt werden, wäre eine entsprechende Werkzeugleiste aber eine zweckmäßige Ergänzung. Diese würde nicht nur die Codierung vereinfachen, sondern auch einen schnellen Überblick zu dem möglichen Markup bereitstellen, ohne den Blick in das Benutzerhandbuch erforderlich zu machen.

Das in FromThePage mit Hilfe der Indexierung durchgeführte Textmining bietet neben einer effizienten thematischen Suche nach dem Vorkommen bestimmter Themen in den Transkripten auch eine Visualisierung der häufig zusammen mit diesen Themen auftauchenden Begriffe in Form eines Graphen. Die Kategorien für die zu indizierenden Themen können projektspezifisch angepasst werden. Auch in diesem Fall wäre aber eine Zugriff des Nutzers auf die Kategorienverwaltung nicht in allen Projekten wünschenswert.

Der Verweis auf Normdaten wird unter FromThePage nicht unterstützt. Für viele historische Quellen wäre eine solche Funktion allerdings zweckmäßig. Auf diese Weise könnten etwa Ortsnamen mit der maßgebenden Ansetzungsform einer Normdatei verlinkt und mit bestehenden Ressourcen vernetzt werden. Momentan wird die Ansetzungsform eines indizierten Begriffes vom Bearbeiter selbst festgelegt. Dessen ungeachtet existieren Pläne und Anfragen verschiedener Interessierter zur Integration von Normdaten. Diese beinhalten Normdaten zur Codierung von Ortsnamen zur dynamischen Erstellung von Einträgen auf geographischen Karten, von Zeitangaben zur Erstellung von Zeitleisten, von Personennamen zum Verweis auf Identifikatoren von FamilySearch, von Artnamen zur Codierung taxonomischer Kategorien sowie von Künstlern zum Verweis auf Informationen aus DBpedia<sup>111</sup> und anderen Linked Open Data Directories<sup>112</sup>.

Die OAI-PMH-Schnittstelle zum Internet Archive bietet eine komfortable Möglichkeit zum Import dort vorliegender Bilder von Quellenmaterial. Darüber hinaus besteht ein Angebot zum Hosting selbst erstellter Scans. Diese Optionen machen eine Teilnahme besonders für kleine Projekte sehr unproblematisch. Die Installation der Software auf einem eigenen Server ist dagegen nicht ohne Kenntnisse in der Systemadministration möglich.

Ein Export von Metadaten z. B. zur Teilnahme an einem größeren Portal wie der Deutschen Digitalen Bibliothek<sup>113</sup> ist gegenwärtig nicht vorgesehen. Eine entsprechende Schnittstelle zum Export der Metadaten zu den enthaltenen Texten im XML-Format könnte aber ohne größeren Programmieraufwand geschaffen werden. Die verwendeten Metadaten basieren zum Teil auf Dublin Core<sup>114</sup>.

Die angebotenen Möglichkeiten zur Textauszeichnung beschränken sich auf ein flaches Markup mit der Codierung grundlegender semantischer und darstellungsbezogener Textmerkmale. In Bezug auf die Reichhaltigkeit der Codierung sind einem Programm mit Verknüpfung von Transkriptionsmodus und Präsentationsebene natürlicherweise Grenzen gesetzt, weil in der Standardkonfiguration nicht für alle erdenklichen Auszeichnungselemente eine entsprechende Stilvorlage für die Darstellung im Web vorgehalten werden kann.

FromThePage wird derzeit von Ben W. Brumfield entwickelt. Zudem erhält das Projekt momentan institutionelle Unterstützung durch (1) die Beisteuerung von Fehlerbehebungen durch Organisationen mit einer eigenen Version von FromThePage auf ihrem

---

<sup>111</sup>Vgl. DBpedia. *DBpedia*. URL: <http://wiki.dbpedia.org> (besucht am 20. 07. 2012).

<sup>112</sup>Vgl. Brumfield, *Re: FromThePage Questions*.

<sup>113</sup>Vgl. Deutsche Digitale Bibliothek, *Deutsche Digitale Bibliothek*.

<sup>114</sup>Vgl. Ben W. Brumfield. *Re: Master Thesis*. E-Mail. 19. Juli 2012.

Server (2) die Erteilung von Aufträgen zur Entwicklung von Anpassungen und Erweiterungen (3) die Bezahlung des Hostings eigener Inhalte auf FromThePage.com (4) Die Erteilung von Aufträgen zur Entwicklung von Anpassungen und Erweiterungen mit Hilfe von beantragten Zuschüssen<sup>115</sup>. Ein Ausbau der genannten Optionen zur institutionellen Zusammenarbeit mit FromThePage im Interesse einer Implementierung von Erweiterungen und der Sicherung des Angebots wäre für alle Beteiligten vorteilhaft.

### 6.1.2 Refine!

Der Editor Refine! wurde im Zusammenhang mit der digitalen Edition des lexikographischen Nachlasses Franz Brümmer<sup>116</sup> (1836-1923) entwickelt<sup>117</sup>. Er entstand aus einer Zusammenarbeit zwischen der Staatsbibliothek zu Berlin, der Humboldt-Universität sowie der Firma 3pc GmbH Neue Kommunikation<sup>118,119</sup>. Im folgenden wird die Funktionalität von Refine! am Beispiel des Transkriptionsprojekts zum Nachlass Franz Brümmer untersucht. Der Editor kann von 3pc für die Anforderungen anderer Projekte individuell angepasst werden.

Das Programm ermöglicht die Transkription von Handschriften als web-basierte Anwendung. Die Texte können mit Hilfe eines WYSIWYG-Editors bearbeitet und indexiert werden. Darüber hinaus fungiert es als Publikationssystem, d. h. die Texte können nach einer Qualitätskontrolle durch das Redaktionsteam ohne den Eingriff eines Administrators auf der Website veröffentlicht werden. Eine Versionskontrolle ermöglicht den Rückgriff auf ältere Bearbeitungsstufen.

Metadaten aus dem Verbundkatalog Kalliope<sup>120</sup> werden importiert und dem TEI-Header der Transkripte hinzugefügt<sup>121</sup>. Die Konvertierung in TEI erfolgt automatisiert und setzt bei den Bearbeitern keine Kenntnisse in diesem XML-Dialekt voraus<sup>122</sup>.

<sup>115</sup>Vgl. Brumfield, *Re: FromThePage Questions*.

<sup>116</sup>Vgl. Staatsbibliothek zu Berlin und Humboldt-Universität zu Berlin, *Nachlass Franz Brümmer*.

<sup>117</sup>Vgl. Björn Martin und Christian Thomas. „Das Wuchern der Archive. Die digitale Edition des Nachlasses Franz Brümmer mit dem Refine!Editor“. In: *editio* 22 (2008), S. 208.

<sup>118</sup>Vgl. 3pc GmbH Neue Kommunikation, *3pc GmbH Neue Kommunikation*.

<sup>119</sup>Vgl. Staatsbibliothek zu Berlin und Humboldt-Universität zu Berlin. *Projektbeschreibung der digitalen Edition des lexikographischen Nachlasses Franz Brümmer*. URL: <http://bruemmer.staatsbibliothek-berlin.de/nlbruemmer/projekt/> (besucht am 30.06.2012).

<sup>120</sup>Vgl. Staatsbibliothek zu Berlin, *Kalliope*.

<sup>121</sup>Vgl. Gregor Middell. *Vortrag auf dem COST Action 32 – Workgroup-2-Treffen in Ancona, Italien (19.-20. Mai 2008)*. URL: [http://bruemmer.staatsbibliothek-berlin.de/nlbruemmer/materialien/2008-05-19\\_COST\\_A32\\_W2\\_Vortrag\\_en.pdf](http://bruemmer.staatsbibliothek-berlin.de/nlbruemmer/materialien/2008-05-19_COST_A32_W2_Vortrag_en.pdf) (besucht am 30.06.2012), S. 5.

<sup>122</sup>Vgl. Staatsbibliothek zu Berlin und Humboldt-Universität zu Berlin, *Projektbeschreibung der digitalen Edition des lexikographischen Nachlasses Franz Brümmer*.

## Projekthintergrund

Franz Brümmer rief Schriftstellerinnen und Schriftsteller in verschiedenen literarischen Periodika zur Abgabe von Auskünften über ihre Biographie und bibliographische Angaben zu ihren Werken auf. Er wertete die ca. 6000 Zuschriften für sein Lexikon der deutschen Dichter und Prosaisten<sup>123</sup> aus und archivierte sie anschließend. Hinzu kommen etwa doppelt so viele Sekundärquellen. Die vielfältigen gesammelten Textmedien umfassen handschriftliche (Selbst-)Auskünfte, Typoskripte, Drucke, Abbildungen und Fotografien<sup>124</sup>.

Der Refine!Editor wurde entsprechend der Bedürfnisse des Nachlasses Franz Brümmer entwickelt. Dabei orientierte man sich an der physischen Ordnung des Nachlasses in Mappen und Kästen, welche so auch in vielen anderen Nachlässen und Sammlungen vorkommt. Deshalb kann die Software auch mit geringen Anpassungen zur Transkription anderer Nachlässe eingesetzt werden<sup>125</sup>.

## Funktionen

Im Transkriptionsmodus wird der Scan der Handschrift in einem Fenster dargestellt und das Transkript in einem zweiten Fenster mit Hilfe eines WYSIWYG-Editors eingegeben (siehe Abbildung 6.4 Seite 50). Die HTML-Codierung kann wahlweise in einem weiteren Fenster aufgerufen und bearbeitet werden. Der WYSIWYG-Editor ermöglicht die Formatierung des Textes z. B. als Fett- oder Kursivdruck sowie mit Durch- oder Unterstreichungen von Wörtern. Ebenso können Sonderzeichen eingefügt und Formatvorlagen z. B. für Überschriften angewendet werden. Die Indexierung kann ebenfalls über ein Symbol in der Werkzeugleiste durchgeführt werden. Dabei öffnet sich ein Fenster, welches die Angabe der Ansetzungsform bzw. die Übernahme einer Ansetzungsformen aus schon indexierten Mappen erlaubt. Im Nachlass Franz Brümmer werden Personen, Orte, Absendedaten, Institutionen, Periodika und Werke indexiert. Metadaten wie z. B. die Signatur des Autographen in der Handschriftenabteilung der Staatsbibliothek zu Berlin sowie PND-Nummer und Bemerkungen können ebenfalls bearbeitet werden.

Bei der Indexierung wird das entsprechende Wort markiert und eine Schaltfläche in der Werkzeugleiste benutzt. Daraufhin öffnet sich ein Dialogfenster zur Auswahl der

---

<sup>123</sup>Vgl. *Lexikon der deutschen Dichter und Prosaisten vom Beginn des 19. Jahrhunderts bis zur Gegenwart*. 6., völlig neu bearbeitete und stark vermehrte Auflage. 8 Bde. [Reprint (1975): Neudeln / Liechtenstein: Kraus]. Leipzig, 1913. URL: <http://archive.org/details/lexikonderdeutsc01bruoft> (besucht am 24.07.2012).

<sup>124</sup>Vgl. Martin und Thomas, „Das Wuchern der Archive“, S. 207.

<sup>125</sup>Vgl. ebd., S. 208.

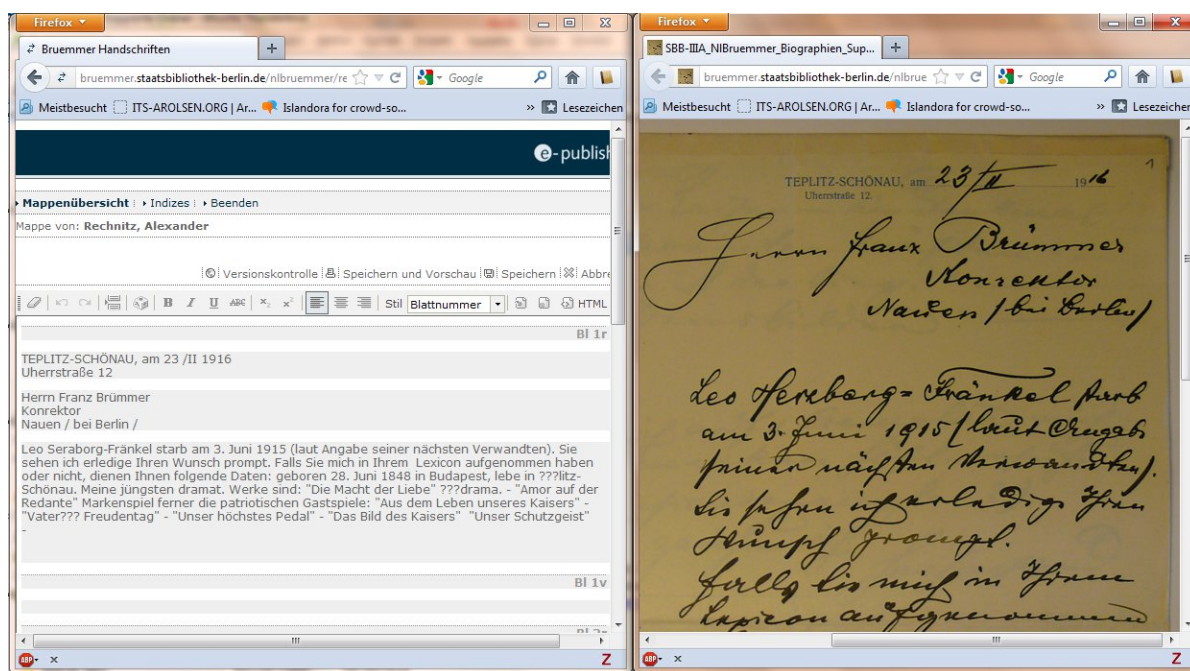


Abbildung 6.4: Refine!: Transkriptionsmodus

Indexkategorie. Die Kategorien wurden entsprechend den Anforderungen für den lexikographischen Nachlass Franz Brummer von der Firma 3pc angepasst. Indexiert werden dabei Personen, Orte, Absendedaten, Institutionen, Periodika und Werke.

Mit Hilfe der Nutzerverwaltung werden einzelne Mappen für einen bestimmten Bearbeiter freigegeben. Nach einem Lektorat durch das Redaktionsteam werden die Transkripte publiziert. Letzteres ist durch eine Verknüpfung des Editionswerkzeugs mit der offiziellen Website ohne weiteren Konvertierungsaufwand möglich<sup>126</sup>.

Der Metadaten austausch mit Kalliope<sup>127</sup> funktioniert in zwei Richtungen. Neue Metadaten aus der Nachlassedition werden von Mitarbeitern der Staatsbibliothek zu Berlin sowie Praktikanten mit Hilfe eines GUI-Clients in der Kalliope-Datenbank verzeichnet. Bereits in dem Verbundkatalog eingetragene Daten können mit einer Z39.50 Schnittstelle und dem Metadatenformat MAB2 abgefragt werden. Eine Umstellung auf- bzw. Ergänzung um das Metadatenformat EAD befindet sich in Vorbereitung und dieses Format wird von Kalliope bereits unterstützt<sup>128</sup>. Ein für die Nachlassedition entwickelter Webservice fungiert als Gateway zwischen Kalliope und der digitalen Edition. Dabei werden Metadaten mittels Z39.50 auf Abruf abgefragt, für spätere Abfragen im Cache

<sup>126</sup>Martin und Thomas, „Das Wuchern der Archive“, Vgl.

<sup>127</sup>Vgl. Staatsbibliothek zu Berlin, *Kalliope*.

<sup>128</sup>Vgl. Christian Thomas. *Re: Fragen zu Refine!* E-Mail. 24. Juli 2012.

gespeichert und nach RDF konvertiert. Schließlich werden die Metadaten mittels HTTP an Clients weitergeleitet<sup>129</sup>.

Der automatische Export nach TEI erzeugt TEI-P5-konforme Dokumente. Dabei werden sowohl das Transkript selbst, die semantischen Auszeichnungen aus der Indexierung und die von Kalliope importierten Metadaten in einer Datei aggregiert. Das Codebeispiel zeigt einen Auszug der aus Kalliope in den TEI-header übertragenen Metadaten (siehe Codebeispiel 6.1 Seite 52). Der header enthält in Bezug auf die Daten aus Kalliope die für alle Mappen in gleicher Weise gültige Beschreibung des Gesamt-Nachlasses sowie die (individuelle) Signatur der jeweiligen Mappe<sup>130</sup>. Bearbeiter und Daten zu der Mappe werden zudem variabel aus einer Datenbank generiert<sup>131</sup>. Neben dem Verfasser und dem Adressaten des Briefes wird im `<editionStmt>` die Bearbeiterin der Transkription angegeben und im `<publicationStmt>` das Datum der Erstveröffentlichung verzeichnet. Die `<sourceDesc>` enthält Titel und Signatur der physischen Vorlage des elektronischen Textes. Zur Anpassung des TEI-Schemas für das Transkriptionsprojekt wurden die folgenden Module ausgewählt: core, tei, header, textstructure, linking sowie namesdates<sup>132</sup>.

---

<sup>129</sup>Vgl. Middell, *Vortrag auf dem COST Action 32 – Workgroup-2-Treffen in Ancona, Italien (19.-20. Mai 2008)*, S. 6.

<sup>130</sup>Vgl. Thomas, *Re: Fragen zu Refine!*

<sup>131</sup>Vgl. ebd.

<sup>132</sup>Vgl. Middell, *Vortrag auf dem COST Action 32 – Workgroup-2-Treffen in Ancona, Italien (19.-20. Mai 2008)*, S. 7.

## Codebeispiel 6.1: Refine!: TEI-header

```
1 <teiHeader type="text">
2   <fileDesc>
3     <titleStmt>
4       <title>Nachl. Franz Brümmer/Biographien, Suppl. I/
        A-C</title>
5       <respStmt>
6         <name>Bachmair, Heinrich F. S.</name>
7         <resp>Verfasser/in</resp>
8       </respStmt>
9       <respStmt>
10        <name>Brümmer, Franz</name>
11        <resp>Adressat/in</resp>
12      </respStmt></titleStmt>
13    <editionStmt>
14      <edition>Digitale Edition</edition>
15      <respStmt>
16        <resp>Editor</resp>
17        <name>Cornelia Gruhn</name>
18      </respStmt>
19    </editionStmt>
20    <!--...-->
21    <sourceDesc>
22      <bibl>
23        <idno>01579614</idno>
24        <idno>Nachl. Franz Brümmer, Biogr. Suppl. I:
          Bachmair, Heinrich Franz</idno>
25      </bibl>
26    </sourceDesc>
27  </fileDesc>
28 </teiHeader>
```

Einzelne Mappen können mit Hilfe der Nutzerverwaltung für Bearbeiter freigeschaltet werden. Mappen, für die ein Bearbeiter nicht angemeldet ist, können zwar betrachtet, aber nicht bearbeitet werden.



Das Hosting von Transkriptionsprojekten kann prinzipiell von 3pc<sup>133</sup> als kommerzieller Dienstleister übernommen werden<sup>134</sup>.

## Standards und Schnittstellen

Die Texte werden zunächst in HTML codiert und können von jedem Nutzer der Website als TEI-P5-konforme Dateien exportiert werden. Beim Import von Metadaten aus Kalliope werden Daten mittels einer Z39.50 Schnittstelle abgefragt, nach RDF/XML konvertiert und in den header der TEI-Dateien eingefügt<sup>135</sup>.

## Systemvoraussetzungen

LAMP: Linux, Apache HTTP Server, MySQL und PHP<sup>136</sup>

## Anwendungsbeispiele

Neben der o.g. digitalen Edition des Nachlasses Franz Brümmer<sup>137</sup> wird der Refine!Editor zur Erstellung der digitalen Edition des Briefnachlasses des Berliner Buchhändlers, Verlegers und Schriftstellers Friedrich Nicolai (1733–1811)<sup>138</sup> eingesetzt.

## Evaluation

Der Editor Refine! besitzt zentrale Funktionen wie einen WYSIWYG-Editor, automatisierte Generierung von TEI-Dateien und eine Schnittstelle zum Metadaten austausch, die ihn für Projekte zum Crowdsourcing von Transkriptionen in Frage kommen lassen. Auf Grund seiner Entwicklung im Rahmen der digitalen Edition des Nachlasses von Franz Brümmer werden allerdings für den Einsatz in anderen Projekten Anpassungen unumgänglich sein. Von besonderem Interesse ist das Programm für Transkriptionsprojekte zu Nachlasseditionen, weil die verwendete Ordnungsstruktur aus Mappen sich auf die meisten Nachlässe anwenden lässt und die Schnittstelle zum Metadaten austausch mit Kalliope nachgenutzt werden könnte. Zudem könnten derartige Projekte von den Erfahrungen aus der Nachlassedition Franz Brümmer profitieren.

---

<sup>133</sup>Vgl. 3pc GmbH Neue Kommunikation, *3pc GmbH Neue Kommunikation*.

<sup>134</sup>Vgl. Thomas, *Re: Fragen zu Refine!*

<sup>135</sup>Vgl. Middell, *Vortrag auf dem COST Action 32 – Workgroup-2-Treffen in Ancona, Italien (19.-20. Mai 2008)*, S. 7.

<sup>136</sup>Vgl. Thomas, *Re: Fragen zu Refine!*

<sup>137</sup>Vgl. Staatsbibliothek zu Berlin und Humboldt-Universität zu Berlin, *Nachlass Franz Brümmer*.

<sup>138</sup>Vgl. Staatsbibliothek zu Berlin und Freie Universität Berlin. *Nachlass Friedrich Nicolai*. URL: <http://nicolai.host6.3-point.de/> (besucht am 02.07.2012).

Der WYSIWYG-Editor kommt der Zusammenarbeit mit Freiwilligen ohne weitere technische Vorkenntnisse entgegen. Durch eine optionale Anzeige und Bearbeitung der HTML-Codierung sind aber auch tiefere Eingriffe in die Textauszeichnung von Bearbeiterseite möglich. Die Arbeit mit zwei parallel geöffneten Fenstern (Editor und Viewer) ist allerdings ein wenig gewöhnungsbedürftig und könnte durch die Integration beider Elemente in einem Fenster benutzerfreundlicher gelöst werden.

Das Konzept der automatisierten Generierung von TEI-Dateien verbindet die Forderungen nach einer für den Freiwilligen einfach durchzuführenden Textauszeichnung und nach einem nachnutzbaren sowie interoperablen Codierungsformat. An dieser Stelle wird „das Beste aus zwei Welten“, d. h. einfach Benutzbarkeit und ein prinzipiell erweiterbarer Standard, mit Hilfe der Programmfunktionen vereint.

Die Integration von Normdaten wie PND und der Datenaustausch mit Kalliope erfüllen Anforderungen zur Integration und Vernetzung von Normdaten besonders für archivische und bibliothekarische Transkriptionsprojekte. Eine Erweiterung dieser Funktionalität etwa um Normdaten für geographische Bezeichnungen könnte diese Funktionalität sinnvoll ergänzen.

Ein Support wird über die Beratung durch die Moderatoren des Brümmer Nachlasses hinaus für den Editor momentan nicht angeboten. Das Programm selbst kann bereitgestellt werden, aber die erforderlichen Anpassungen müssten nutzerseitig bzw. durch eine Auftragsvergabe realisiert werden<sup>139</sup>. Eine Weiterentwicklung wäre im Interesse der Interoperabilität der Software wünschenswert. Die Chancen hierfür sind derzeit aber nicht abzusehen.

Das Redaktionssystem erlaubt eine Freischaltung einzelner Transkriptionen nach einer Qualitätskontrolle durch das Moderatorenteam ohne den Eingriff eines Administrators nötig zu machen. Damit ist das Programm für Projekte mit einem Bedarf nach einem entsprechenden editorischen Workflow gut geeignet. Die Verknüpfung von Back- und Front-End wirkt sich positiv auf die Benutzerfreundlichkeit auf dieser Ebene aus und ermöglicht eine effiziente Handhabung des Publikationsprozesses.

Zur vollen Crowdsourcing-Funktionalität fehlt dem Programm noch eine Wiedergabemöglichkeit für Bearbeiterstatistiken. Bei der digitalen Edition des Nachlasses von Franz Brümmer ist diese Forderung nicht so vordringlich, weil in der Hauptsache mit Studentengruppen im Rahmen von Seminaren an der Transkription gearbeitet wird. Für andere Projekte mit der Einbeziehung größerer Gruppen von Freiwilligen wäre dieses aber wünschenswert.

---

<sup>139</sup>Vgl. Thomas, *Re: Fragen zu Refine!*

### 6.1.3 Wikisource

Als Schwesterprojekt von Wikipedia bemüht sich Wikisource<sup>140</sup> um die Publikation gemeinfreier oder unter einer freien Lizenz stehender Quellentexte. Vorzugsweise werden Dokumente gesammelt, die als historische Quellen gelten können. Die Textvorlagen werden als Scans oder Digitalfotos hochgeladen, wenn notwendig transkribiert und für den Leser erschlossen. Zu den ergänzend eingebrachten Informationen zählen z. B. Einführungen, Strukturierungen der Texte, bibliographische Angaben und Anmerkungen<sup>141</sup>. In dieser Arbeit wird schwerpunktmäßig die deutschsprachige Version von Wikisource behandelt.

Für den Transkriptionsmodus wird in der deutschen Sprachversion für Briefe und Flugblätter eine Erweiterung zu MediaWiki<sup>142</sup> mit dem Namen Proofread Page<sup>143</sup> genutzt<sup>144</sup>. Scans werden Seite an Seite mit der Textbox für die Transkription angezeigt. Der Editor bietet eine Werkzeugleiste zur Textauszeichnung im Wiki Markup und mit Hilfe von HTML-Tags sowie eine Funktion zur Texterkennung für lateinische Schrift und Frakturschrift. Verweise auf Wikipedia-Artikel können mit Wikilinks erstellt werden.

#### Projekthintergrund

Das Projekt wurde am 24. November 2003 unter dem Namen „Project Sourceberg“ gegründet und gliedert sich in verschiedene Sprachversionen. Die Sprachversionen werden in Bezug auf Auswahl, Aufbereitung und Wiedergabe der Texte unterschiedlich verwaltet. Das deutsche Wikisource hat seit seiner Neukonzeptionierung 2006 den Anspruch, wissenschaftlich verwertbare Quellen im Internet aufzubereiten. Letzteres soll durch eine korrekte Angabe der Textquellen, das Bereitstellen von Scans und ein kollaboratives Korrekturlesen erreicht werden<sup>145</sup>.

Über die Funktionen im Transkriptionsmodus hinaus bietet das Portal Möglichkeiten zur Diskussion inhaltlicher und technischer Fragen. Eigene Projekte können eingebracht

---

<sup>140</sup>Vgl. Wikimedia Foundation, *Wikisource*.

<sup>141</sup>Vgl. Wikipedia Community. *Über Wikisource*. URL: [http://de.wikisource.org/wiki/Wikisource:%C3%9Cber\\_Wikisource](http://de.wikisource.org/wiki/Wikisource:%C3%9Cber_Wikisource) (besucht am 03.07.2012).

<sup>142</sup>Vgl. Wikimedia Foundation. *MediaWiki*. URL: [http://www.mediawiki.org/wiki/Main\\_Page](http://www.mediawiki.org/wiki/Main_Page) (besucht am 03.07.2012).

<sup>143</sup>Vgl. ThomasV. *Proofread Page*. URL: [http://www.mediawiki.org/wiki/Extension:Proofread\\_Page](http://www.mediawiki.org/wiki/Extension:Proofread_Page) (besucht am 03.07.2012).

<sup>144</sup>Ben W. Brumfield. *Wikisource for Manuscript Transcription*. Blog-Eintrag. 21. Juli 2010. URL: <http://manuscripttranscription.blogspot.de/2009/07/wikisource-for-manuscript-transcription.html> (besucht am 12.07.2012), Vgl.

<sup>145</sup>Vgl. Wikipedia Community. *Wikipedia-Artikel zu Wikisource*. URL: <http://de.wikipedia.org/wiki/Wikisource> (besucht am 03.07.2012).

und Scans hochgeladen werden. Wenn ein Projekt abgeschlossen ist und keine Korrekturen mehr nötig sind, kann es gegen weitere Veränderungen geschützt werden.

Unter die von der deutschen Wikisource gesammelten Texte, die frei von Urheberrechten sind (i.d.R. 70 Jahre nach Tod des Autors) oder unter einer freien Lizenz stehen, fallen neben bereits anderweitig veröffentlichten Publikationen auch unveröffentlichte Texte, die z. B. in Manuskriptform vorliegen. Weiterhin sollte es sich um historische Quellen handeln, d.h. für eine Aufnahme bei Wikisource sollten die Texte grundsätzlich älter als 70 Jahre sein. Zudem wird in Bezug auf die Wahl der Textgrundlage auf der Website bemerkt, dass ihre Auswahl sorgfältig erfolgen und die wissenschaftliche Verwertbarkeit gesichert sein sollte<sup>146</sup>.

Als historische Quellen werden in der deutschen Sprachversion von Wikisource grundsätzlich auch Quellen wie Tagebücher angesehen, die für familiengeschichtlich orientierte Projekte von Interesse sein können. Sie sollten allerdings eine (historische) Relevanz besitzen und verifizierbar sein. Zudem wird es von der Wikisource Community begrüßt, wenn schon vor dem Einstellen Freiwillige zur Transkription und zum Korrekturlesen bereitstehen<sup>147</sup>.

## Funktionen

Die Transkription wird in eine Klartextbox eingegeben und der Scan kann in vertikaler oder horizontaler Ausrichtung zur Textbox angezeigt werden (siehe Abschnitt 6.5 Seite 57). Die in MediaWiki erlaubten HTML-Tags sowie die MediaWiki Syntax können entweder direkt in den Text geschrieben oder mit Hilfe einer Werkzeugleiste eingefügt werden. Es gibt Schaltflächen für den Fett- und Kursivdruck, zur Vergabe von Formatvorlagen für Überschriften, für Tabellen, zur Einbettung von Bildern, zur Eingabe von Sonderzeichen usw. Der Transkriptionsmodus ist mit der Präsentationsebene verknüpft und erlaubt eine Vorschauanzeige des gerenderten Textes. Die Darstellung von Marginalien und unterschiedlichen Schriftarten zur Nachempfindung des ursprünglichen Layouts wird ebenso unterstützt (siehe Abschnitt 6.6 Seite 57).

Die Arbeit mit Hyperlinks ist im Rahmen einer Verlinkung der Seitenzahlen mit den entsprechenden Scans möglich. Ferner werden Fußnoten als Hyperlinks eingefügt. Eine automatische Erstellung von Indizes kann derzeit nicht vorgenommen werden, aber es

---

<sup>146</sup>Vgl. Wikipedia Community, *Über Wikisource*.

<sup>147</sup>Jens Brokfeld. *Quellen zur Familiengeschichtsforschung auf Wikisource?* Forumsbeitrag. 2. Aug. 2012. URL: [https://de.wikisource.org/wiki/Wikisource:Skriptorium#Quellen\\_zur\\_Familiengeschichtsforschung\\_auf\\_Wikisource.3F](https://de.wikisource.org/wiki/Wikisource:Skriptorium#Quellen_zur_Familiengeschichtsforschung_auf_Wikisource.3F) (besucht am 02.08.2012), Vgl.

Benutzerkonto anlegen Anmelden

Seite Diskussion Scan Lesen Bearbeiten Versionsgeschichte Suche

## Bearbeiten von „Seite:Hexengeständnisse 1564 Blatt 149r.jpg“

Achtung: Du bearbeitest den Artikel unangemeldet. Statt eines Benutzernamens wird deine aktuelle IP-Adresse in der Versionsgeschichte aufgezeichnet. (Zum Anmelden klicke einfach oben rechts auf „Anmelden“.)

F K OROCR

Erweitert Sonderzeichen Hilfe

Erstlich bekent Anna Beglerin von werleschwang daß sy bei den 20 jaren ein vnholden gewesen sey vnd sie sich daß Erstmall hinder den laidigen teuffel Ergeben. von der vrsach wegen, das die herschafft hat allen vnderthannen anzaigen lassen war od(er) wölch am Juden legen, daß sy die selben solltten in Ainer genannten Zeit, von dem Juden wider Ledig machen. Dammals were sy gedachte Böglerin, vnd Jr man auch an dem Juden gelegen, vnd sy souil vnmuets darab genomen, das Er von hauß hat lassen wöllen, In dem ward sie gen Zußerhausen,

Brand Gerichtet worden Anno 1564.

Abbildung 6.5: Wikisource: Transkriptionsmodus

nsen, nach der Oder hin, mit schönen Wiesen, und einen hohen breiten und mit . Unser Marsch führte uns bis nahe an die Vorstädte Crohsens, doch von dort Neubrück<sup>[6]</sup> paßten wir am 28<sup>ten</sup> ejßdem<sup>[7]</sup> den nicht ganz unbedeutenden brück liegt sehr romantisch; von der rechten Seite mit bedeutenden Anhöhen merwerken mit PappelAlléen versehen, die sich mit der Straße den Fluß h Nettkow<sup>[8]</sup>, einem Flecken der Herzogin von Curland gehörig, betraten wir ndenburger, und laßen keinen bedeutenden Übergang aus einer so verschiedenen chtbarer wird, der Weinstock sich im Ganzen mehr angesiedelt befindet, und der t, nur dann erst vermißt man lebhafter die Mark Brandenburg und überzeugt sich,

Marsch durch  
einen Theil  
der Marck und  
Nieder  
Schlesien.

Abbildung 6.6: Wikisource: Textdarstellung mit Marginalien

wird von den Initiatoren des Projektes „Field Notes of Junius Henderson“<sup>148</sup> (siehe Abschnitt Anwendungsbeispiele) im Zusammenhang mit der Annotierung von Artnamen und anderen taxonomischen Kategorien daran gearbeitet. Die Erweiterungen und Ziele für die neuen Funktionen werden auf einem sogenannten Portal dokumentiert<sup>149</sup>. Ein Index kann momentan lediglich händisch erstellt werden und eine Verlinkung der aufgenommenen Begriffe ist in diesem Fall nicht möglich. Der Benutzer muss im Text zu den im Index angegebenen Seiten scrollen, um den gesuchten Begriff in seinem Kontext lesen zu können<sup>150</sup>.

Des Weiteren wird die Arbeit mit Wikilinks unterstützt. So wird bei der Auszeichnung etwa eines Artnamens eine Verlinkung auf den entsprechenden Wikipedia Artikel erstellt. Dabei wird in der Textauszeichnung der wissenschaftlichen Artname sowie die ursprüngliche Schreibweise im Text angegeben: `[[w:Passer domesticus|English sparrows]]`<sup>151</sup>. Zudem wird beim Schieben des Mauszeigers über den Link (mouseover) der wissenschaftliche Artname eingeblendet.

In der Wikisource Community werden kontinuierlich neue Vorlagen entwickelt. Ein weiteres Beispiel stellt die Verlinkung eines Ortsnamens zu OpenStreetMaps<sup>152</sup> dar. In doppelten geschweiften Klammern wird dabei der Typ der Verlinkung (hier: place), der Ortsname in Ansetzungsform und in der Vorlageform angegeben. Eine beispielhafte Auszeichnung dieser Art sieht folgendermaßen aus: `{{place|Florissant Lake|Florissant Lake basin}}`. In der Leseversion des Textes (Präsentationsebene) werden bei Benutzung dieser Vorlage in einer kleinen Box Links zu Wikimedia Commons<sup>153,154</sup> und OpenStreetMaps angezeigt. Ein Klick auf das Symbol für OpenStreetMaps öffnet eine Karte mit dem entsprechenden Kartenausschnitt.

Beim Anlegen einer neuen Seite können Metadaten eingegeben werden. Diese werden in einer Textbox am rechten Rand der jeweiligen Seite wiedergegeben. Es existieren Vorlagen für beschreibende Informationen zu einem neuen Text (Titel, Autor, Original-

<sup>148</sup>Vgl. Wikisource Community. *Field Notes of Junius Henderson*. URL: [http://en.wikisource.org/wiki/Field\\_Notes\\_of\\_Junius\\_Henderson](http://en.wikisource.org/wiki/Field_Notes_of_Junius_Henderson) (besucht am 12.07.2012).

<sup>149</sup>Vgl. Wikisource Community. *Portal:Taxonomy*. URL: <http://en.wikisource.org/wiki/Portal:Taxonomy> (besucht am 12.07.2012).

<sup>150</sup>Vgl. Brumfield, *Wikisource for Manuscript Transcription*.

<sup>151</sup>Vgl. Ben W. Brumfield. *Developments in Wikisource/ProofreadPage for Transcription*. URL: <http://manuscripttranscription.blogspot.de/2011/12/developments-in-wikisourceproofreadpage.html> (besucht am 12.07.2012).

<sup>152</sup>Vgl. OpenStreetMap Community. *OpenStreetMap*. URL: <http://www.openstreetmap.org/> (besucht am 13.07.2012).

<sup>153</sup>Medienarchiv für alle Wikimedia-Projekte

<sup>154</sup>Vgl. Wikimedia Foundation. *Wikimedia Commons*. URL: <http://commons.wikimedia.org/wiki/Hauptseite?uselang=de> (besucht am 13.07.2012).

herkunft etc.), einem neuen Autor (Nachname, Vorname, Geburtsdatum etc.), für eine neue Themenseite (Thema, Kurzbeschreibung, Bild etc.) sowie für eine neue Ortsseite (Ortsname, Kurzbeschreibung, Alternativnamen etc.)<sup>155</sup>. Der Code für die Metadaten zu einem Personennamen wird hier am Beispiel des Schriftstellers Rudolf Gottschall wiedergegeben (siehe Abschnitt 6.2 Seite 59). In Zeile 16 wird auch die PND zu dem Autor referenziert. Weitere Möglichkeiten zum Verweis auf Normdaten bietet die Arbeit mit GND und VD17.

#### Codebeispiel 6.2: Wikisource: Metadaten

```

1 {{Personendaten
2 |NACHNAME=Gottschall
3 |ANFANGSBUCHSTABE=G
4 |VORNAMEN=Rudolf
5 |ALTERNATIVNAMEN=
6 |KURZBESCHREIBUNG=Deutscher Dramatiker, Epiker, Erzähler,
   Literarhistoriker und Kritiker.
7 |SONSTIGES=
8 |GEBURTSDATUM=30. September 1823
9 |GEBURTSORT=Breslau
10 |STERBEDATUM=21. März 1909
11 |STERBEORT=[[Leipzig]]
12 |BILD=Rudolf Gottschall.jpg
13 |WIKIPEDIA=Rudolf Gottschall
14 |WIKIQUOTE=
15 |COMMONS=
16 |PND=119449072
17 }}
```

Im Workflow zum Korrekturlesen kann das Dokument als unkorrigiert, korrigiert oder fertig gekennzeichnet werden. Auf einer Diskussionsseite werden Probleme oder Kommentare besprochen. Eine Versionskontrolle ist ebenfalls vorhanden.

Neben der Volltextsuche existieren verschiedene Hilfsmittel zur Suche nach einem bestimmten Thema oder einer Quellengattung. Dazu zählen der Systematische Einstieg (Kategorien: Fach, Entstehungszeit, Entstehungsort, Sprache, Textgattung), eine Auflistung der vorhandenen Gedichte sowie alphabetische Listen der Autoren und Werke. Auf

<sup>155</sup>Vgl. Wikisource Community. *Wikisource:Neue Seite anlegen*. URL: [http://de.wikisource.org/wiki/Wikisource:Neue\\_Seite\\_anlegen](http://de.wikisource.org/wiki/Wikisource:Neue_Seite_anlegen) (besucht am 13.07.2012).

der Hauptseite werden kürzlich hinzugefügte Quellen, aktuelle Hinweise und eine Auswahl aus dem Bestand vorgestellt. Zudem können zufällige Seiten mit Anzeige von Scan und Transkription einen Eindruck von den bereits vorhandenen Inhalten verschaffen.

Im Rahmen der Statistik wird auf der Hauptseite die Anzahl der verfügbaren Werke sowie die Anzahl der in den verschiedenen Quellengattungen vorliegenden Dokumente verzeichnet. Zur Motivationssteigerung der Freiwilligen werden die sogenannten Korrekturen des Monats auf einer speziellen Seite vorgestellt und besonders aktiven Bearbeitern ein „Blümchen-Award“ verliehen<sup>156</sup>

## Standards und Schnittstellen

Für die Textauszeichnung wird Wikitext eingesetzt. Im Rahmen der Weitergabe von Metadaten an Web-Portale wird an den technischen Voraussetzungen für eine Datenlieferung etwa an das Zentrale Verzeichnis Digitalisierter Drucke<sup>157</sup> gearbeitet. Im Vordergrund steht dabei die Abbildung des in Wikisource verwendeten Metadatenformats auf Standardformate wie Dublin Core oder MAB sowie die Integration einer OAI-PMH-Schnittstelle<sup>158</sup>.

## Systemvoraussetzungen

Alternativ zu einem Hosting durch Wikisource kann die freie Software auch auf eigenen Servern installiert werden. Zur Installation von Proofread Page wird zunächst MediaWiki<sup>159</sup> benötigt. Letzteres erfordert in der Standard-Installation einen Web-Server, PHP und einen Datenbank-Server<sup>160</sup>. Die Installation und Konfiguration von Proofread Page als Erweiterung von MediaWiki wird hier<sup>161</sup> beschrieben. In einem Blog-Eintrag zur lokalen Installation wird die Konfiguration der nötigen Erweiterungen erläutert<sup>162</sup>.

---

<sup>156</sup>Vgl. Wikisource Community. *Korrekturen des Monats*. URL: <http://de.wikisource.org/wiki/Vorlage:Reviewtext> (besucht am 12.07.2012).

<sup>157</sup>Vgl. Zentrales Verzeichnis Digitalisierter Drucke. *Zentralen Verzeichnisses Digitalisierter Drucke*. URL: <http://www.zvdd.de/> (besucht am 26.07.2012).

<sup>158</sup>Vgl. Wikisource Community. *Wikisource:Metadaten*. URL: <http://de.wikisource.org/wiki/Wikisource:Metadaten> (besucht am 26.07.2012).

<sup>159</sup>Vgl. Wikimedia Foundation, *MediaWiki*.

<sup>160</sup>Vgl. Wikimedia Foundation. *Installationsanforderungen für MediaWiki*. URL: [http://www.mediawiki.org/wiki/Manual:Installation\\_requirements](http://www.mediawiki.org/wiki/Manual:Installation_requirements) (besucht am 12.07.2012).

<sup>161</sup>ThomasV. *Beschreibung der Installation und Konfiguration von Proofread Page*. URL: [http://www.mediawiki.org/wiki/Extension:Proofread\\_Page](http://www.mediawiki.org/wiki/Extension:Proofread_Page) (besucht am 12.07.2012), Vgl.

<sup>162</sup>Vgl. Roderick D.M. Page. *Setting up a local Wikisource*. URL: <http://iphylo.blogspot.de/2010/03/setting-up-local-wikisource.html> (besucht am 26.07.2012).



## Anwendungsbeispiele

Das Projekt „Field Notes of Junius Henderson“<sup>163</sup> hat die Transkription der Aufzeichnungen des Naturforschers Junius Henderson (1905–1931) zum Ziel. Die Fortschritte bei der Transkription und bei projektspezifischen Anpassungen werden in einem Blog dokumentiert<sup>164</sup>. Einige der o. g. Vorlagen<sup>165</sup> wurden im Rahmen dieses Projekts erstellt. Ein Artikel zur wissenschaftlichen Verwendung von Wikisource im Rahmen dieses Projekts beschreibt die entwickelten Möglichkeiten der Annotierung und Extraktion von Artbeschreibungen en detail<sup>166</sup>.

Das Nationalarchiv der Vereinigten Staaten (National Archives and Records Administration<sup>167</sup>) hat im Sommer 2011 ein Transkriptionsprojekt in Zusammenarbeit mit Wikisource gestartet. Dabei wurden digitale Text- und Audiodokumente aus den Beständen des Archivs auf Wikimedia Commons hochgeladen und auf Wikisource zur Transkription bereitgestellt. Mehr als 20.000 Seiten wurden hochgeladen und bis jetzt sind 45 Dokumente vollständig transkribiert, validiert und die Transkriptionen dem NARA Online-Katalog hinzugefügt worden<sup>168</sup>.

## Evaluation

die deutsche Sprachversion von Wikisource beruht als Gemeinschaftsprojekt auf der Mitarbeit von Freiwilligen sowohl im Hinblick auf die technische Weiterentwicklung als auch die inhaltliche Arbeit an neuen Projekten mit der Transkription und Annotation von historischen Quellen. Mit dieser Ausgangssituation genießt das Projekt die Unterstützung einer engagierten Nutzer-Community, die neue Entwicklungen vorantreibt und projektspezifische Anpassungen realisiert. Die Kooperation mit institutionellen Partnern ist aber ebenso zum beiderseitigen Vorteil möglich und bietet etwa für Archive, Bibliotheken und Museen eine Plattform zur Publikation und Transkription ihrer historischen

<sup>163</sup>Vgl. Wikisource Community, *Field Notes of Junius Henderson*.

<sup>164</sup>Vgl. Rob Guralnick und Andrea Thomer. *So You Can Think You Can Digitize*. Blog. URL: <http://soyouthinkyoucandigitize.wordpress.com/> (besucht am 13.07.2012).

<sup>165</sup>Verlinkung von Artnamen mit entsprechendem Wikipedia-Artikel und Verlinkung von Ortsnamen mit OpenStreetMap

<sup>166</sup>Vgl. Andrea Thomer u. a. „From documents to datasets: A MediaWiki-based method of annotating and extracting species observations in century-old field notebooks“. In: *ZooKeys* 209 (2012), 235–253. DOI: 10.3897/zookeys.209.3247. URL: <http://www.pensoft.net/journals/zookeys/article/3247/> (besucht am 26.07.2012).

<sup>167</sup>Vgl. National Archives and Records Administration. *National Archives and Records Administration*. URL: [http://outreach.wikimedia.org/wiki/GLAM/Model\\_projects/Improving\\_the\\_quality\\_of\\_OCR](http://outreach.wikimedia.org/wiki/GLAM/Model_projects/Improving_the_quality_of_OCR) (besucht am 14.07.2012).

<sup>168</sup>Vgl. Wikipedia Community. *GLAM Model projects*. URL: [http://outreach.wikimedia.org/wiki/GLAM/Model\\_projects/Improving\\_the\\_quality\\_of\\_OCR](http://outreach.wikimedia.org/wiki/GLAM/Model_projects/Improving_the_quality_of_OCR) (besucht am 14.07.2012).

Quellen (siehe Anwendungsbeispiele). Der hohe Bekanntheitsgrad des Mutterprojektes Wikipedia kann darüber hinaus zur Öffentlichkeitswirksamkeit von Transkriptionsprojekten auf Wikisource beitragen.

Die Übersichtlichkeit von Wikisource könnte noch verbessert werden, aber die kollaborative Arbeitsweise und die Integration vieler verschiedener Projekte erschwert die Umsetzung solcher Forderungen.

Die Textauszeichnung mit ProofreadPage ermöglicht neben der Nachempfindung bestimmter Layout-Merkmale der Vorlage auch die Verlinkung der Inhalte mit anderen Ressourcen etwa der Wikipedia oder OpenStreetMap. Dieses Vorgehen trägt zur Vernetzung der Inhalte bei und bietet für den Nutzer eines Editionstextes ein reichhaltiges Hilfsmittel für die Recherche.

Die Vergabe von Metadaten zu Textseiten, Autoren, Themenseiten und Ortsseiten erhöht die Aussagekraft der gespeicherten Inhalte für den Nutzer. Darüber hinaus sind die begonnenen Bemühungen um eine Weitergabe der Metadaten an Web-Verzeichnisse eine folgerichtige Maßnahme zur Ermöglichung gezielter Recherchen sowie zur Vernetzung vorhandener Datenbestände. Eine Weiterentwicklung dieser Funktionen würde im Besonderen Bibliotheken und Archive bei ihrem Bestreben nach der Verwendung einheitlicher Standards unterstützen und schließlich auch dem Nutzer im Rahmen einer zielgerichteten Recherche mit möglichst vielen für sein Thema relevanten Suchergebnissen zugutekommen.

Der in den Programmfunktionen integrierte Workflow zum Korrekturlesen mit der Kennzeichnung von Transkriptionen als unkorrigiert, korrigiert oder fertig stellt einen Versuch zur Einführung einer Qualitätskontrolle in die kollaborative Arbeitsweise von Wikisource dar. Dieses Hilfsmittel ist für die Erstellung qualitativ hochwertiger Transkripte im Rahmen von Wikisource unerlässlich.

#### **6.1.4 Scripto**

Die freie Software Scripto ist ein reines Transkriptionswerkzeug, d. h. es stellt Transkriptionsfunktionen, eine Versionskontrolle und weitere Kontrollfunktionen zur Verwaltung der Transkriptionen zur Verfügung. Neben der Transkription von textuellen Quellen kann mit dem Plugin auch der Inhalt von Audio- und Videodateien erfasst werden. Scrip-

to ist eine PHP Bibliothek (library) auf Basis von MediaWiki<sup>169</sup>. Es kann als Plugin in bestehende CMS wie Omeka<sup>170</sup>, WordPress<sup>171</sup> und Drupal<sup>172</sup> eingebunden werden<sup>173</sup>.

## Projekthintergrund

Die Beta-Version von Scripto wurde auf Grundlage einer Anwendung der Alpha-Version im Rahmen des Projekts Papers of the War Department (1784–1800)<sup>174</sup> entwickelt<sup>175</sup>. Verantwortlich für die Entwicklung des Transkriptionswerkzeugs ist das Ray Rosenzweig Center for History and New Media<sup>176</sup>.

## Funktionen

Nach der Integration von Scripto in ein CMS können die dort vorhandenen digitalen Objekte mit dem Plugin eingelesen und kollaborativ transkribiert werden. Scripto enthält eine Benutzerschnittstelle zur Transkription sowie zum Management der Transkripte. Über ein API ist das Programm mit MediaWiki verbunden. Auch in MediaWiki können weitere Einstellungen zur Verwaltung der Transkripte sowie für administrative Aufgaben vorgenommen werden<sup>177</sup> (siehe Abbildung 6.7 Seite 64).

Die Funktionsbereiche von Scripto gliedern sich in die für Moderatoren oder Editoren des Transkriptionsprojekts vorgesehenen Aufgaben und die für Bearbeiter zugänglichen Funktionen.

Im Transkriptionsmodus wird das aus der im CMS verwalteten digitalen Sammlung ausgewählte Objekt (im Fall von textuellen Quellen) in einem Viewer mit Zoomfunktion angezeigt. Bei mehrseitigen Dokumenten können die vorhandenen Scans durchgeblättert werden. In der Klartextbox zur Eingabe der Transkription kann der Text auch mit Wikitext sowie den von MediaWiki unterstützten HTML-Tags ausgezeichnet werden (siehe Abbildung 6.8 Seite 64).

---

<sup>169</sup>Vgl. Center for History and New Media, *Scripto*.

<sup>170</sup>Vgl. Roy Rosenzweig Center for History and New Media, *Omeka*.

<sup>171</sup>Vgl. inpsyde, *WordPress*.

<sup>172</sup>Vgl. Drupal Community, *Drupal*.

<sup>173</sup>Vgl. Center for History and New Media. *About Scripto*. URL: <http://scripto.org/documentation/about/> (besucht am 21. 07. 2012).

<sup>174</sup>Vgl. Roy Rosenzweig Center for History and New Media. *Papers of the War Department (1784-1800)*. URL: <http://wardepartmentpapers.org/index.php> (besucht am 14. 07. 2012).

<sup>175</sup>Vgl. Konrad Lawson. *Crowdsourcing Transcription: FromThePage and Scripto*. Blog-Eintrag. 23. Jan. 2012. URL: <http://chronicle.com/blogs/profhacker/crowdsourcing-transcription-fromthepage-and-scripto/38028> (besucht am 14. 07. 2012).

<sup>176</sup>Vgl. Roy Rosenzweig Center for History and New Media. *Roy Rosenzweig Center for History and New Media*. URL: <http://chnm.gmu.edu/> (besucht am 21. 07. 2012).

<sup>177</sup>Vgl. Center for History and New Media, *About Scripto*.

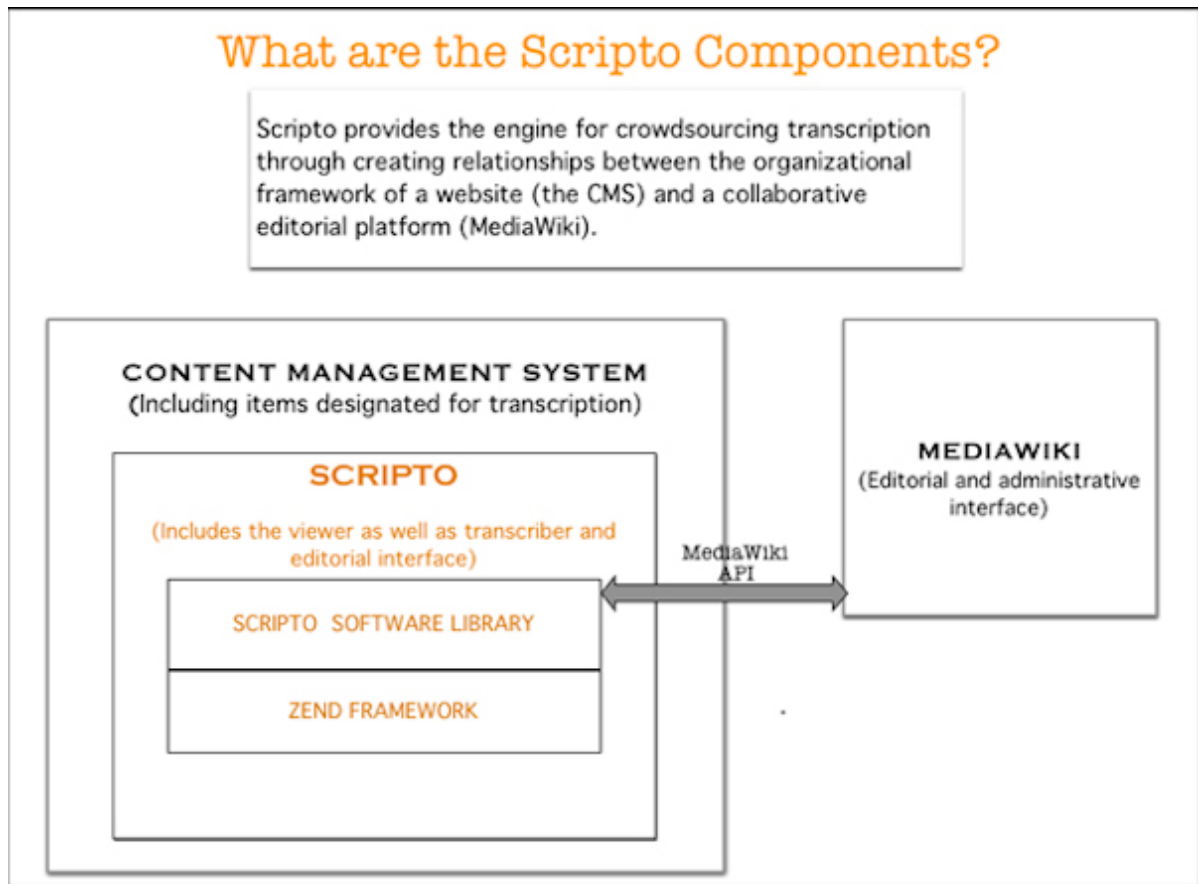


Abbildung 6.7: Scripto: Komponenten



Abbildung 6.8: Scripto: Transkriptionsmodus

| Recent Changes   |                              |                             |         |                |  |
|--|------------------------------|-----------------------------|---------|----------------|--|
| Changes  | Document Page Name           | Changed on                  | Changed | Changed By     | Document Title                                     |
| Edited<br>( <a href="#">diff</a>   <a href="#">hist</a> )  | <a href="#">letter2.pdf</a>  | 21:22:43<br>Jun 10,<br>2012 | +30     | 173.48.204.14  | <a href="#">Letter from Mar... her sister, Sue</a> |
| Created<br>( <a href="#">diff</a>   <a href="#">hist</a> ) | <a href="#">letter2.pdf</a>  | 21:19:26<br>Jun 10,<br>2012 | +367    | 173.48.204.14  | <a href="#">Letter from Mar... her sister, Sue</a> |
| Edited<br>( <a href="#">diff</a>   <a href="#">hist</a> )  | <a href="#">letter1.pdf</a>  | 21:14:54<br>Jun 10,<br>2012 | +351    | 173.48.204.14  | <a href="#">Letter from Mar... her sister, Sue</a> |
| Edited<br>( <a href="#">diff</a>   <a href="#">hist</a> )  | <a href="#">IMG_0013.jpg</a> | 16:17:56<br>Jun 05,<br>2012 | +6      | 81.234.227.215 | <a href="#">V Mail Folmer S... . Alfred Madsen</a> |

Abbildung 6.9: Scripto: Versionsgeschichte

Weiterhin kann die Versionsgeschichte vom Bearbeiter eingesehen werden. Dabei wird ein Vergleich der aktuellen Transkription mit einer früheren Version ermöglicht und Datum und Uhrzeit der Veränderung, die Identität des Bearbeiters, das Ausmaß der Veränderung sowie die Art der getätigten Aktion (erstellt, bearbeitet oder gesichert) ausgegeben.

Zum Austausch der Bearbeiter über schwer lesbare Wörter o. ä. steht eine Diskussionsfunktion bereit. Die Anzeige der Diskussionen zu einem digitalen Objekt wird mit denselben Metadatenkategorien wie zur Versionsgeschichte der Transkriptionen wiedergegeben. Ebenso ist die Anzeige der letzten Veränderungen an allen digitalen Objekten eines Projektes möglich (siehe Abbildung 6.9 Seite 65). Durch die Aufnahme von Objekten auf eine Merkliste (Watchlist) können angemeldete Bearbeiter die Veränderungen an für sie interessanten Transkriptionen im Auge behalten.

Die Moderatoren des Projektes können nach Anmeldung mittels der Omeka-Administratorschnittstelle auf eine Reihe von editorischen Funktionen zurückgreifen. Sie haben in diesem Modus Zugriff auf alle Informationen, die auch den Bearbeitern zur Verfügung stehen. Dazu zählt die Versionskontrolle und die Anzeige der Diskussionen

zu bestimmten Objekten. Ferner können sie Transkriptionen bearbeiten und an Diskussionen teilnehmen sowie eine eigene Merkliste anlegen.

Zu den Funktionen, die ausschließlich den Moderatoren vorbehalten sind, gehören: Das Öffnen einer Transkription in MediaWiki für weitere Einstellungen, die Sicherung bestimmter Transkriptionen vor weiteren Veränderungen und der Import der Texte in das Omeka-Front-End.

Im Fall von Omeka werden die Metadaten zu den digitalen Objekten im Dublin-Core-Standard abgelegt. Zu den verfügbaren Datenfeldern gehören Title, Subject, Description, Creator, Source, Publisher, Dates, Contributor, Rights, Relation, Format, Language, Type, Identifier und Coverage<sup>178</sup>. Mit Hilfe dieser Angaben können digitale Objekte auf einer Omeka-Website z. B. nach dem entsprechenden Format (Buch, Manuskript, Bild, Video etc.) geordnet werden und darüber hinaus erhält der Nutzer eine Beschreibung der zentralen Merkmale einer ausgewählten Ressource. Die Metadaten können im XML-Format exportiert werden und ein Harvesting nach OAI-PMH ist ebenso möglich<sup>179</sup>.

### Standards und Schnittstellen

Als Format zur Textcodierung wird Wikitext eingesetzt. Eine Exportfunktion für Metadaten ist von dem verwendeten CMS abhängig. Im Fall von Omeka werden die Metadaten im Dublin Core Standard<sup>180</sup> abgelegt.

### Systemvoraussetzungen

Die Installation von Scripto im Rahmen eines bestehenden CMS erfordert eine für den Einsatz mit dem Plugin konfigurierte Instanz von MediaWiki. Bei der Integration in WordPress und Drupal muss darüber hinaus die Software Zend Framework<sup>181</sup> vorhanden sein<sup>182</sup>.

---

<sup>178</sup>Vgl. Roy Rosenzweig Center for History and New Media. *Working with Dublin Core*. URL: [http://omeka.org/codex/Working\\_with\\_Dublin\\_Core](http://omeka.org/codex/Working_with_Dublin_Core) (besucht am 28.07.2012).

<sup>179</sup>Vgl. Roy Rosenzweig Center for History and New Media. *Omeka Feature List*. URL: <http://omeka.org/files/docs/Featurelist.pdf> (besucht am 28.07.2012).

<sup>180</sup>Dublin Core Metadata Initiative. *Dublin Core*. URL: <http://dublincore.org/> (besucht am 21.07.2012), Vgl.

<sup>181</sup>Vgl. Zend Technologies. *Zend Framework*. URL: <http://framework.zend.com/> (besucht am 21.07.2012).

<sup>182</sup>Vgl. Center for History and New Media, *About Scripto*.

## Anwendungsbeispiele

Scripto wird in seiner Alpha-Version von dem Projekt „Papers of the War Department“<sup>183</sup> eingesetzt (s. o.). Es beinhaltet die Online-Publikation von 55.000 Dokumenten des Kriegsministeriums der Vereinigten Staaten aus den Jahren 1784 bis 1800. Die Manuskripte des Kriegsministeriums wurden am 8. November 1800 bei einem Brand beschädigt bzw. vernichtet. Nach einer Wiederherstellung der Bestände durch Konsultation verschiedener Archive in den Vereinigten Staaten, England, Frankreich und Schottland sollen sie nun für Wissenschaftler, Studenten und die interessierte Öffentlichkeit online zugänglich gemacht werden. Dabei werden die Scans mit Metadaten verlinkt und transkribiert<sup>184</sup>. Das Projekt wird vom Roy Rosenzweig Center for History and New Media<sup>185</sup> durchgeführt und erhält Zuschüsse von der National Historical Publications and Records Commission sowie dem National Endowment for the Humanities.

## Evaluation

Der bei Scripto verfolgte Ansatz einer Integration des Transkriptionswerkzeugs in ein bestehendes CMS ist in besonderem Maße zur Transkription von Objekten aus digitalen Sammlungen geeignet, bei denen der Schwerpunkt auf dem Inhalt der Transkripte liegt und die Textcodierung eine untergeordnete Rolle spielt. Die überschaubare Benutzerschnittstelle sorgt für kurze Einarbeitungszeiten und eignet sich damit für Crowdsourcing-Projekte mit einem heterogenen Kreis von Freiwilligen.

Die über den reinen Transkriptionsmodus hinaus bereitgestellten Funktionen wie Versionskontrolle, Diskussionsmöglichkeiten und Freischaltung der Transkriptionen (Import in das CMS) werden in ähnlicher Weise auch von den meisten anderen betrachteten Programmen realisiert.

Die Verwendung von Scripto in Kombination mit Omeka ist besonders für Anwendungsfälle mit einem bildzentrierten Ansatz empfehlenswert. Dabei kommen die Programmeigenschaften zur Verlinkung und Beschreibung einzelner Elemente (items) mit Metadaten zum Tragen. Ein Element kann auch mehrere Dateien enthalten, die sich vom Nutzer durchblättern lassen. Davon abgesehen bilden die Elemente allerdings isolierte Einheiten, d. h. nicht ein zusammenhängender Text, sondern die einzelnen Elemente stellen das Ordnungsprinzip in dem CMS dar. Für textzentrierte Ansätze ergeben sich

---

<sup>183</sup>Vgl. Roy Rosenzweig Center for History and New Media, *Papers of the War Department (1784-1800)*.

<sup>184</sup>Vgl. ebd.

<sup>185</sup>Vgl. Roy Rosenzweig Center for History and New Media, *Roy Rosenzweig Center for History and New Media*.

daraus einige Schwierigkeiten. So lassen sich die Scans nicht Seite an Seite mit den Transkriptionen darstellen und die Transkriptionen lassen sich bei einer Einbindung in die sogenannten exhibits (Internetseiten zur Darstellung einer Kombination verschiedener items zusammen mit einem narrativen Text) nicht durchsuchen<sup>186</sup>. Indessen werden auch für textzentrierte Ansätze Erweiterungen für das CMS entwickelt. So stellen das TEI Boilerplate project<sup>187</sup> und das TEI Display Plugin<sup>188</sup> Versuche dar Textdaten im TEI-Format mit Hilfe von Omeka zu publizieren<sup>189</sup>. Ein mögliches Szenario zur Anwendung dieser Erweiterungen wäre ein Workflow mit der Transkription von Manuskripten mit Hilfe von Scripto als erstem Arbeitsschritt und der darauf folgenden Erstellung eines Editionstextes im TEI-Format.

Für weitergehende Anforderungen wie die Erstellung von Indizes oder Textkommentaren ist das Programm ohne weitere Anpassungen nicht geeignet. Eine Erweiterung der Funktionalität in dieser Hinsicht ergibt allerdings lediglich in Anwendungsszenarien Sinn, die auf grundsätzliche Programmeigenschaften wie die CMS-Integration nicht verzichten können und mit dem Ansatz einer flachen Textauszeichnung ihre Ziele erreichen wollen. Ein Ansatz zur Generierung von Indizes wäre die Verwendung der Dublin Core Metadaten zur Speicherung von Personen- und Ortsnamen in Verbindung mit einigen Programmierarbeiten zur Erstellung der Indizes. Damit wäre allerdings nicht die Referenzierung der exakten Textstellen, in denen die entsprechenden Begriffe auftauchen, möglich, sondern ausschließlich die Aufnahme der Begriffe in die Metadaten zu dem digitalen Objekt als Ganzem. Ein weiterer sinnvoller Ansatz zur Realisierung von Zeitleisten und Georeferenzierung der digitalen Objekte ist mit dem Plugin Neatline<sup>190</sup> gegeben<sup>191</sup>.

Über die vorhandenen Möglichkeiten zur Anzeige von Bearbeiterstatistiken hinaus wäre eine Funktion zur Visualisierung des Transkriptionsfortschritts etwa durch einen Fortschrittsbalken eine sinnvolle Erweiterung.

---

<sup>186</sup>Vgl. Arno Bosse. *Creating a resource center for texts*. Forumsbeitrag. URL: <http://omeka.org/forums/topic/creating-a-ressource-center-for-texts> (besucht am 28.07.2012).

<sup>187</sup>Vgl. John Walsh, Grant Simpson und Saeed Moaddelli. *TEI Boilerplate*. URL: <http://dcl.slis.indiana.edu/teibp/> (besucht am 28.07.2012).

<sup>188</sup>Vgl. Roy Rosenzweig Center for History and New Media. *Plugins/TeiDisplay*. URL: <http://omeka.org/codex/Plugins/TeiDisplay> (besucht am 28.07.2012).

<sup>189</sup>Vgl. Bosse, *Creating a resource center for texts*.

<sup>190</sup>Vgl. Scholars' Lab. *Neatline*. URL: <http://neatline.org/> (besucht am 27.07.2012).

<sup>191</sup>Vgl. Omeka Development Team. *Using Omeka for a digital editon*. Forumsbeitrag. URL: <http://omeka.org/forums/topic/using-omeka-for-a-digital-edition> (besucht am 27.07.2012).



### 6.1.5 T-PEN

T-PEN (Transcription for Paleographical and Editorial Notation) ist ein Online-Werkzeug für Bearbeiter mit unterschiedlichen technischen Vorkenntnissen. Es unterstützt Optionen zur Codierung mit einem XML-TEI-Assistenten und zur reinen Transkription ohne Textauszeichnung. Transkriptionen können erstellt und bearbeitet sowie auf verschiedene Arten angezeigt werden. Sie können als PDF, RTF und XML exportiert oder direkt an eine teilnehmende Institution zur weiteren Aufbereitung weitergeleitet werden. T-PEN kann automatisch Spalten und Zeilen im Layout des Scans analysieren und macht auf diese Weise eine Verknüpfung der Zeilen im transkribierten Text mit den entsprechenden Zeilen im Scan möglich<sup>192</sup>. Der Transkriptionsmodus ist nicht mit einer Präsentationsebene für die Texte verknüpft, d. h. es handelt sich bei der Funktionalität um die Erstellung von Transkriptionen, die nach dem Export in einem weiteren Schritt mit anderen Tools für die Präsentation im Internet oder für andere Zwecke aufbereitet werden sollen.

Ferner können eigene Projekte angelegt und Scans hochgeladen werden. Die Projektmanagementfunktionen umfassen Einstellungen für öffentliche Projekte (Crowdsourcing) sowie für einen begrenzten Bearbeiterkreis. In den Transkriptionsmodus integrierte Hilfsmittel wie Wörterbücher, ein elektronischer Text der lateinischen Bibel (Vulgata), Verzeichnisse mittelalterlicher Abkürzungen usw. können projektspezifisch durch weitere Hilfsmittel ergänzt werden. Ebenso kann für ein Projekt ein eigenes XML-Schema importiert und die Transkriptionen dagegen validiert werden<sup>193</sup>.

#### Projekthintergrund

Das Projekt T-PEN wird vom Center for Digital Theology<sup>194</sup> der Saint Louis University koordiniert und hat Zuschüsse von der Andrew W. Mellon Foundation<sup>195</sup> und dem National Endowment for the Humanities<sup>196</sup> erhalten<sup>197</sup>.

---

<sup>192</sup>Vgl. Center for Digital Theology, *T-PEN*.

<sup>193</sup>Vgl. Center for Digital Theology. *T-PEN*. Blog. URL: <http://digital-editor.blogspot.de/> (besucht am 16.07.2012).

<sup>194</sup>Vgl. Center for Digital Theology. *Center for Digital Theology*. URL: <http://www.slu.edu/x27122.xml> (besucht am 15.07.2012).

<sup>195</sup>Vgl. Andrew W. Mellon Foundation. *Andrew W. Mellon Foundation*. URL: <http://www.mellon.org/> (besucht am 15.07.2012).

<sup>196</sup>Vgl. National Endowment for the Humanities. *National Endowment for the Humanities*. URL: <http://www.neh.gov/> (besucht am 15.07.2012).

<sup>197</sup>Vgl. Center for Digital Theology. *About T-PEN*. URL: <http://t-pen.org/TPEN/about.jsp> (besucht am 16.07.2012).

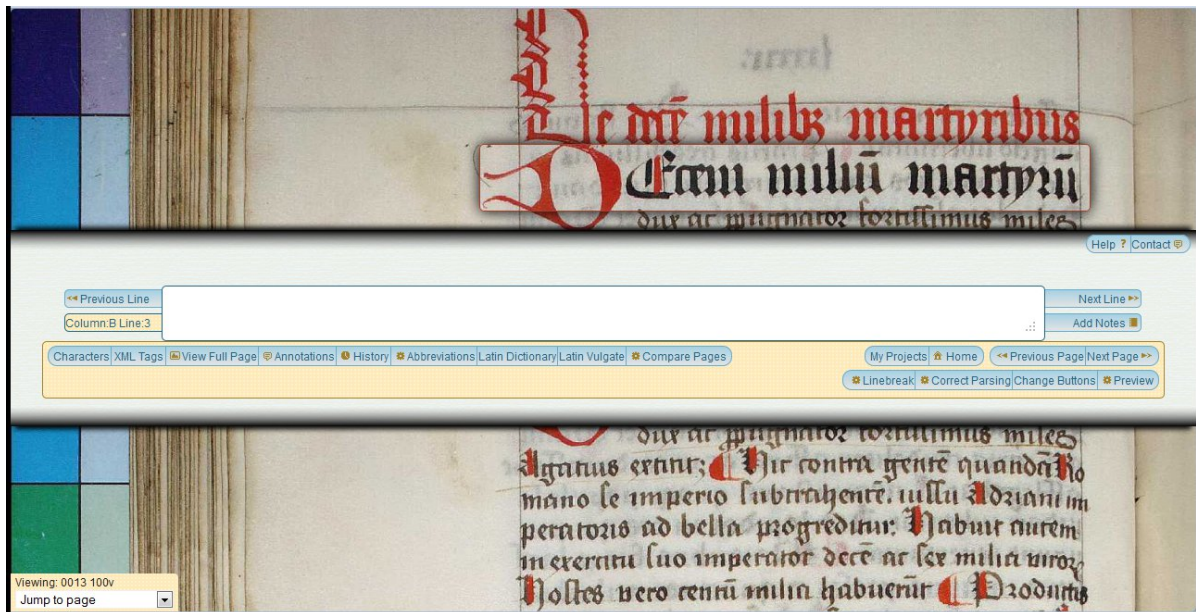


Abbildung 6.10: T-PEN: Transkriptionsmodus

## Funktionen

Die Darstellung im Transkriptionsmodus ist zeilenbasiert, d. h. die aktuell zu bearbeitende Zeile wird im Scan hervorgehoben und die Textbox direkt darunter angezeigt, sodass beide immer im Blick des Bearbeiters sind und kein Scrollen notwendig ist (siehe Abbildung 6.10 Seite 70). Dieser Modus wird durch die automatische Zeilen- und Spaltenerkennung der Software ermöglicht, welche bei Schwierigkeiten mit der automatischen Erkennung auch vom Benutzer angepasst werden kann. Neben der Transkriptionsbox wird das Bild der Handschrift und eine Navigationsschaltfläche zur Wahl der Seiten des Manuskripts dargestellt.

Die Werkzeugleiste kann projektspezifisch durch Angabe der URL des entsprechenden Hilfsmittels angepasst werden. T-PEN wurde ursprünglich zur Transkription mittelalterlicher Handschriften entwickelt und die standardmäßige Werkzeugleiste spiegelt diese Ausrichtung wider (siehe Abbildung 6.11 Seite 71). Dort steht die Eingabe von projektspezifisch anpassbaren Sonderzeichen, von XML-Tags entsprechend dem importierten Schema, eine Option zur Anzeige der kompletten Manuskriptseite, eine Funktion zur Annotierung des Scans, ein Zugriff auf die Versionskontrolle, ein Verzeichnis mittelalterlicher Abkürzungen, ein lateinisches Wörterbuch sowie eine Funktion zum Vergleich von



Abbildung 6.11: T-PEN: Werkzeugleiste

Glyphen auf verschiedenen Seiten des Manuskripts (Ziel: Analyse von Veränderungen der Handschriften; derzeit eine experimentelle Funktion<sup>198</sup>) zur Auswahl.

Die Funktionen zum Projektmanagement (siehe Abbildung 6.12 Seite 72) beinhalten im Rahmen der Projektseite (Reiter „Projects“) Funktionen zur Erstellung neuer Projekte, zum Hochladen neuer Scans, zur Aufnahme neuer Manuskripte in bestehende Projekte, zur Anpassung der Metadaten des entsprechenden Projekts, zur Veränderung der Reihenfolge der Manuskripte sowie zur Löschung unerwünschter Scans. Die Manuskriptverwaltung (Reiter „Manuscripts“) ermöglicht das Öffnen eines Manuskripts im Transkriptionsmodus, das Hochladen von Transkriptionen im Klartextformat und die Korrektur der automatischen Zeilenerkennung. Das Kollaborations-Management (Reiter „Collaboration“) umfasst Einstellungen zum Projektteam mit Möglichkeiten zum Hinzufügen und Löschen von Mitgliedern, eine Funktion zum Einladen neuer Team-Mitglieder und ein Protokoll zur Dokumentation der Aktivitäten in dem jeweiligen Projekt. In den Optionen (Reiter „Options“) können XML-Schemata eingebunden und die Transkriptionen validiert werden. Zudem wird die Möglichkeit angeboten, die Schaltflächen für Sonderzeichen und XML-Tags projektspezifisch anzupassen sowie Transkriptionshilfsmittel für das Projekt auszuwählen oder durch Angabe der entsprechenden URL neu hinzuzufügen. In der Kategorie Export (Reiter „Export“) können Transkriptionen in PDF, RTF, XML exportiert werden. Eine Vorschauanzeige des Dokuments ist ebenso möglich wie Einstellungen zur Anpassung des Layout. Des Weiteren können Projekte vollständig öffentlich geführt werden, d. h. die Projekteinstellungen und die Transkriptionen können von jedermann bearbeitet werden.

## Standards und Schnittstellen

Es existiert eine Importfunktion für XML-Schemata und eine Exportmöglichkeit in den Formaten RTF, PDF und XML. Ferner können abgeschlossene Transkriptionen direkt an teilnehmende Institutionen weitergeleitet werden.

<sup>198</sup>Vgl. Center for Digital Theology. *T-PEN Version 2.0*. Blog-Eintrag. 5. Mai 2012. URL: <http://digital-editor.blogspot.de/2012/05/t-pen-version-20.html> (besucht am 16.07.2012).

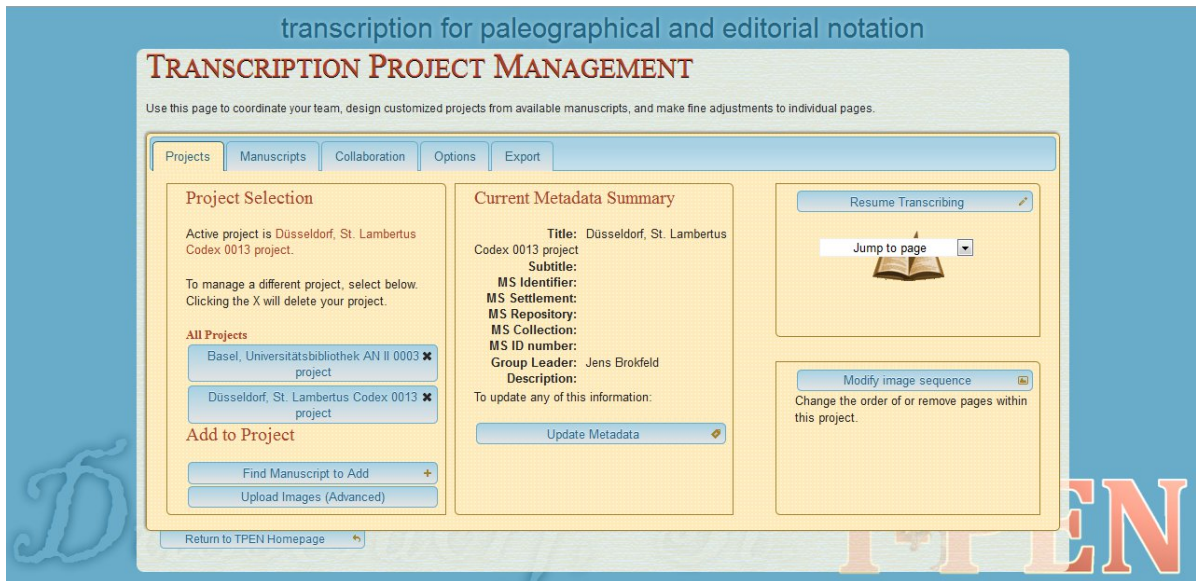


Abbildung 6.12: T-PEN: Projektmanagement

### Systemvoraussetzungen

Das Programm ist derzeit lediglich in der gehosteten Version verfügbar. Es soll aber als freie Software unter der Lizenz ECL v.2.0 veröffentlicht werden.

### Anwendungsbeispiele

Das Projekt Carolingian Canon Law setzt T-PEN zur Transkription der Hauptwerke der karolingischen kanonischen Rechtsprechung ein. Dabei werden Verbindungen zu anderen Texten kanonischen Rechts aufgezeigt, die von den karolingischen Lesern benutzt wurden. Das Projekt zeichnet das Ausmaß der Abweichungen von den geläufigen Gesetzestexten auf und nimmt historische sowie bibliographische Annotationen in den Editionstext auf<sup>199</sup>.

### Evaluation

Auf Grund seiner hohen Anpassungsfähigkeit in Bezug auf das verwendete TEI-Schema und die eingesetzten Hilfsmittel ist T-PEN besonders für wissenschaftlich orientierte Projekte mit einem Bedarf nach detaillierter Auszeichnung von Textmerkmalen prädes-

<sup>199</sup>Vgl. University of Kentucky. *Carolingian Canon Law*. URL: <http://ccl1.rch.uky.edu/> (besucht am 28.07.2012).

tiert. Auch für Transkriptionsprojekte in Gedächtnisorganisationen ist der Einsatz des Programms bei Vorliegen ähnlicher Rahmenbedingungen bedenkenswert.

Für Crowdsourcing-Projekte in Archiven, Bibliotheken und Museen fehlt allerdings die wichtige Komponente einer motivationssteigernden Wiedergabe von Bearbeiterstatistiken etwa im Rahmen von Ranglisten oder Fortschrittsbalken. Zudem ist T-PEN ein reines Transkriptionswerkzeug, das die Aufbereitung der Textdaten für eine Online-Präsentation nicht beinhaltet. Die XML-Dateien müssten zu diesem Zweck erst in ein CMS exportiert oder mit anderen Tools für eine Publikation im Internet aufbereitet werden.

Mit der Wahl von TEI als Codierungsformat liegen zahlreiche Möglichkeiten zur semantischen Auszeichnung der Texte vor. Auf diese kann später für unterschiedliche Anwendungsfälle zurückgegriffen werden, so etwa zur Generierung von Indizes aus entsprechend ausgezeichneten Orts- und Personennamen oder der Weitergabe von den enthaltenen Metadaten zur Beteiligung an Portalen.

Der Transkriptionsmodus bietet vielfältige Möglichkeiten zur Bearbeitung und Darstellung der Transkripte. Die ursprüngliche Konzeption des Werkzeugs für mittelalterliche Texte stellt auf Grund seiner hohen Anpassungsfähigkeit für die Arbeit mit Manuskripten aus anderen Epochen keine maßgebliche Hürde dar. So können Schaltflächen zur Eingabe von Sonderzeichen projektspezifisch angepasst und andere Transkriptionshilfsmittel durch Angabe der entsprechenden URL eingebunden werden.

Die zeilenbasierte Transkriptionsansicht ermöglicht ein komfortables Arbeiten, das im Vergleich zu den anderen vorgestellten Ansätzen eine Besonderheit darstellt. Der Bearbeiter kann damit sowohl die Vorlage als auch das Transkript im Auge behalten und sich Zeile für Zeile durch das Manuskript arbeiten. Die optionale Anzeige der gesamten Manuskriptseite parallel zur zeilenbasierten Ansicht ermöglicht dabei einen Blick auf den weiteren Kontext der zu bearbeitenden Zeile.

Es existiert eine Funktion zur Erstellung vollständig öffentlicher Projekte, in denen sowohl Projekteigenschaften und Transkriptionen von jederman bearbeitet werden können. Damit werden Crowdsourcing-Projekte mit einem offenen Bearbeiterkreis unterstützt. Die Zugriffsrechte können in Bezug auf Lesezugriff, Projektmanagement-Einstellungen, Zugang zum Transkriptionsmodus sowie Genehmigung zum Export der Texte festgelegt werden. Auf diese Weise ist ein große Bandbreite von Konfigurationen möglich, welche auch die Realisierung privater Projekte ohne Lesezugriff für Externe erlaubt.

Allein die Öffnung von Projekten für einen heterogenen Bearbeiterkreis wird allerdings in den meisten Fällen nicht zur Beteiligung vieler Freiwilliger an einem Projekt führen.

Dafür ist neben der Öffentlichkeitsarbeit auch ein gewisser spielerischer Charakter der Website notwendig. Neben den genannten Bearbeiterstatistiken sind hier auch Themenseiten mit weiteren Projektinformationen in leicht verständlicher Form und Diskussionsforen zu nennen. Hierfür bietet T-PEN in der gehosteten Version keine Gelegenheit. Letzteres rührt von der Konzeption des Programms für Bearbeiter aus dem Bereich der Digital Humanities her und ist in diesem Sinne auch folgerichtig. Ein wissenschaftlich orientiertes Projekt wird keinen großen Wert auf den spielerischen Umgang mit den angebotenen Werkzeugen legen, sondern die Programmeigenschaften schwerpunktmäßig an der exakten Abbildung der Textmerkmale der behandelten Manuskripte ausrichten. Im Rahmen der Einbeziehung einer breiteren Öffentlichkeit auch in wissenschaftliche Projekte wird dieser Ansatz aber nur in engen Grenzen zu einer Beteiligung größerer Kreise führen. Die Installation von T-PEN auf einem eigenen Server könnte in dieser Hinsicht größere Freiheiten zur Einbeziehung öffentlichkeitswirksamer Funktionen erlauben.

### 6.1.6 Bentham Transcription Desk

Das Transkriptionswerkzeug „Bentham Transcription Desk“ stellt eine projektspezifisch angepasste Version von MediaWiki dar. Im Transkriptionsmodus wird eine Klartextbox zur Eingabe des Transkriptionstextes und grundlegender Textauszeichnungselemente in TEI verwendet. Die Codierung wird mit Hilfe einer Werkzeugleiste in den Text eingefügt. Bei Abschluss der Transkription wird der Text in einer mit Hilfe von CSS gerenderten Version neben dem Scan zur Überprüfung des Layout angezeigt<sup>200</sup>.

#### Projekthintergrund

Das Projekt „Transcribe Bentham“<sup>201</sup> verfolgt das Ziel einer Transkription der unveröffentlichten Werke des Philosophen und Reformers Jeremy Bentham (1748–1832). Die entstandenen Transkriptionen fließen in die editorische Arbeit zur Publikation der gedruckten Fassung von Benthams Gesammelten Werken ein und werden im digitalen Bentham Repository<sup>202</sup> des University College London zugänglich gemacht. Das Projekt wurde durch einen einjährigen Zuschuss vom Arts and Humanities Research Coun-

---

<sup>200</sup>Vgl. Causer, Tonra und Wallace, „Transcription maximized; expense minimized? crowdsourcing and editing The Collected Works of Jeremy Bentham“, S. 121-123.

<sup>201</sup>Vgl. University College London, *Transcribe Bentham*.

<sup>202</sup>Vgl. University College London. *Digital Bentham Repository*. URL: [http://digitool-b.lib.ucl.ac.uk:8881/R/?func=collections&collection\\_id=1867](http://digitool-b.lib.ucl.ac.uk:8881/R/?func=collections&collection_id=1867) (besucht am 22.07.2012).

cil<sup>203</sup> ermöglicht. Die Software „Bentham Transcription Desk“ wurde vom University of London Computer Center<sup>204</sup> entwickelt<sup>205</sup>.

## Funktionen

Neben dem Transkriptionsmodus verfügt das Programm über Funktionen zur Darstellung der Transkriptionsrichtlinien und Projektdokumentation sowie zur Diskussion zwischen Benutzern und zum Austausch mit dem Moderatorenteam sowie zur Integration von Social Media<sup>206</sup>.

Im Transkriptionsmodus (siehe Abbildung 6.13 Seite 76) wird der Text in eine Klartextbox eingegeben und eine Werkzeugleiste ermöglicht das Einfügen von Elementen zur Textauszeichnung mit TEI-Tags. Diese beinhalten Strukturinformationen wie Zeilenumbrüche, Seitenumbrüche, Überschriften und Absätze. Zudem werden linguistische Merkmale wie Anmerkungen, ungewöhnliche Schreibweisen und fremdsprachige Textpassagen codiert. Gleiches werden Besonderheiten der Textkomposition wie Hinzufügungen und Durchstreichungen sowie interpretatorische Entscheidungen über fragliche Lesarten und schwer lesbare Wörter vermerkt. Die Auswahl der Tags orientiert sich dabei an den am häufigsten in Benthams Manuskripten vorkommenden Charakteristika und Überlegungen über die Wichtigkeit dieser Merkmale in Bezug auf die Codierung. Die verwendete Syntax zur Textauszeichnung stellt eine vereinfachte Schreibweise dar, bei der die Repräsentation von Attributen und Attributwerten nach Möglichkeit vermieden wird<sup>207</sup>.

Die vorhandenen Manuskripte werden nach Thema, Datum und Schwierigkeitsgrad klassifiziert. Für das Transkriptionsprojekt wird der Schwerpunkt allerdings auf die frühen, „einfachen“ Handschriften gelegt<sup>208</sup>. Zudem existiert eine Liste mit noch zu transkribierenden Handschriften. Zur Fortführung begonnener Transkriptionen können sich Bearbeiter gezielt ihre eigenen Beiträge anzeigen lassen. Eine weitere Option ist die

<sup>203</sup>Vgl. Arts and Humanities Research Council. *Arts and Humanities Research Council*. URL: <http://www.ahrc.ac.uk/Pages/default.aspx> (besucht am 22.07.2012).

<sup>204</sup>Vgl. University of London Computer Center. *University of London Computer Center*. URL: <http://www.ulcc.ac.uk/> (besucht am 22.07.2012).

<sup>205</sup>Vgl. Causer, Tonra und Wallace, „Transcription maximized; expense minimized? crowdsourcing and editing The Collected Works of Jeremy Bentham“, S. 120-122.

<sup>206</sup>Vgl. ebd., S. 122.

<sup>207</sup>Vgl. ebd., S. 123.

<sup>208</sup>Vgl. Martin Moyle, Justin Tonra und Valerie Wallace. „Manuscript Transcription by Crowdsourcing: Transcribe Bentham“. In: *Liber Quarterly* 20.3/4 (2011), S. 347-256. URL: <http://liber.library.uu.nl/index.php/lq/article/view/7999/8329> (besucht am 24.07.2012), S. 351.



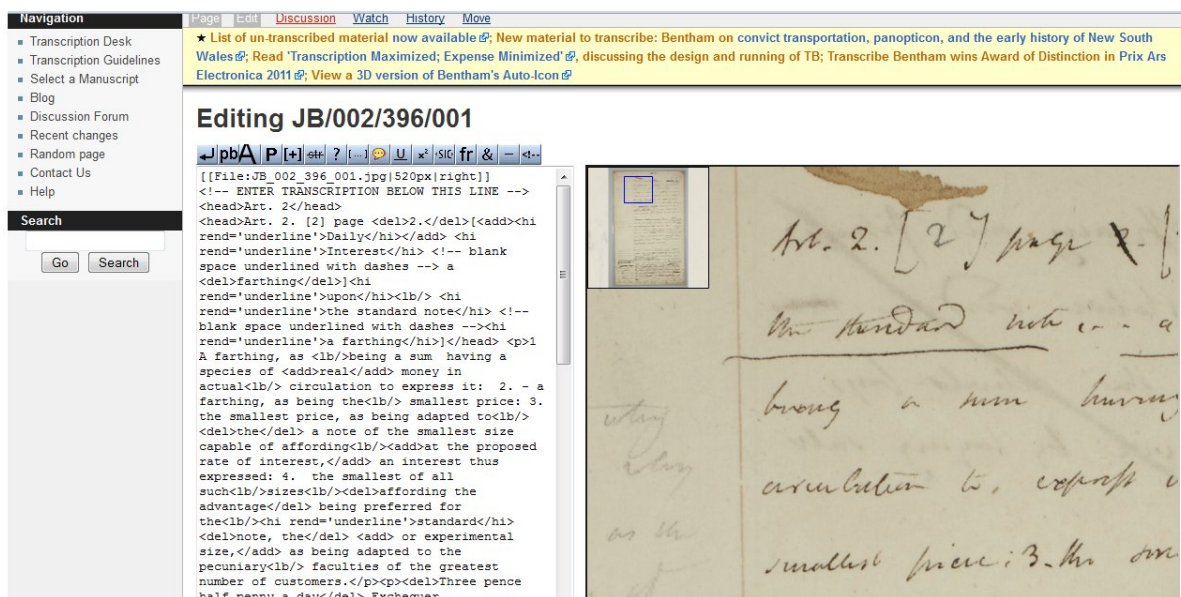


Abbildung 6.13: Bentham Transcription Desk: Transkriptionsmodus

Merkliste (watchlist), auf die der Bearbeiter Seiten aufnehmen kann deren Bearbeitungsfortschritt er mitverfolgen möchte.

Im digitalen Bentham Repositorium<sup>209</sup> werden die Scans als Bilddateien sowie die fertigen Transkriptionen in einer mit CSS formatierten Ansicht bereitgestellt (siehe Abbildung 6.14 Seite 77). In Zukunft soll im Interesse einer verbesserten Suchfunktion bei der Recherche auf die TEI-Codierung zurückgegriffen werden können. Damit wäre etwa die Suche nach allen Beispielen für die Durchstreichung eines bestimmten Begriffes oder die gezielte Suche in den Marginalien möglich<sup>210</sup>.

Die vor einigen Jahren zusammengetragenen Metadaten zu den einzelnen Manuskriptseiten werden mit den dazugehörigen Scans in den Transcription Desk und das digitale Repositorium hochgeladen<sup>211</sup>. Sie können von Wiki-Administratoren und Systementwicklern angepasst werden<sup>212</sup>. Das Codebeispiel (siehe Beispiel 6.3 Seite 77) zeigt die Datenfelder eines zufälligen Metadatensatzes. Daraus wird die Strukturierung des Bestandes nach Kästen (box\_number) und Manuskriptseiten (folio\_number) ersichtlich. Zudem wird das Blatt u. a. einer Kategorie zugeordnet und den Metadaten werden im

<sup>209</sup>Vgl. University College London. *Digital Bentham Repository*. URL: [http://digitool-b.lib.ucl.ac.uk:8881/R&?local\\_base=BENTHAM](http://digitool-b.lib.ucl.ac.uk:8881/R&?local_base=BENTHAM) (besucht am 01.08.2012).

<sup>210</sup>Vgl. Causer, Tonra und Wallace, „Transcription maximized; expense minimized? crowdsourcing and editing The Collected Works of Jeremy Bentham“, S. 125.

<sup>211</sup>Vgl. Tim Causer. *RE: Questions on Bentham Transcription Desk*. E-Mail. 31. Juli 2012.

<sup>212</sup>Vgl. University College London. *Help: Wiki Overview*. URL: [http://www.transcribe-bentham.dalcc.ac.uk/td/Help:Wiki\\_Overview](http://www.transcribe-bentham.dalcc.ac.uk/td/Help:Wiki_Overview) (besucht am 28.07.2012).



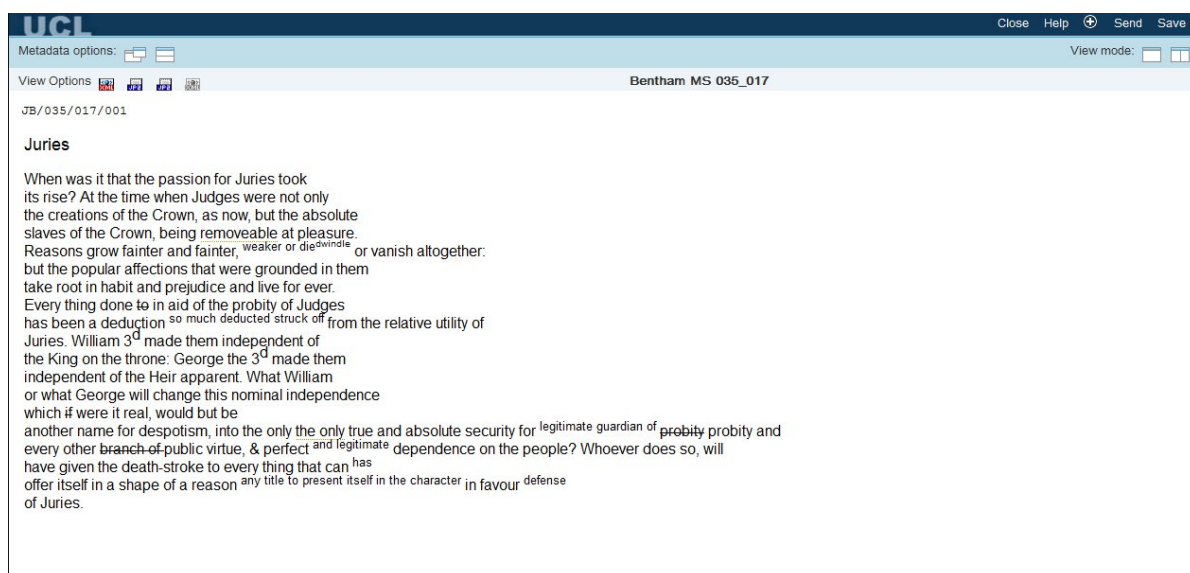


Abbildung 6.14: Bentham Transcription Desk: Digitales Bentham Repositorium

Text vorkommende Datumsangaben sowie Überschriften hinzugefügt. Zur Identifizierung benötigte Daten werden ebenfalls abgebildet (`id_number`, `image_number`, `identifizier`).

Die Transkriptionen im Projekt „Transcribe Bentham“ durchlaufen eine Qualitätskontrolle durch das Moderatorenteam. Nachdem ein Transkript durch den Bearbeiter zur Kontrolle freigegeben wird, überprüfen es die Moderatoren auf textuelle Genauigkeit und Konsistenz der Codierung. Anschließend entscheiden die Moderatoren, ob der Text gegen Veränderungen gesichert wird oder weiterhin für Bearbeitungen zugänglich bleiben soll<sup>213</sup>. Das von Causer et al. publizierte Prozessdiagramm stellt diese Aktivitäten in ihrem organisatorischen Zusammenhang dar<sup>214</sup> (siehe Abbildung 6.15 Seite 79). Das Layout ist vom Autor dieser Arbeit selbst in Anlehnung an die in der Zeitschrift „Literary and Linguistics Computing“ veröffentlichte Fassung erstellt worden. Die abgeschlossenen Transkriptionen werden mit dem oXygen-XML-Editor in XML-Dateien konvertiert und anschließend im digitalen Bentham Repositorium bereitgestellt<sup>215</sup>. Über diesen Workflow hinaus besteht für Benutzer die Möglichkeit die Transkriptionen anderer Freiwilliger zu kontrollieren und zu ergänzen<sup>216</sup>.

<sup>213</sup>Vgl. Causer, Tonra und Wallace, „Transcription maximized; expense minimized? crowdsourcing and editing The Collected Works of Jeremy Bentham“, S. 128.

<sup>214</sup>Vgl. ebd., S. 128.

<sup>215</sup>Vgl. ebd., S. 125.

<sup>216</sup>Vgl. ebd., S. 122.

## Codebeispiel 6.3: Bentham Transcription Desk: Metadaten

```
1 {{Infobox Folio New
2 | box_number = 035
3 | folio_number = 202
4 | second_folio_number =
5 | category = text sheet
6 | rectoverso = recto
7 | creator =
8 | recipient =
9 | penner = jeremy bentham
10 | marginals =
11 | corrections =
12 | date_1 = 1808-05-30
13 <!--...-->
14 | estimated_date =
15 | info_in_main_headings_field =
16 | main_headings = constitutional code; evidence; procedure code
17 | sub_headings = jury analysed / disadvantages insepar. / misprision
18 | marginal_summary_numbering = not numbered
19 | number_of_pages = 1
20 | page_numbering = e3
21 | titles =
22 | watermarks =
23 | paper_producer =
24 | paper_produced_in_year =
25 | notes_public =
26 | id_number = 10795
27 | image_number = 001
28 | identifier = JB/035/202/001
29 | next = JB/035/203/001
30 | prev = JB/035/201/001
31 }}
```

Informationsseiten zum Thema Paläographie und Video-Tutorien zum Umgang mit dem Transkriptionswerkzeug ergänzen den Internetauftritt. Neben einem Diskussionsforum können registrierte Nutzer auch persönliche Profile anderer Freiwilliger einsehen. Ein

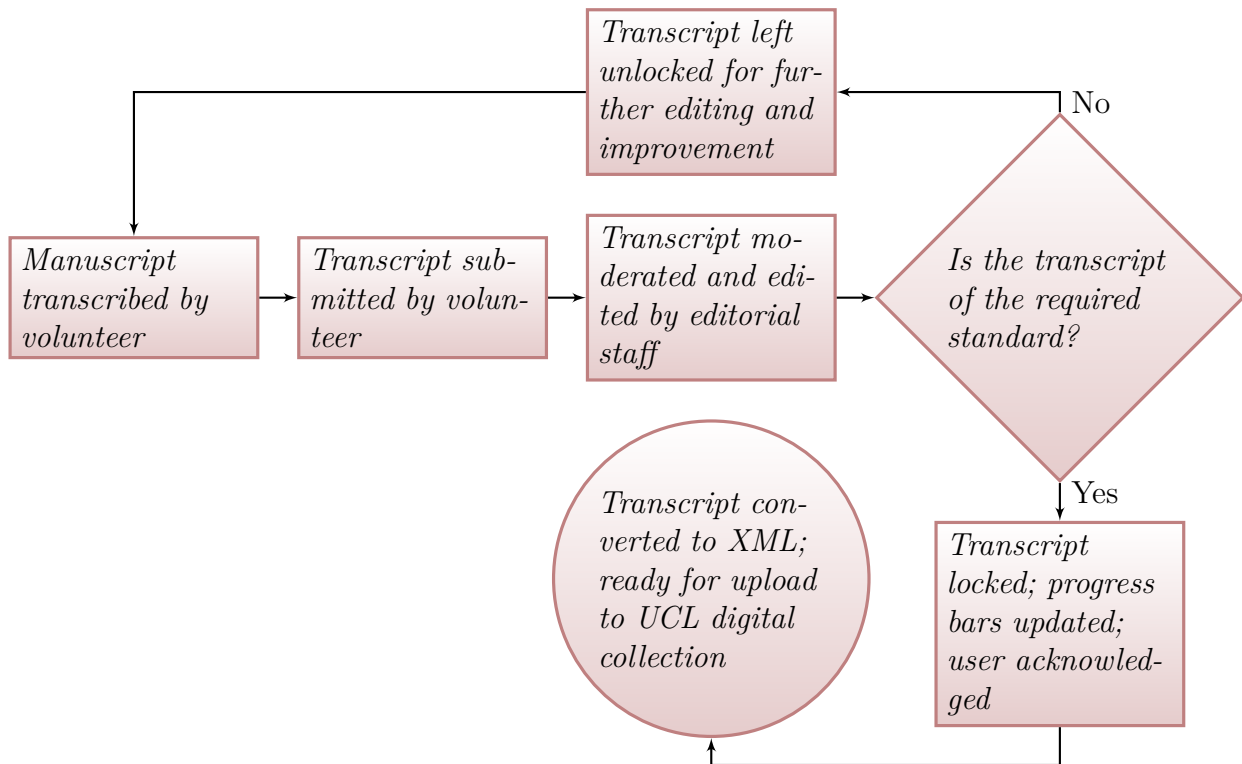


Abbildung 6.15: Bentham Transcription Desk: Moderations-Workflow

Punktesystem und eine Rangliste der aktivsten Bearbeiter dienen zur weiteren Motivation (siehe Abbildung 6.16 Seite 80). Die Fortschritte des Transkriptionsprojekts werden mit Hilfe eines „Benthamometer“ visualisiert<sup>217</sup>. Zudem informiert ein Blog regelmäßig über die Fortschritte des Projekts.

### Standards und Schnittstellen

Zur Textcodierung wird eine Auswahl von TEI-Tags mit vereinfachter Syntax eingesetzt. Das dem Transcription Desk zu Grunde liegende MediaWiki besitzt eine Schnittstelle zum Datenimport. Diese wird bei Transcribe Bentham zum Hochladen der Metadaten und Scans genutzt.

### Systemvoraussetzungen

Zuerst wird eine Installation von MediaWiki benötigt. Die verwendeten Erweiterungen sind JBTEIToolbar, TEITags (Unterstützung für TEI markup) und JBZV (Unterstüt-

<sup>217</sup>Vgl. Moyle, Tonra und Wallace, „Manuscript Transcription by Crowdsourcing: Transcribe Bentham“, S. 353.










| Transcribe Bentham Top Contributors   |   |
|---|---|
|    | Diane Folan (Prodigy) 96,300 points       |
| benprowiki_1050_m.jpg   | Lea Stern (Prodigy) 79,325 points         |
|    | JFoxe (Master) 63,725 points              |
| benprowiki_1441_m.jpg   | Petergh (Acolyte) 22,275 points           |
|    | Jillybean (Acolyte) 20,475 points         |
| benprowiki_951_m.jpg  | RexL (Scribe) 13,550 points               |
|    | Mfoutz (Scribe) 13,100 points             |
| benprowiki_1301_m.jpg   | Jancoopes (Scribe) 10,375 points          |
| benprowiki_1364_m.jpg   | Ohsoldgirl (Novice) 8450 points           |
|    | Clarabloomer (Apprentice) 8225 points     |
|   | Calico-pie 6400 (Apprentice) points       |
| benprowiki_766_m.jpg  | Carno (Apprentice) 6350 points            |
| benprowiki_1384_m.jpg   | Charlotte bronte (Apprentice) 6275 points |
| benprowiki_655_m.jpg  | Lidunn (Apprentice) 5275 points           |
|  | Duyfken (Apprentice) 5000 points          |
| benprowiki_631_m.jpg  | PeaceAndJoy (Novice) 4925 points          |
| benprowiki_1402_m.jpg   | OlgaNM (Novice) 4800 points               |
| benprowiki_1390_m.jpg   | Sh4mth (Novice) 3325 points               |
|  | MatthewA (Novice) 3275 points             |
|  | BCRideout (Novice) 3225 points            |
| Find out how to become a Master transcriber!  |   |
| Last updated on 20/7/2012   |   |

Abbildung 6.16: Bentham Transcription Desk: Rangliste der aktivsten Bearbeiter

zung für den zoombaren Viewer). Hinzu kommen noch andere Erweiterungen, die je nach projektspezifischer Anpassung mehr oder weniger optional sind: ProgressBar, VideoFlash, AWC Forum (optional), Notitle, RSSReader, DiscussionThreading, ReCaptcha, SemanticMediaWiki (optional), SocialProfile, GroupPermissionsManager, UserMerge und MassEditRegex<sup>218</sup>.

### Anwendungsbeispiele

Neben dem Projekt „Transcribe Bentham“ (siehe Projekthintergrund), für das die Software entwickelt wurde, setzt das Staatsarchiv von Victoria (Public Records Office Victoria) den Transcription Desk zur Transkription und Textauszeichnung ihrer Bestände ein<sup>219</sup>. Für diesen Anwendungsfall wurde die Werkzeugleiste zur Annotierung von Personennamen, Ortsnamen etc. angepasst<sup>220</sup>. Das Projekt befindet sich noch im Aufbau.

### Evaluation

Der Schwerpunkt des Programms liegt auf der Transkription sowie der Codierung und Verwaltung der Transkripte. Eine Verknüpfung von Transkriptionsmodus und Präsentationsebene ist nicht gegeben, d. h. der Workflow (s. o.) beinhaltet zwar die Publikation der Texte im digitalen Bentham Repository, aber dieses ist nicht Teil des Transcription Desk. Die Online-Veröffentlichung wird über den Zwischenschritt einer Konvertierung der Dateien in XML mit dem oXygen Editor und das anschließende Hochladen in das Repository realisiert. Der Einsatz des Programms für andere Transkriptionsprojekte zur Erstellung einer digitalen Edition ist deshalb vom Vorhandensein bzw. dem Aufbau einer vergleichbaren Publikationsplattform abhängig. Wenn die Transkripte lediglich die Grundlage für eine gedruckte Edition bilden sollen, entfällt der letzte Schritt des Workflow mit dem Hochladen in ein Repository. Stattdessen müssten die XML-Dateien durch eine XSLT-Transformation für den Satz der gedruckten Edition vorbereitet bzw. zu diesem Zweck in ein anderes Format wie .doc konvertiert werden<sup>221</sup>.

Der Transkriptionsmodus erleichtert die Arbeit mit TEI-Tags durch eine vereinfachte Syntax und das Einfügen der Elemente mit Hilfe einer Werkzeugleiste. Die Einführung

---

<sup>218</sup>Vgl. University College London. *Installationsanweisungen für „Bentham Transcription Desk“*. URL: <http://code.google.com/p/tb-transcription-desk/wiki/Instructions> (besucht am 28.07.2012).

<sup>219</sup>Vgl. Public Records Office Victoria. *PROV Transcription Pilot*. URL: <http://prov.versi.edu.au/> (besucht am 28.07.2012).

<sup>220</sup>Vgl. Causer, *RE: Questions on Bentham Transcription Desk*.

<sup>221</sup>Vgl. Causer, Tonra und Wallace, „Transcription maximized; expense minimized? crowdsourcing and editing The Collected Works of Jeremy Bentham“, S. 125.

einer optionalen Bearbeitung im WYSIWYG-Modus wäre eine sinnvolle Erweiterung im Sinne der Usability<sup>222</sup>. Die Entwicklung des WYSIWYG-Editors soll im Oktober 2012 beginnen und die Publikation des ebenfalls unter freier Lizenz stehenden Moduls soll Mitte 2013 erfolgen<sup>223</sup>. Die Auswahl einer begrenzten Anzahl möglicher Elemente zur Textauszeichnung ist für Crowdsourcing-Projekte ein zweckmäßiger Ansatz. Diese wird aber in Bezug auf die zu codierenden Textmerkmale für unterschiedliche Manuskriptarten differieren und müsste in den meisten Fällen projektspezifisch angepasst werden. Für Projekte mit stark verschachtelten Dokumentstrukturen könnte sich zudem eine Hilfsfunktion zur automatischen Validitätskontrolle der Textauszeichnung eignen. Dabei würde bei dem Versuch ein Element innerhalb eines übergeordneten anderen Elements einzufügen eine Fehlermeldung angezeigt, wenn das Element entsprechend dem verwendeten XML-Schema nicht in dieser Umgebung eingesetzt werden darf.

Im Projekt „Transcribe Bentham“ wird von vielen motivationssteigernden Maßnahmen wie der Wiedergabe von Bearbeiterstatistiken, Fortschrittsbalken, Benutzerprofilen, Diskussionsforen, einem Blog und einer Rangliste der aktivsten Bearbeiter auf Grundlage eines Punktesystems Gebrauch gemacht. Diese Inhalte lassen sich mit dem Programm einbinden und bieten damit eine gute Ausgangslage für Crowdsourcing-Projekte mit einem heterogenen Bearbeiterkreis.

---

<sup>222</sup>Vgl. Causer, Tonra und Wallace, „Transcription maximized; expense minimized? crowdsourcing and editing The Collected Works of Jeremy Bentham“, S. 131.

<sup>223</sup>Vgl. Causer, *RE: Questions on Bentham Transcription Desk*.

# 7 Programmvergleich

## 7.1 Programmvergleich anhand der Prüfkriterien

Neben den Gemeinsamkeiten in Bezug auf die Auswahlkriterien (siehe Abschnitt 4 Seite 24) haben alle Programme folgende Funktionen gemeinsam:

- Editor zur Texteingabe und zoombarer Viewer
- Versionskontrolle
- Möglichkeiten zur Textauszeichnung (bei starken Unterschieden in der Zahl der möglichen Elemente)

Die Tabellen zum Programmvergleich anhand der Prüfkriterien bieten einen Überblick zum Funktionsumfang sowie Vergleichsmöglichkeiten zwischen den Werkzeugen (siehe Tabelle 1 auf Seite 85 sowie Tabelle 2 auf Seite 87). Die Programmeigenschaften werden in kurzgefasster Form wiedergegeben.

| Auffindbarkeit |  |  | Langlebigkeit  |  | Intellektuelle Redlichkeit                                 |  |
|----------------|--|--|--|--|--|--|
| Werkzeugname   | Schnittstellen   | Normdaten  | Hosting  | Lizenz   | Codierungsformat   | Indexierungsfunktion   |
| FromThePage    | <ul style="list-style-type: none"> <li>Datenimport aus dem Internet Archive</li> </ul>   | <ul style="list-style-type: none"> <li>Nein (aber geplant)</li> </ul>  | <ul style="list-style-type: none"> <li>Ja</li> </ul>   | <ul style="list-style-type: none"> <li>AGPL 3.0</li> </ul> | <ul style="list-style-type: none"> <li>Wikitext</li> </ul> | <ul style="list-style-type: none"> <li>Ja</li> </ul>   |
| Refine!        | <ul style="list-style-type: none"> <li>Datenaustausch mit Kalliope</li> </ul>  | <ul style="list-style-type: none"> <li>PND</li> <li>Integration von normierten Daten aus Kalliope</li> </ul> | <ul style="list-style-type: none"> <li>prinzipiell Ja (durch 3pc als kommerzieller Dienstleister)</li> </ul> | <ul style="list-style-type: none"> <li>GPL</li> </ul>      | <ul style="list-style-type: none"> <li>HTML/TEI</li> </ul> | <ul style="list-style-type: none"> <li>Ja (Indexierung mit Hilfe der Werkzeugeleiste)</li> <li>Nein</li> </ul> |
| Wikisource     | <ul style="list-style-type: none"> <li>Datenimport und -Export im Rahmen der von MediaWiki bereitgestellten Funktionalität</li> <li>OAL-PMH-Schnittstelle zum Metadatenabgleich geplant</li> </ul> | <ul style="list-style-type: none"> <li>PND, GND, VDI7</li> </ul>   | <ul style="list-style-type: none"> <li>Ja</li> </ul>   | <ul style="list-style-type: none"> <li>GPL 2.0</li> </ul>  | <ul style="list-style-type: none"> <li>Wikitext</li> </ul> | <ul style="list-style-type: none"> <li>Ja (händische Erstellung von Indizes)</li> <li>Ja</li> </ul>            |
| Scripto        | <ul style="list-style-type: none"> <li>abhängig vom verwendeten CMS. Im Fall von Omeka: Metadatenabgleich über OAL-PMH-Schnittstelle</li> </ul>  | <ul style="list-style-type: none"> <li>Nein</li> </ul>   | <ul style="list-style-type: none"> <li>abhängig vom verwendeten CMS. Im Fall von Omeka: Ja</li> </ul>        | <ul style="list-style-type: none"> <li>GPL 3.0</li> </ul>  | <ul style="list-style-type: none"> <li>Wikitext</li> </ul> | <ul style="list-style-type: none"> <li>Nein</li> </ul>   |

(Tabelle wird fortgesetzt)



| Auffindbarkeit             |   |                          | Langlebigkeit |             | Intellektuelle Redlichkeit |                          |
|----------------------------|---|--------------------------|---------------|-------------|----------------------------|--------------------------|
| Werkzeugname               | Schnittstellen  | Normdaten                | Hosting       | Lizenz      | Codierungsformat           | Indexierungsfunktion     |
| T-PEN                      | • Import von TEI-Schemata   | • Ja (im Rahmen von TEI) | • Ja          | • ECL v.2.0 | • TEI                      | • Ja (im Rahmen von TEI) |
|                            | • Export von Transkripten in XML, PDF und RTF   |                          |               |             |                            |                          |
|                            | • Weiterleitung von Transkripten an teilnehmende Institutionen                        |                          |               |             |                            |                          |
| Bentham Transcription Desk | • Datenimport und -Export im Rahmen der von Mediawiki bereitgestellten Funktionalität |                          | • Ja          | • GPL 2.0   | • TEI                      | • Ja (im Rahmen von TEI) |

Tabelle 7.1: Programmvergleich: Teil 1

| Gebrauchstauglichkeit |               |  |  | Besonderheiten   |
|-----------------------|---------------|--|--|--|
| Werkzeugname          | Editor        | Bearbeitungsstatistiken                            | Benutzerverwaltung                     | Verknüpfung v. Front- u. Back-End  |
| FromThePage           | • Klartextbox | • Fortschrittsbalken für Transkriptionsfortschritt | • Registrierungsformular zur Mitarbeit | • Ja   |
|                       |               | • Top Bearbeiter                                   | • private Transkriptionsprojekte       | • Visualisierung häufig zusammen vorkommender Themen mit Hilfe eines Graphen |
|                       |               | • Top Indexierer                                   | • Lesezugang nach Einladung            | • frei wählbare Kategorien für hierarchisch strukturierte Indexierung        |
|                       |               | • Top Editoren                                     |  |  |

(Tabelle wird fortgesetzt)

| Gebrauchstauglichkeit |  |  |  | Besonderheiten  |
|-----------------------|--|--|--|---|
| Werkzeugname          | Editor   | Bearbeitungsstapen                                     | Benutzerverwaltung   | Verknüpfung v. Front- u. Backend  |
| Refine!               | <ul style="list-style-type: none"> <li>WYSIWYG und optionale Anzeige sowie Bearbeitung von HTML</li> </ul>                           | <ul style="list-style-type: none"> <li>Nein</li> </ul> | <ul style="list-style-type: none"> <li>Zuordnung einzelner Mappen zu bestimmten Bearbeitern</li> <li>Sicherung einzelner Mappen vor Veränderung</li> </ul>   | <ul style="list-style-type: none"> <li>Ja (Freischaltung einzelner Transkriptionen nach Qualitätskontrolle durch Redaktionsteam)</li> <li>automatisierte Generierung von TEL-Dateien</li> </ul> |
| Wikisource            | <ul style="list-style-type: none"> <li>Klartextbox mit Werkzeugleiste</li> </ul>   | <ul style="list-style-type: none"> <li>Ja</li> </ul>   | <ul style="list-style-type: none"> <li>Registrierung sowie anonyme Mitarbeit</li> </ul>  | <ul style="list-style-type: none"> <li>Ja</li> <li>große Community zur inhaltlichen und technischen Weiterentwicklung</li> </ul>  |
| Scripto               | <ul style="list-style-type: none"> <li>Klartextbox (Werkzeugleiste kann eingebunden werden)</li> </ul>                               | <ul style="list-style-type: none"> <li>Nein</li> </ul> | <ul style="list-style-type: none"> <li>abhängig vom verwendeten CMS</li> </ul>   | <ul style="list-style-type: none"> <li>Ja (Import der Transkripte in das CMS durch Moderatoren)</li> <li>CMS-Integration</li> </ul>   |
| T-PEN                 | <ul style="list-style-type: none"> <li>zeilenbasierte Transkription in Textbox mit Syn-taxhervorhebung und Werkzeugleiste</li> </ul> | <ul style="list-style-type: none"> <li>Nein</li> </ul> | <ul style="list-style-type: none"> <li>Einladen von Bearbeitern zu einem Projektteam</li> <li>Hinzufügen und Löschen von Team-Mitgliedern</li> <li>private Projekte</li> <li>Projekte mit reinem Lesezugang für Externe</li> </ul> | <ul style="list-style-type: none"> <li>Import von TEL-Schemata</li> <li>zeilenbasierte Transkription</li> <li>Integration von Hilfsmitteln wie Lexika (erweiterbar)</li> </ul>                  |

(Tabelle wird fortgesetzt)

| Gebrauchstauglichkeit      |  |  |  |   | Besonderheiten   |
|----------------------------|--|--|--|---|--|
| Werkzeugname               | Editor   | Bearbeitungsstatistiken  | Benutzerverwaltung   | Verknüpfung v. Front- u. Backend  |  |
| Bentham Transcription Desk | <ul style="list-style-type: none"> <li>Klartextbox mit Werkzeugleiste</li> </ul> | <ul style="list-style-type: none"> <li>Punktesystem</li> <li>Top Bearbeiter</li> <li>Fortschrittsbalken</li> </ul> | <ul style="list-style-type: none"> <li>Registrierungsformular zur Mitarbeit</li> <li>Editorischer Workflow (Freigabe von Transkripten zur Qualitätskontrolle)</li> </ul> | <ul style="list-style-type: none"> <li>Nein (Workflow im Bentham-Projekt: Export der Transkripte, Umwandlung in XML mit oXygen und Hochladen der XML-Dateien in das digitale Bentham Repository)</li> </ul> | <ul style="list-style-type: none"> <li>TEI-Auszeichnung mit ausgewählten Elementen über eine Werkzeugleiste</li> </ul> |

Tabelle 7.2: Programmvergleich: Teil 2

## 7.2 Programmvergleich anhand der Zielgruppe

Die analysierten Programme besitzen individuelle Schwerpunkte und sind damit für bestimmte Gruppen von Anwendungsfällen mehr oder weniger gut geeignet. Um die Schwerpunkte zu untersuchen und miteinander zu vergleichen werden hier modellhaft einige Zielgruppen mit ihren hypothetischen Anforderungen definiert. Diesen Gruppen werden wiederum für jedes Programm die zur Erfüllung der Anforderungen zweckmäßigen Programmeigenschaften zugeordnet (siehe Abschnitt 7.3 Seite 91). In einem Säulendiagramm visualisiert zählt jede einer Zielgruppe zugeordnete Eigenschaft als Gewichtungspunkt und lässt die Säule um eine Einheit „anwachsen“. Weil diese Darstellung eine Abstraktion von den realen Projektbedingungen darstellt, dürfen die aufgezeigten Tendenzen nur als Hinweis auf die allgemeine Schwerpunktsetzung des Tools genommen werden. Der Kontext eines konkreten Projekts kann durchaus eine bestimmte Eigenschaft größten Wert legen und dadurch andere Aspekte zurückstellen. Dieses kann bei der vorgenommenen Verallgemeinerung der Zielgruppen nicht berücksichtigt werden. Für die Verteilung der Gewichtungseinheiten wird hier für jede Programmeigenschaft der gleiche Wert von einer Einheit angesetzt. Das Layout der verwendeten Diagramme ist mit Hilfe der Tutorials der „Statistiker-WG“ erstellt worden<sup>224</sup>.

### Definition der Zielgruppen

#### Wissenschaft

Die wissenschaftlich orientierte Zielgruppe zeichnet sich durch hohe Anforderungen an die Reichhaltigkeit der Textauszeichnung aus. Ebenso wird ein verhältnismäßig hoher Grad an Anpassungsfähigkeit in Bezug auf die Programmfunktionen gefordert. Die Benutzerschnittstelle sollte intuitiv zu bedienen sein, aber für Projekte in dieser Gruppe ist im Interesse einer höheren Anpassungsfähigkeit des Programms eine entsprechend komplexere Benutzerführung akzeptabel.

#### Familiengeschichte

In familiengeschichtlich orientierten Transkriptionsprojekten spielt vor allem die einfache Benutzbarkeit des Transkriptionsmodus und komfortable Angebote zur Initiierung eigener Projekte eine Rolle. Der Schutz sensibler persönlicher Daten kann in familienge-

---

<sup>224</sup>Vgl. Statistiker-WG. *Statistiker WG – pgf/TikZ-Tutorials*. URL: <http://www.pgf.statistiker-wg.de/> (besucht am 02.08.2012).

schichtlichen Projekten eine wichtige Rolle spielen. Die Verknüpfung von Transkriptionsmodus und Präsentationsebene erleichtert den Workflow für Projekte in dieser Kategorie.

### **Archive**

Die Archive legen ebenfalls wissenschaftliche Maßstäbe an ihre Editionen an, aber es handelt sich in vielen Fällen um die Transkription umfangreicher Bestände, bei denen eine flache semantische Auszeichnung für die spätere (wissenschaftliche) Nutzung ausreicht. Weiterhin sind Archive um die Integration und Erweiterung bestehender Normdaten bemüht. Die Benutzerschnittstelle sollte überschaubar sein, damit sich der Bearbeiter vornehmlich mit den Schwierigkeiten der Transkription selbst beschäftigen kann.

### **Bibliotheken**

Die Bibliotheken legen ebenfalls großen Wert auf die Referenzierung und Anreicherung von Normdaten. Der Verweis auf bibliographische Daten aus eigenen Katalogen spielt gleichermaßen eine wichtige Rolle. Ferner handelt es sich bei den zu transkribierenden Dokumentengruppen häufig um Nachlässe und Sammlungen, die gemeinsame Strukturmerkmale aufweisen: Sie sind nach Kästen und Mappen geordnet.

### **Museen**

Museen sammeln vielfältige Arten von Gegenständen und setzen dabei in der Regel ihren Schwerpunkt nicht wie die vorangehenden Gruppen auf Papierdokumente als historische Quellen. Für Transkriptionsprojekte kommen in Museen neben textuellen Quellen auch Audio- und Videodokumente in Frage. Weiterhin verfügen viele Museen über handschriftliche Eingangskataloge und Karteikarten oder Notizbücher zur Beschreibung von Museumsobjekten. Neben der Transkription dieser Unterlagen ist für Museen auch ein Verweis vom Transkript auf das entsprechende Objekt wichtig.

| Werkzeugname | Wissenschaft  | Familiengeschichte                                     | Archive  | Bibliotheken   | Museen  |
|--------------|---|--|--|--|---|
| FromThePage  | 1. Indexierung nach frei wählbaren Kategorien                               | 1. Private Transkriptionsprojekte                      | 1. überschaubare Benutzer-schnittstelle                | 1. Import von Daten aus dem Internet Archive           | 1. überschaubare Benutzer-schnittstelle                               |
|              | 2. Analysewerkzeug für indexierte Themen (Visualisierung mit einem Graphen) | 2. Verknüpfung von Back- und Front-End                 | 2. Ausgabe von Bearbeitertatsachen                     | 2. überschaubare Benutzer-schnittstelle                | 2. Ausgabe von Bearbeitertatsachen                                    |
| Refine!      | 3. Freischaltung der Transkriptionen nach Kontrolle durch Moderatoren       | 3. Hostung   | 3. Verknüpfung von Back- und Front-End                 | 3. Ausgabe von Bearbeitertatsachen                     | 3. Verknüpfung von Back- und Front-End                                |
|              |   | 4. Überschaubarer Benutzer-schnittstelle               | 4. Hosting   | 4. Verknüpfung von Back- und Front-End                 | 4. Hosting  |
| Wikisource   | 1. Verweis auf Normdaten  | 5. Ausgabe von Bearbeitertatsachen                     |  | 5. Hosting   |   |
|              | 2. Automatisierte Generierung von TEL-Dateien                               |  | 1. Verknüpfung von Back- und Front-End                 | 1. Verknüpfung von Back- und Front-End                 | 1. Verknüpfung von Back- und Front-End                                |
| Scripto      | 3. Freischaltung der Transkriptionen nach Kontrolle durch Moderatoren       |  | 2. Automatisierte Generierung von TEL-Dateien          | 2. Automatisierte Generierung von TEL-Dateien          | 2. Verweis auf Normdaten  |
|              |   |  | 3. Verweis auf Normdaten                               | 3. Verweis auf Normdaten                               | 3. Werkzeugeleiste zur Textauszeichnung                               |
|              | 4. Verweis auf Normdaten  |  | 4. Werkzeugeleiste zur Textauszeichnung                | 4. Werkzeugeleiste zur Textauszeichnung                |   |
|              | 5. inhaltliche Vernetzung mit anderen Wikipedia-Ressourcen                  |  | 5. Besondere Eignung für Nachlasseditionen             |  |   |
|              | 1. vielfältige Diskussionsfunktionen  | 1. Verknüpfung von Back- und Front-End                 | 1. Verknüpfung von Back- und Front-End                 | 1. Verknüpfung von Back- und Front-End                 | 1. Verknüpfung von Back- und Front-End                                |
|              | 2. Qualitätskontrolle   | 2. überschaubare Benutzer-schnittstelle                | 2. überschaubare Benutzer-schnittstelle                | 2. überschaubare Benutzer-schnittstelle                | 2. überschaubare Benutzer-schnittstelle                               |
|              | 3. Kollaborative Erweiterung und Anpassung des Funktionsumfangs             | 3. Hosting   | 3. hoher Bekanntheitsgrad                              | 3. hoher Bekanntheitsgrad                              | 3. hoher Bekanntheitsgrad   |
|              | 4. Verweis auf Normdaten  | 4. Hochladen eigener Scans                             | 4. aktive Community                                    | 4. aktive Community                                    | 4. aktive Community   |
|              | 5. inhaltliche Vernetzung mit anderen Wikipedia-Ressourcen                  | 5. Hosting   | 5. Hosting   | 5. Hosting   | 5. Hosting  |
|              |   |  | 6. Verweis auf Normdaten                               | 6. Verweis auf Normdaten                               | 6. Verweis auf Normdaten  |
|              | 1. Import der Transkripte in CMS durch Moderatoren                          | 1. überschaubare Benutzer-schnittstelle                | 1. überschaubare Benutzer-schnittstelle                | 1. überschaubare Benutzer-schnittstelle                | 1. überschaubare Benutzer-schnittstelle                               |
|              | ermöglicht Qualitätskontrolle   | 2. Import der Transkripte in CMS durch Moderatoren     | 2. Import der Transkripte in CMS durch Moderatoren     | 2. Import der Transkripte in CMS durch Moderatoren     | 2. Import der Transkripte in CMS durch Moderatoren                    |
|              |   | 3. Transkription digitaler Objekte aus bestehendem CMS | 3. Transkription digitaler Objekte aus bestehendem CMS | 3. Transkription digitaler Objekte aus bestehendem CMS | 3. Transkription digitaler Objekte aus bestehendem CMS                |
|              |   |  |  |  | 4. neben textuellen Quellen auch Transkription von Multimedia-dateien |

(Tabelle wird fortgesetzt)

| Werkzeugname               | Wissenschaft  | Familien-geschichte                      | Archive                                    | Bibliotheken                                  | Museen  |
|----------------------------|---|--|--|---|---|
| T-PEN                      | 1. Volle TEI-Unterstützung durch Import entsprechender XML-Schemata | 1. Hochladen eigener Scans               | 1. Integration von Transkriptionsprojekten | 1. Integration von Transkriptionshilfsmitteln | 1. Integration von Transkriptionshilfsmitteln |
|                            | 2. Integration von Transkriptionshilfsmitteln                       | 2. Private Transkriptionsprojekte        | 2. Funktionen zum Projektmanagement        | 2. Funktionen zum Projektmanagement           | 2. Funktionen zum Projektmanagement           |
|                            | 3. Assistent für TEI-Markup und Sonderzeichen                       | 3. Hosting                               | 3. Hosting                                 | 3. Hosting                                    | 3. Hosting                                    |
|                            | 4. Anpassbarer Datenexport  |  |  |   |   |
|                            | 5. Funktionen zum Projektmanagement                                 |  |  |   |   |
| Bentham Transcription Desk | 1. TEI Markup   | 1. überschaubare Benutzerschnittstelle   | 1. überschaubare Benutzerschnittstelle     | 1. überschaubare Benutzerschnittstelle        | 1. überschaubare Benutzerschnittstelle        |
|                            | 2. Workflow zur Qualitätskontrolle                                  | 2. Wiedergabe von Bearbeitersstatistiken | 2. Workflow zur Qualitätskontrolle         | 2. Workflow zur Qualitätskontrolle            | 2. Workflow zur Qualitätskontrolle            |
|                            |   | 3. Wiedergabe von Bearbeitersstatistiken | 3. Wiedergabe von Bearbeitersstatistiken   | 3. Wiedergabe von Bearbeitersstatistiken      | 3. Wiedergabe von Bearbeitersstatistiken      |

Tabelle 7.3: Zielgruppenspezifische Verteilung der Programmeigenschaften

## FromThePage

Das Programm bietet auch für wissenschaftlich orientierte Transkriptionsprojekte wertvolle Funktionen. Allerdings werden für die meisten Projekte in diesem Bereich die Möglichkeiten zur Textauszeichnung zu gering ausfallen. Der Bereich der Familiengeschichtsforschung zeichnet sich in der Gewichtung der zielgruppenspezifischen Merkmale deutlich ab. Das Werkzeug wurde ursprünglich zu diesem Zweck konzipiert und seine Vorteile für derartige Projekte sind auf eine gezielte Programmentwicklung für diesen Bereich zurückzuführen. Die für familiengeschichtlich orientierte Projekte zu verzeichnenden Programmfunktionen spielen auch in den Bereichen Archive, Bibliotheken und Museen eine wichtige Rolle. Hinzu kommt im Fall der Bibliotheken noch die Schnittstelle zum Internet Archive, weil hier viele Drucke schon in digitalisierter Form vorliegen.

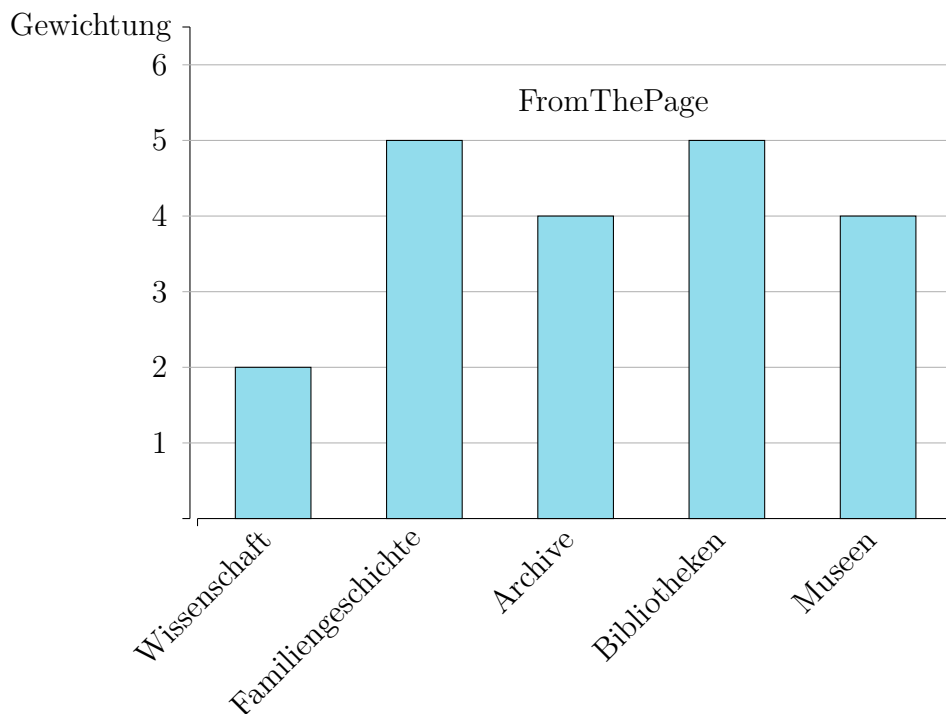


Diagramm 7.1: FromThePage: Zielgruppenspezifische Eigenschaften

## Refine!

Das Programm stellt eine bedenkenswerte Option auch für wissenschaftlich orientierte Projekte dar, wobei in diesem Fall lediglich auf eine beschränkte Auswahl von TEI-Elementen zurückgegriffen werden kann. Für wissenschaftliche Projekte, die eine Edition



mit flacher Textcodierung realisieren wollen, ist zudem die Möglichkeit einer Qualitätskontrolle (Redaktionssystem) sowie der Verweis auf Normdaten vorteilhaft. Im Bereich Familiengeschichte ist das Programm weniger sinnvoll einzusetzen. Hier gibt es andere Programme, die mehr im Hinblick auf die spezifischen Anforderungen hin konzipiert sind. Den stärksten Ausschlag gibt es in der Gewichtung der Funktionsmerkmale für Archive und Bibliotheken. Diese ähneln sich stark in ihren Anforderungen und sollten das Programm bei einer Auswahlentscheidung berücksichtigen. Zudem eignet sich das Werkzeug auf Grund seiner Entwicklung im Zusammenhang mit einer Nachlassedition in besonderem Maße für diese, häufig in Bibliotheksbeständen vorhandene, Dokumentgruppe. Für einige Projekte im Museumsbereich könnte die TEI-Auszeichnung schon unverhältnismäßig sein, denn Eingangskataloge etc. müssen in der Regel nicht nach philologisch exakten Methoden erfasst werden. Für andere Projekte wie etwa die Transkription von Forschungstagebüchern in einem Naturkundemuseum wäre der Einsatz des Programms aber wieder überlegenswert.

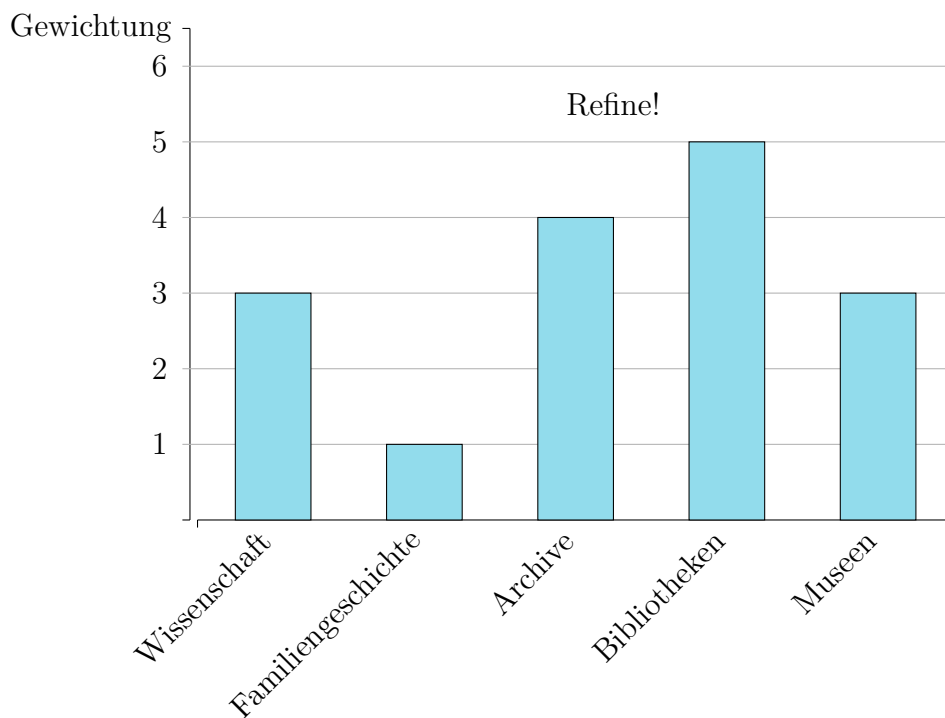


Diagramm 7.2: Refine!: Zielgruppenspezifische Eigenschaften

### **Wikisource**

Die deutsche Sprachversion von Wikisource bemüht sich um die Edition wissenschaftlich verwertbarer Texte. Diesem Anspruch wird man auch in verschiedenen Einzelprojekten, wie das Beispiel „Field Notes of Junius Henderson“ aus der englischen Sprachversion zeigt, gerecht (siehe Abschnitt 6.1.3 Seite 61). Dennoch sind die Möglichkeiten der verwendeten Textauszeichnung beschränkt und können nicht alle wissenschaftlichen Ansprüche im Rahmen einer reichhaltigen Textcodierung erfüllen. Die kollaborative Weiterentwicklung von Programmfunktionen verspricht aber diesen Umstand zumindest für wissenschaftliche Projekte ohne eine explizite Notwendigkeit zur Arbeit mit TEI abzumildern. Für den Bereich Familiengeschichtsforschung ist die Möglichkeit zum Hosting eigener Scans und die Funktionalität des Editors ProofreadPage vorteilhaft. Optionen zum sensiblen Umgang mit Personendaten oder das Anlegen privater Projekte sind dagegen in der gehosteten Version nicht vorhanden. Archive, Bibliotheken und Museen können bei der Kooperation mit Wikisource in gleicher Weise von Programmeigenschaften wie der überschaubaren Benutzeroberfläche von ProofreadPage und dem Verweis auf Normdaten profitieren. Zudem spielt die Öffentlichkeitswirksamkeit des Schwesterprojekts von Wikipedia eine Rolle und die Suche nach Freiwilligen wird durch eine aktive Community erleichtert. Die Hosting-Variante bietet sich besonders für kleinere Institutionen an, aber sie wird auch z. B. vom Nationalarchiv der Vereinigten Staaten genutzt (siehe Abschnitt 6.1.3 Seite 61).

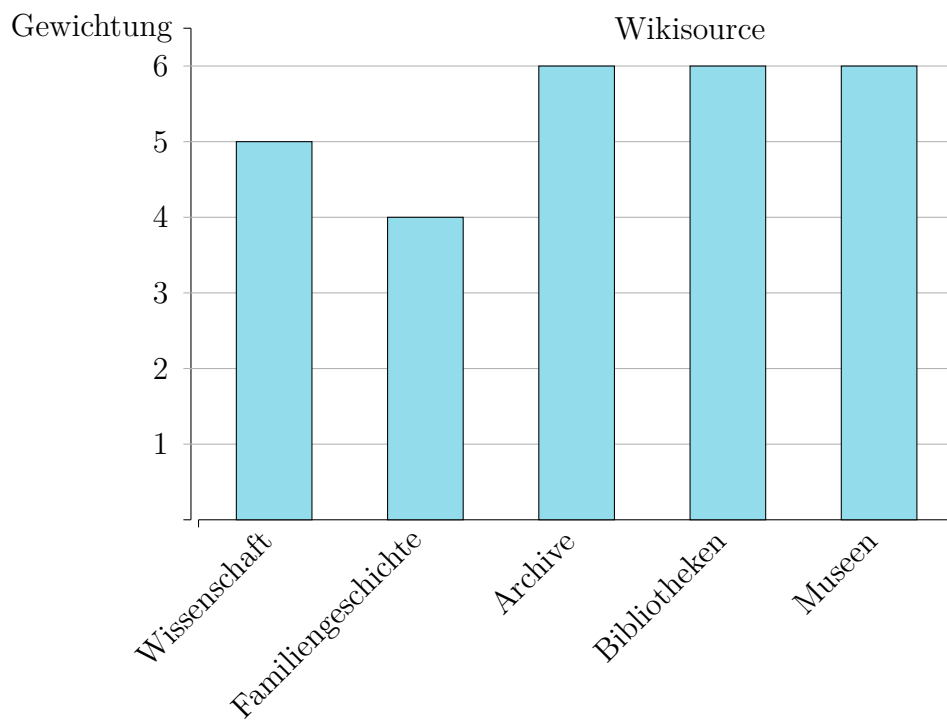


Diagramm 7.3: Wikisource: Zielgruppenspezifische Eigenschaften

### Scripto

Im Vergleich zu den anderen betrachteten Programmen beschränkt sich Scripto auf grundlegende Funktionen zur Transkription selbst und zur Verwaltung der Transkripte. Dieser Ansatz kann für viele Projekte ein Vorteil sein, so etwa wenn die digitalen Objekte eines bestehenden CMS transkribiert werden sollen. Viele der im Kontext der anderen Werkzeuge beschriebenen Funktionen müssten bei der Arbeit mit Scripto von dem CMS übernommen werden. Auf Grund der beschränkten Möglichkeiten zur Textauszeichnung ist das Programm für wissenschaftlich orientierte Editionen lediglich als eine Möglichkeit anzusehen die reinen Transkriptionen mit Mitteln des Crowdsourcing zu erstellen und die entstandenen Texte zur Grundlage für eine mit anderen Hilfsmitteln zu gestaltende Edition zu machen<sup>225</sup>. Bei Projekten im Bereich der Familiengeschichtsforschung werden häufig Angebote mit einer Möglichkeit zum Hosting gesucht. Letzteres kann für Scripto als reines Plugin nicht in die Bewertung eingehen, aber ein solcher Ansatz wäre z. B. im Rahmen von Omeka.net<sup>226</sup> realisierbar. Bei Vorhandensein einer digitalen Sammlung in einem CMS wäre der Plugin-Charakter für Archive, Bibliotheken und Museen in gleicher Weise ein Vorteil. Zudem kann Scripto zur Erfassung des Inhalts von Multimediadateien eingesetzt werden.

---

<sup>225</sup>Vgl. Center for History and New Media, *About Scripto*.

<sup>226</sup>Vgl. Roy Rosenzweig Center for History and New Media. *omeka.net*. URL: <http://www.omeka.net/> (besucht am 26.07.2012).

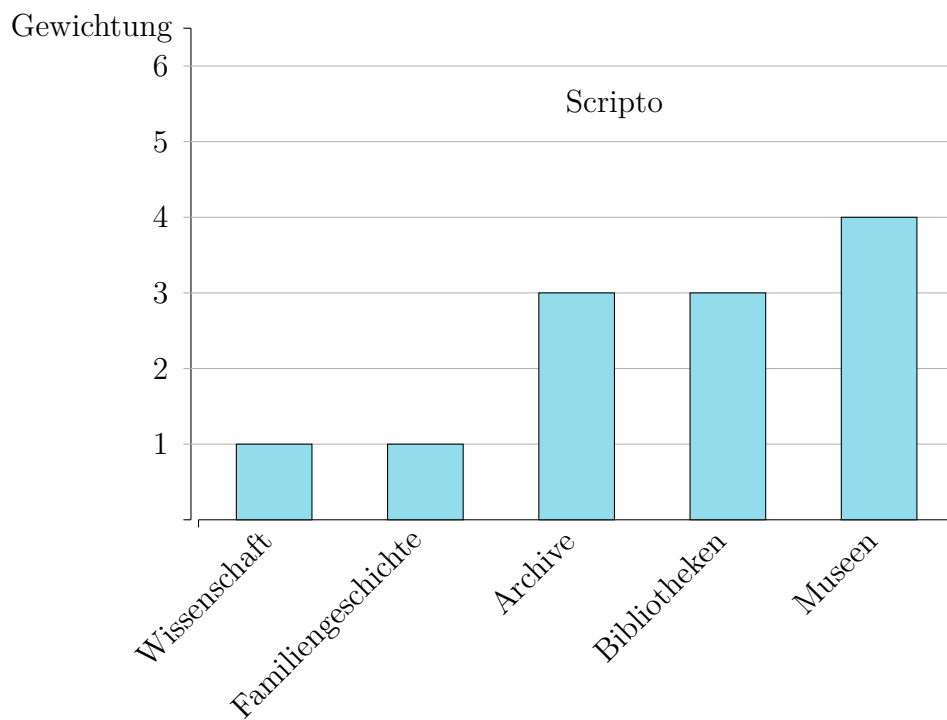


Diagramm 7.4: Scripto: Zielgruppenspezifische Eigenschaften

## **T-PEN**

Das Programm ist als Transkriptionswerkzeug für die Digital Humanities konzipiert. Letzters macht sich auch in der Gewichtung der projektspezifischen Programmeigenschaften bemerkbar. Der Schwerpunkt liegt auf der Anpassungsfähigkeit für verschiedene TEI-Schemata und der reichhaltigen Textauszeichnung. Ebenso können die Werkzeu-  
gleiste sowie die Transkriptionshilfsmittel projektspezifisch angepasst werden. Die Erstellung von Transkripten mit T-PEN kann die Grundlage für eine weitere Analyse und Aufbereitung der Textdaten bieten und das Werkzeug verfügt mit anpassbaren Exportfunktionen über einen komfortablen Ausgangspunkt für diese Arbeit. Der Einsatz des Programms für die Familiengeschichtsforschung ist auf Grund der fehlenden Funktionen zur Erstellung einer Präsentationsebene zur Web-Publikation nur eingeschränkt möglich. Die vielfältigen Anpassungsmöglichkeiten in Bezug auf die Textcodierung werden von den meisten Projekten in diesem Bereich nicht benötigt werden. Archive, Bibliotheken und Museen sollten das Transkriptionswerkzeug bei der Auswahlentscheidung berücksichtigen, wenn bei den konkreten Projekten großer Wert auf eine detaillierte Erfassung von Textmerkmalen gelegt wird. In diesem Fall kommt auch ihnen die Integration von Transkriptionshilfsmitteln zugute. Die Erstellung einer Präsentationsebene müsste dann allerdings in einem zweiten Arbeitsschritt erfolgen. Nachteilig wird sich für familiengeschichtlich orientierte Projekte und Gedächtnisorganisationen die fehlende Präsentation von Bearbeiterstatistiken auswirken.

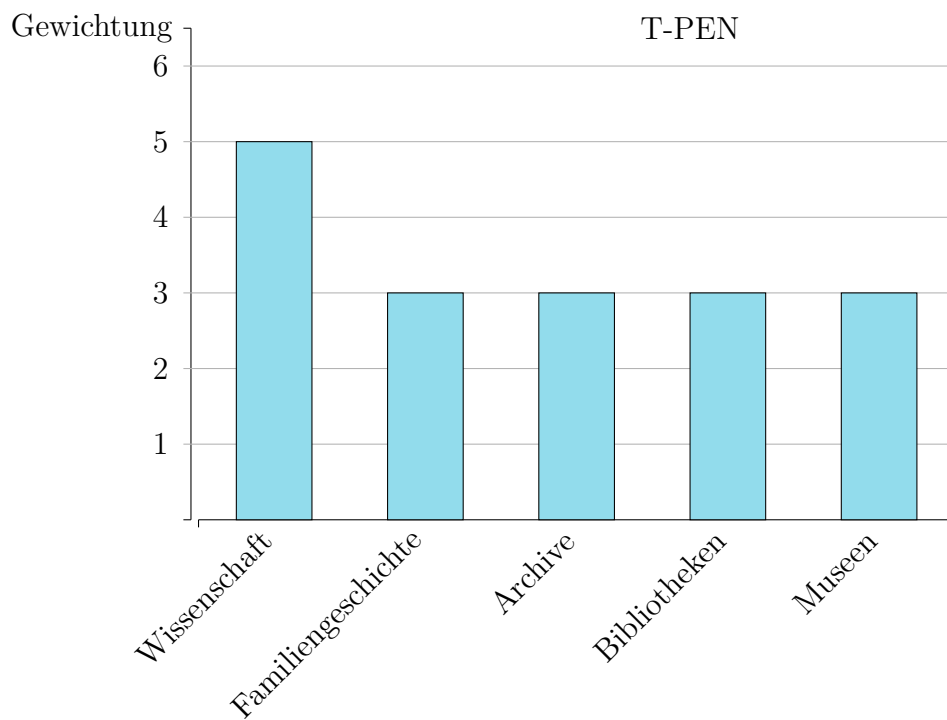


Diagramm 7.5: T-PEN: Zielgruppenspezifische Eigenschaften

### Bentham Transcription Desk

Das Transkriptionswerkzeug erlaubt durch das Freischalten von Transkriptionen zur Qualitätskontrolle einen Workflow, der in Projekten in den Bereichen Wissenschaft und Gedächtnisorganisationen gefordert sein kann. In familiengeschichtlich orientierten Vorhaben kommt dieser Vorteil nicht so sehr zum Tragen, weil die Qualitätskontrolle hier eher kollaborativ und nicht mit Hilfe eines Moderatorenteams durchgeführt wird. Zudem fehlt die für diese Projekte wichtige direkte Verknüpfung von Back- und Front-End. Die Wiedergabe von Bearbeiterstatistiken als wichtiges Mittel zur Motivation von Freiwilligen wird unterstützt und spielt besonders in der Familiengeschichtsforschung sowie bei den Archiven, Bibliotheken und Museen eine wichtige Rolle. Die Vergabe ausgewählter TEI-Elemente mit Hilfe einer Werkzeugleiste ist vor allem für Wissenschaft und Gedächtnisorganisationen interessant. Sie müsste allerdings in den meisten Fällen projektspezifisch angepasst werden.

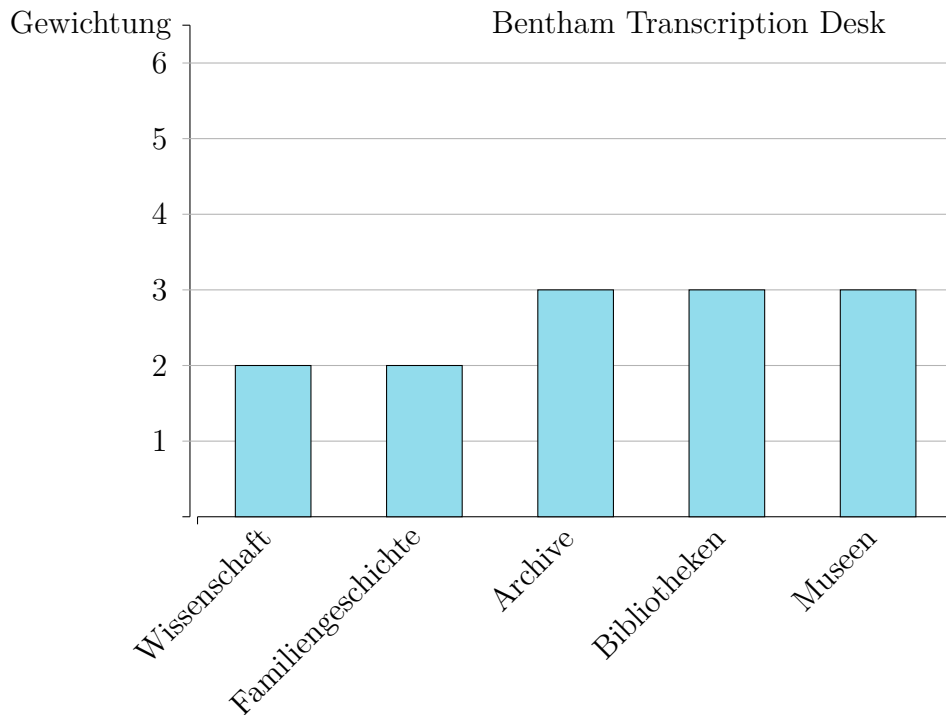


Diagramm 7.6: Bentham Transcription Desk: Zielgruppenspezifische Eigenschaften



## 8 Fazit

Die im Rahmen der vorliegenden Arbeit durchgeführte Analyse von Editionswerkzeugen zur nutzergenerierten Transkription handschriftlicher Quellen bestätigt auf der einen Seite die eingangs von Ben W. Brumfield zitierte Bemerkung, dass der Entwicklung eines einzelnen Werkzeugs für die verschiedenen Ansprüche von Transkriptionsprojekten der gegenwärtige Arbeitsaufwand der Entwickler für die Anpassung der Funktionalität in Bezug auf die Anforderungen ihrer eigenen Manuskripte entgegensteht<sup>227</sup>. Auf der anderen Seite konnte gezeigt werden, in welchem Rahmen die projektspezifischen Weiterentwicklungen auch für andere Projekte vermehrt nutzbar gemacht werden können und welche Programmeigenschaften für bestimmte Anwenderkreise besondere Wichtigkeit besitzen.

Die Anstrengungen zur Entwicklung eines Werkzeugs mit einem hohen Maß an Interoperabilität, d. h. Einsetzbarkeit unter heterogenen Rahmenbedingungen, führt namentlich bei den Programmen FromThePage, Wikisource, Scripto und T-PEN zu überzeugenden Lösungen für bestimmte Anwendungsfälle. Während FromThePage vornehmlich für Projekte in der Familiengeschichtsforschung eine leicht zu bedienende Benutzeroberfläche zur Verfügung stellt und durch die Verknüpfung von Transkriptionsmodus und Präsentationsebene eine unverzügliche Publikation der Transkripte erlaubt, überzeugt Wikisource besonders für Projekte in Gedächtnisorganisationen mit Möglichkeiten zur Verknüpfung mit anderen Ressourcen und der kollaborativen Entwicklung von Programmfunktionen sowie der Einbeziehung der Community in den mit einer Qualitätskontrolle verbundenen Transkriptionsworkflow. Im Fall von Scripto wurden durch die Integrierbarkeit des Plugin in verschiedene CMS die Weichen für seinen Einsatz in bestehenden digitalen Sammlungen gestellt. Im Zusammenhang mit Omeka bietet es gute Voraussetzungen zur Publikation einer reichhaltigen Sammlung digitaler Objekte, wobei das Ordnungsprinzip des CMS bildzentriert ist und die weiteren Entwicklungen für seinen Einsatz im Rahmen textzentrierter Projekte abgewartet werden müssen. T-PEN besitzt schließlich seine Stärken im Bereich wissenschaftlich orientierter Transkriptionsprojekte, die mit

---

<sup>227</sup>Vgl. Brumfield, 2010: *The Year of Crowdsourcing Transcription*.

dem Programm auf eine von den anderen Werkzeugen bei weitem nicht erreichte Anpassungsfähigkeit im Rahmen der verwendeten TEI-Elemente und Werkzeugleisten ohne den Eingriff eines Administrators zurückgreifen können.

Das Werkzeug Refine! ist mit seiner Orientierung an Nachlasseditionen vornehmlich für Bibliotheken interessant. Im Vergleich zu den vorhergehenden Tools werden Anpassungen allerdings in höherem Maße notwendig sein, weil es mehr auf den ursprünglichen Entwicklungszweck zugeschnitten ist. Zudem wäre die Bereitstellung eines technischen Support für potentielle Anwender von großer Wichtigkeit. Bentham Transcription Desk zeichnet sich besonders durch die Integration von Ranglisten nach einem Punktesystem, Diskussionsforen und anderen interaktiven Elementen aus. Auch die Verwendung ausgewählter TEI-Tags wird auf vergleichbare Anwendungsfälle übertragbar sein. Die Anpassungsfähigkeit der Programmfunktionen ohne tiefgehende technische Vorkenntnisse bleibt allerdings auch in diesem Fall hinter den erstgenannten Programmen zurück. Die angekündigten weiteren Entwicklungen in dieser Hinsicht<sup>228</sup> würden für die Anwender einen bedeutenden Mehrwert generieren.

Die Weiterentwicklungen auf dem Markt der Transkriptionswerkzeuge könnten in einem Ausbau einiger der hier vorgestellten Lösungsansätze bestehen. Wünschenswert wäre etwa die Arbeit an den Schnittstellen zur Einspeisung der codierten Transkripte in ein CMS, sodass der Übergang vom Transkriptionsmodus zur Präsentationsebene verbessert wird. Dabei wäre im Interesse vieler Editionsprojekte auch an der Umsetzung textzentrierter Präsentationsformen zu arbeiten. Eine Bereitstellung von Stilvorlagen für die standardmäßig im Transkriptionsmodus angebotenen Auszeichnungselemente würde auch kleineren Projekten zu einer unproblematischen und in Bezug auf das Layout homogenen Präsentationsform für ihre Texte verhelfen. Offene Arbeitsumgebungen wie Wikisource ermöglichen projektspezifische Weiterentwicklungen mit Hilfe von engagierten Freiwilligen und gewinnen auch durch die Mitarbeit institutioneller Partner. Themenseiten zur Diskussion der Anforderungen der Codierung bestimmter Textmerkmale sind ein gutes Mittel, um die Anstrengungen über den Bedarf eines konkreten Projekts hinaus auch für andere Vorhaben greifbar zu machen. Neben in Bezug auf die Textcodierung anspruchsvolleren Projekten bieten sich für kleinere Vorhaben mit einem Schwerpunkt auf der Transkription der Inhalte ihrer Manuskripte schon jetzt einige Möglichkeiten zur Teilnahme an gehosteten Angeboten, die grundlegende Funktionen der Transkription und Präsentation zur Verfügung stellen. Hier existieren noch Möglichkeiten, um den

---

<sup>228</sup>Vgl. Moyle, Tonra und Wallace, „Manuscript Transcription by Crowdsourcing: Transcribe Bentham“, S. 349.

Anwenden die Konzentration auf die inhaltlichen Aspekte ihrer Projekte zu erleichtern. Dazu zählt die die Individualisierung ihres Internetauftritts mit Themenseiten, Transkriptionsrichtlinien, Blogs und Social Media. Schließlich dürfen die Freiwilligen bei der Konzeption eines Transkriptionsprojektes nicht vergessen werden und es reicht nicht aus, die grundsätzliche Funktionalität zur Arbeit mit den Handschriften zur Verfügung zu stellen. Vielmehr muss das Interesse der Teilnehmer durch inhaltliche Angebote zum Kontext der Manuskripte sowie spielerische Aspekte und Maßnahmen zur Identifikationsbildung gewonnen werden.

## 9 Literaturverzeichnis

### Monographien, Zeitschriftenartikel, Online-Dokumente

*Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities.*

22. Okt. 2003. URL:

[http://www.zim.mpg.de/openaccess-berlin/berlin\\_declaration.pdf](http://www.zim.mpg.de/openaccess-berlin/berlin_declaration.pdf) (besucht am 08.07.2012).

Causer, Tim, Justin Tonra und Valerie Wallace. „Transcription maximized; expense minimized? crowdsourcing and editing The Collected Works of Jeremy Bentham“.

In: *Literary and Linguistic Computing* 27.2 (2012), S. 119–137. DOI:

10.1093/llc/fqs004. eprint: <http://llc.oxfordjournals.org/content/early/2012/03/28/llc.fqs004.full.pdf+html>. URL: [http:](http://llc.oxfordjournals.org/content/early/2012/03/28/llc.fqs004.abstract)

[//llc.oxfordjournals.org/content/early/2012/03/28/llc.fqs004.abstract](http://llc.oxfordjournals.org/content/early/2012/03/28/llc.fqs004.abstract) (besucht am 21.06.2012).

Entlich, Richard. „Handwriting Recognition for Historical Documents“. In: *RLG*

*DigiNews* 8.1 (2004). URL: <http://chnm.gmu.edu/digitalhistory/links/cached/chapter3/link3.33b.RLGhandwriting.html> (besucht am 14.05.2012).

Estellés-Arolas, Enrique und Fernando González-Ladrón-de Guevara. „Towards an integrated crowdsourcing definition“. In: *Journal of Information Science* 38.2 (2012), S. 189–200. URL:

<http://www.crowdsourcing-blog.org/wp-content/uploads/2012/02/Towards-an-integrated-crowdsourcing-definition-Estell%C3%A9s-Gonz%C3%A1lez.pdf> (besucht am 05.07.2012).

Feldmann, Bianca. „OCR von Handschriften. Ein Forschungsüberblick“. In: *Fundus – Forum für Geschichte und ihre Quellen* (1), S. 107–143. URL:

<http://webdoc.gwdg.de/edoc/p/fundus/1/feldmann.pdf> (besucht am 14.05.2012).

Holley, Rose. „Crowdsourcing: How and Why Should Libraries Do It?“ In: *D-Lib Magazine* 16.3/4 (2012). URL:

- <http://www.dlib.org/dlib/march10/holley/03holley.html> (besucht am 05.07.2012).
- International Organization for Standardization (ISO). *ISO 9241-11: Ergonomic requirements for office work with visual display terminals (VDTs) - Guidance on usability*. URL:  
<http://www.it.uu.se/edu/course/homepage/acsd/vt09/ISO9241part11.pdf>  
(besucht am 11.07.2012).
- Jannidis, Fotis, Hrsg. *Bewertungskriterien für elektronische Editionen*. IASL Diskussionsforum online. 1999. URL:  
<http://iasl.uni-muenchen.de/discuss/lisforen/jannidis.htm> (besucht am 29.04.2012).
- Lexikon der deutschen Dichter und Prosaisten vom Beginn des 19. Jahrhunderts bis zur Gegenwart*. 6., völlig neu bearbeitete und stark vermehrte Auflage. 8 Bde. [Reprint (1975): Neudeln / Liechtenstein: Kraus]. Leipzig, 1913. URL:  
<http://archive.org/details/lexikonderdeutsc01bruoft> (besucht am 24.07.2012).
- Märtn, Björn und Christian Thomas. „Das Wuchern der Archive. Die digitale Edition des Nachlasses Franz Brümmer mit dem Refine!Editor“. In: *editio* 22 (2008).
- Middell, Gregor. *Vortrag auf dem COST Action 32 – Workgroup-2-Treffen in Ancona, Italien (19.-20. Mai 2008)*. URL: [http://bruemmer.staatsbibliothek-berlin.de/nlbruemmer/materialien/2008-05-19\\_COST\\_A32\\_W2\\_Vortrag\\_en.pdf](http://bruemmer.staatsbibliothek-berlin.de/nlbruemmer/materialien/2008-05-19_COST_A32_W2_Vortrag_en.pdf)  
(besucht am 30.06.2012).
- Moyle, Martin, Justin Tonra und Valerie Wallace. „Manuscript Transcription by Crowdsourcing: Transcribe Bentham“. In: *Liber Quarterly* 20.3/4 (2011), S. 347–256. URL: <http://liber.library.uu.nl/index.php/lq/article/view/7999/8329>  
(besucht am 24.07.2012).
- Pape, Sebastian, Christof Schöch und Lutz Wegner. „TEICHI and the Tools Paradox“. In: *Journal of the Text Encoding Initiative* (2 2012). URL:  
<http://jtei.revues.org/432> (besucht am 05.05.2012).
- Romary, Laurent. „Questions & Answers for TEI Newcomers“. In: *Jahrbuch für Computerphilologie* 10 (2008). URL:  
<http://computerphilologie.de/jg08/romary.pdf> (besucht am 29.04.2012).
- San Diego Natural History Museum. *FromThePage: User Guide*. URL:  
<http://files.balboaparkonline.org/psully/FromThePageUserGuide.pdf>  
(besucht am 26.06.2012).

Sperberg-McQueen, C. M. „Textual Criticism and the Text Encoding Initiative“. In: *MLA '94*. 1994. URL: <http://xml.coverpages.org/sperb-mla94.html> (besucht am 28.04.2012).

Thomer, Andrea u. a. „From documents to datasets: A MediaWiki-based method of annotating and extracting species observations in century-old field notebooks“. In: *ZooKeys* 209 (2012), 235–253. DOI: 10.3897/zookeys.209.3247. URL: <http://www.pensoft.net/journals/zookeys/article/3247/> (besucht am 26.07.2012).

## Webseiten

3pc GmbH Neue Kommunikation. *3pc GmbH Neue Kommunikation*. URL: <http://3pc.de/> (besucht am 30.06.2012).

Andrew W. Mellon Foundation. *Andrew W. Mellon Foundation*. URL: <http://www.mellon.org/> (besucht am 15.07.2012).

Arts and Humanities Research Council. *Arts and Humanities Research Council*. URL: <http://www.ahrc.ac.uk/Pages/default.aspx> (besucht am 22.07.2012).

Bibliothèque Nationale de France. *gallica – Bibliothèque Numérique*. URL: <http://gallica.bnf.fr/> (besucht am 25.06.2012).

Bosse, Arno. *Creating a resource center for texts*. Forumsbeitrag. URL: <http://omeka.org/forums/topic/creating-a-ressource-center-for-texts> (besucht am 28.07.2012).

Brandenburgisches Landeshauptarchiv. *Brandenburgisches Landeshauptarchiv*. URL: <http://www.landeshauptarchiv-brandenburg.de/> (besucht am 16.07.2012).

Brokfeld, Jens. *Quellen zur Familiengeschichtsforschung auf Wikisource?* Forumsbeitrag. 2. Aug. 2012. URL: [https://de.wikisource.org/wiki/Wikisource:Skriptorium#Quellen\\_zur\\_Familiengeschichtsforschung\\_auf\\_Wikisource.3F](https://de.wikisource.org/wiki/Wikisource:Skriptorium#Quellen_zur_Familiengeschichtsforschung_auf_Wikisource.3F) (besucht am 02.08.2012).

Brumfield, Ben W. *2010: The Year of Crowdsourcing Transcription*. Blog-Eintrag. 2. Feb. 2011. URL: <http://manuscripttranscription.blogspot.de/2011/02/2010-year-of-crowdsourcing.html> (besucht am 09.05.2012).

– *Collaborative Manuscript Transcription*. Blog. URL: <http://manuscripttranscription.blogspot.de/> (besucht am 09.05.2012).

- Brumfield, Ben W. *Developments in Wikisource/ProofreadPage for Transcription*. URL: <http://manuscripttranscription.blogspot.de/2011/12/developments-in-wikisourceproofreadpage.html> (besucht am 12.07.2012).
- *FromThePage Installationsanweisungen*. URL: <https://github.com/benwbrum/fromthepage/wiki> (besucht am 28.06.2012).
  - *My Goals for FromThePage*. Blog-Eintrag. 15. Feb. 2011. URL: <http://manuscripttranscription.blogspot.de/2011/02/my-goals-for-fromthepage.html> (besucht am 26.06.2012).
  - *Transcription Tool List*. Blog-Eintrag. 11. Apr. 2012. URL: <http://manuscripttranscription.blogspot.de/2012/04/crowdsourced-transcription-tool-list.html> (besucht am 14.05.2012).
  - *Transcription Tools at TCDL 2012*. URL: <http://manuscripttranscription.blogspot.de/2012/05/transcription-tools-at-tcdl2012.html> (besucht am 06.07.2012).
  - *Wikisource for Manuscript Transcription*. Blog-Eintrag. 21. Juli 2010. URL: <http://manuscripttranscription.blogspot.de/2009/07/wikisource-for-manuscript-transcription.html> (besucht am 12.07.2012).
- Brumfield, Julia Ann Craddock. *Julia Brumfield Diaries*. Hrsg. von Ben W. Brumfield. URL: [http://beta.fromthepage.com/JuliaBrumfield?ol=s\\_sp\\_diaries](http://beta.fromthepage.com/JuliaBrumfield?ol=s_sp_diaries) (besucht am 30.06.2012).
- Center for Digital Theology. *About T-PEN*. URL: <http://t-pen.org/TPEN/about.jsp> (besucht am 16.07.2012).
- *Center for Digital Theology*. URL: <http://www.slu.edu/x27122.xml> (besucht am 15.07.2012).
  - *T-PEN*. URL: <http://t-pen.org/TPEN/> (besucht am 13.06.2012).
  - *T-PEN*. Blog. URL: <http://digital-editor.blogspot.de/> (besucht am 16.07.2012).
  - *T-PEN Version 2.0*. Blog-Eintrag. 5. Mai 2012. URL: <http://digital-editor.blogspot.de/2012/05/t-pen-version-20.html> (besucht am 16.07.2012).
- Center for History and New Media. *About Scripto*. URL: <http://scripto.org/documentation/about/> (besucht am 21.07.2012).
- *Scripto*. URL: <http://scripto.org/> (besucht am 14.07.2012).
- Centre national de la recherche scientifique. *Isidore*. URL: <http://www.rechercheisidore.fr/> (besucht am 25.06.2012).

- DBpedia. *DBpedia*. URL: <http://wiki.dbpedia.org> (besucht am 20.07.2012).
- Deutsche Digitale Bibliothek. *Deutsche Digitale Bibliothek*. URL: <http://www.deutsche-digitale-bibliothek.de/> (besucht am 25.06.2012).
- Deutsches Textarchiv. *DTAQ: Kollaborative Qualitätssicherung im Deutschen Textarchiv*. URL: <http://kaskade.dwds.de/dtaq/about> (besucht am 23.06.2012).
- Dublin Core Metadata Initiative. *Dublin Core*. URL: <http://dublincore.org/> (besucht am 21.07.2012).
- Europeana Foundation. *Europeana*. URL: <http://www.europeana.eu/portal/aboutus.html> (besucht am 25.06.2012).
- FamilySearch. *FamilySearch Indexing*. URL: <https://familysearch.org/volunteer/indexing> (besucht am 05.07.2012).
- Graf, Klaus. *Archivalia*. Blog. URL: <http://archiv.twoday.net/> (besucht am 05.06.2012).
- Guralnick, Rob und Andrea Thomer. *So You Can Think You Can Digitize*. Blog. URL: <http://soyouthinkyoucandigitize.wordpress.com/> (besucht am 13.07.2012).
- Humboldt-Universität zu Berlin. *H-Soz-u-Kult*. URL: <http://hsozkult.geschichte.hu-berlin.de/> (besucht am 05.06.2012).
- Internet Archive. *Internet Archive*. URL: <http://archive.org/> (besucht am 27.06.2012).
- JISC Collections. *JISC Collections*. URL: <http://www.jisc-collections.ac.uk/> (besucht am 25.06.2012).
- Kentucky, University of. *Carolingian Canon Law*. URL: <http://ccl.rch.uky.edu/> (besucht am 28.07.2012).
- Klauber, Laurence M. *Field notes of herpetologist Lawrence Klauber*. Hrsg. von San Diego Natural History Museum. URL: <http://fromthepage.bpoc.org/> (besucht am 30.06.2012).
- Landesarchiv Nordrhein-Westfalen. *Archive in NRW*. URL: <http://www.archive.nrw.de/> (besucht am 25.06.2012).
- Lawson, Konrad. *Crowdsourcing Transcription: FromThePage and Scripto*. Blog-Eintrag. 23. Jan. 2012. URL: <http://chronicle.com/blogs/profhacker/crowdsourcing-transcription-fromthepage-and-scripto/38028> (besucht am 14.07.2012).
- London, University College. *Digital Bentham Repository*. URL: [http://digitool-b.lib.ucl.ac.uk:8881/R&?local\\_base=BENTHAM](http://digitool-b.lib.ucl.ac.uk:8881/R&?local_base=BENTHAM) (besucht am 01.08.2012).



- LWL-Institut für westfälische Regionalgeschichte und LWL-Archivamt für Westfalen. *Digitale Westfälische Urkunden-Datenbank*. URL: [http://www.lwl.org/westfaelische-geschichte/portal/Internet/urkunden\\_datenbank/haupt.php?urlNeu=Ja](http://www.lwl.org/westfaelische-geschichte/portal/Internet/urkunden_datenbank/haupt.php?urlNeu=Ja) (besucht am 13.06.2012).
- Matthews, Zenas. *Zenas Matthews' 1846 U.S.-Mexico War Diary and Service Papers*. Hrsg. von Southwestern University Special Collections. URL: <http://beta.fromthepage.com/ZenasMatthews> (besucht am 30.06.2012).
- Monasterium-Konsortium. *Monasterium.net*. URL: <http://monasterium.net> (besucht am 13.06.2012).
- National Archives and Records Administration. *National Archives and Records Administration*. URL: [http://outreach.wikimedia.org/wiki/GLAM/Model\\_projects/Improving\\_the\\_quality\\_of\\_OCR](http://outreach.wikimedia.org/wiki/GLAM/Model_projects/Improving_the_quality_of_OCR) (besucht am 14.07.2012).
- National Endowment for the Humanities. *National Endowment for the Humanities*. URL: <http://www.neh.gov/> (besucht am 15.07.2012).
- National Library of Australia. *Australian Newspaper Digitisation Program*. URL: <http://www.nla.gov.au/ndp/> (besucht am 05.07.2012).
- Omeka Development Team. *Using Omeka for a digital edition*. Forumsbeitrag. URL: <http://omeka.org/forums/topic/using-omeka-for-a-digital-edition> (besucht am 27.07.2012).
- Online Computer Library Center. *ArchiveGrid*. URL: <http://www.archivegrid.org/> (besucht am 25.06.2012).
- OpenStreetMap Community. *OpenStreetMap*. URL: <http://www.openstreetmap.org/> (besucht am 13.07.2012).
- Page, Roderick D.M. *Setting up a local Wikisource*. URL: <http://iphylo.blogspot.de/2010/03/setting-up-local-wikisource.html> (besucht am 26.07.2012).
- ProQuest. *ProQuest*. URL: <http://www.proquest.co.uk/en-UK/> (besucht am 25.06.2012).
- Roy Rosenzweig Center for History and New Media. *Omeka Feature List*. URL: <http://omeka.org/files/docs/Featurelist.pdf> (besucht am 28.07.2012).
- *omeka.net*. URL: <http://www.omeka.net/> (besucht am 26.07.2012).
  - *Papers of the War Department (1784-1800)*. URL: <http://wardepartmentpapers.org/index.php> (besucht am 14.07.2012).

- Roy Rosenzweig Center for History and New Media. *Plugins/TeiDisplay*. URL:  
<http://omeka.org/codex/Plugins/TeiDisplay> (besucht am 28.07.2012).
- Roy Rosenzweig Center for History and New Media. URL: <http://chnm.gmu.edu/>  
(besucht am 21.07.2012).
- *Working with Dublin Core*. URL:  
[http://omeka.org/codex/Working\\_with\\_Dublin\\_Core](http://omeka.org/codex/Working_with_Dublin_Core) (besucht am 28.07.2012).
- Scholars' Lab. *Neatline*. URL: <http://neatline.org/> (besucht am 27.07.2012).
- Staatsbibliothek zu Berlin. *Kalliope*. URL:  
<http://kalliope.staatsbibliothek-berlin.de> (besucht am 25.06.2012).
- Staatsbibliothek zu Berlin, Deutsches Dokumentationszentrum für Kunstgeschichte -  
Bildarchiv Foto Marburg und Bayerische Staatsbibliothek München. *Manuscripta  
Mediaevalia*. URL: <http://www.manuscripta-mediaevalia.de/> (besucht am  
25.06.2012).
- Staatsbibliothek zu Berlin und Freie Universität Berlin. *Nachlass Friedrich Nicolai*.  
URL: <http://nicolai.host6.3-point.de/> (besucht am 02.07.2012).
- Staatsbibliothek zu Berlin und Humboldt-Universität zu Berlin. *Nachlass Franz  
Brümmer*. URL: <http://bruemmer.staatsbibliothek-berlin.de/nlbruemmer/>  
(besucht am 21.06.2012).
- *Projektbeschreibung der digitalen Edition des lexikographischen Nachlasses Franz  
Brümmers*. URL:  
<http://bruemmer.staatsbibliothek-berlin.de/nlbruemmer/projekt/> (besucht  
am 30.06.2012).
- Statistiker-WG. *Statistiker WG – pgf/TikZ-Tutorials*. URL:  
<http://www.pgf.statistiker-wg.de/> (besucht am 02.08.2012).
- SurveyMonkey. *Survey Monkey*. URL: <http://de.surveymonkey.com/> (besucht am  
24.06.2012).
- TextGrid. *TextGrid Repository*. URL: <http://www.textgridrep.de/> (besucht am  
25.06.2012).
- The European Library. *The European Library*. URL:  
<http://www.theeuropeanlibrary.org/> (besucht am 25.06.2012).
- ThomasV. *Beschreibung der Installation und Konfiguration von Proofread Page*. URL:  
[http://www.mediawiki.org/wiki/Extension:Proofread\\_Page](http://www.mediawiki.org/wiki/Extension:Proofread_Page) (besucht am  
12.07.2012).

University College London. *Digital Bentham Repository*. URL: [http://digitool-b.lib.ucl.ac.uk:8881/R/?func=collections&collection\\_id=1867](http://digitool-b.lib.ucl.ac.uk:8881/R/?func=collections&collection_id=1867) (besucht am 22.07.2012).

– *Help: Wiki Overview*. URL:

[http://www.transcribe-bentham.da.ulcc.ac.uk/td/Help:Wiki\\_Overview](http://www.transcribe-bentham.da.ulcc.ac.uk/td/Help:Wiki_Overview) (besucht am 28.07.2012).

– *Installationsanweisungen für „Bentham Transcription Desk“*. URL:

<http://code.google.com/p/tb-transcription-desk/wiki/Instructions> (besucht am 28.07.2012).

– *Transcribe Bentham*. URL: <http://www.ucl.ac.uk/Bentham-Project/> (besucht am 05.07.2012).

University of London Computer Center. *University of London Computer Center*. URL:

<http://www.ulcc.ac.uk/> (besucht am 22.07.2012).

Victoria, Public Records Office. *PROV Transcription Pilot*. URL:

<http://prov.versi.edu.au/> (besucht am 28.07.2012).

Walsh, John, Grant Simpson und Saeed Moaddelli. *TEI Boilerplate*. URL:

<http://dcl.slis.indiana.edu/teibp/> (besucht am 28.07.2012).

Wikimedia Foundation. *Installationsanforderungen für MediaWiki*. URL:

[http://www.mediawiki.org/wiki/Manual:Installation\\_requirements](http://www.mediawiki.org/wiki/Manual:Installation_requirements) (besucht am 12.07.2012).

– *Wikimedia*. URL: <http://www.wikimedia.org/> (besucht am 25.06.2012).

– *Wikimedia Commons*. URL:

<http://commons.wikimedia.org/wiki/Hauptseite?uselang=de> (besucht am 13.07.2012).

– *Wikisource*. URL: [http://en.wikisource.org/wiki/Main\\_Page](http://en.wikisource.org/wiki/Main_Page) (besucht am 21.06.2012).

Wikipedia Community. *GLAM Model projects*. URL: [http://outreach.wikimedia.org/wiki/GLAM/Model\\_projects/Improving\\_the\\_quality\\_of\\_OCR](http://outreach.wikimedia.org/wiki/GLAM/Model_projects/Improving_the_quality_of_OCR) (besucht am 14.07.2012).

– *Homepage von Wikipedia*. URL: <http://www.wikipedia.org/> (besucht am 05.07.2012).

– *Über Wikisource*. URL:

[http://de.wikisource.org/wiki/Wikisource:%C3%9Cber\\_Wikisource](http://de.wikisource.org/wiki/Wikisource:%C3%9Cber_Wikisource) (besucht am 03.07.2012).

Wikipedia Community. *Wikipedia-Artikel zu Wikisource*. URL:

<http://de.wikipedia.org/wiki/Wikisource> (besucht am 03.07.2012).

Wikisource Community. *Field Notes of Junius Henderson*. URL:

[http://en.wikisource.org/wiki/Field\\_Notes\\_of\\_Junius\\_Henderson](http://en.wikisource.org/wiki/Field_Notes_of_Junius_Henderson) (besucht am 12.07.2012).

– *Korrekturen des Monats*. URL:

<http://de.wikisource.org/wiki/Vorlage:Reviewtext> (besucht am 12.07.2012).

– *Portal:Taxonomy*. URL: <http://en.wikisource.org/wiki/Portal:Taxonomy> (besucht am 12.07.2012).

– *Wikisource:Metadaten*. URL:

<http://de.wikisource.org/wiki/Wikisource:Metadaten> (besucht am 26.07.2012).

– *Wikisource:Neue Seite anlegen*. URL:

[http://de.wikisource.org/wiki/Wikisource:Neue\\_Seite\\_anlegen](http://de.wikisource.org/wiki/Wikisource:Neue_Seite_anlegen) (besucht am 13.07.2012).

World Wide Web Consortium (W3C). *World Wide Web Consortium*. URL:

<http://www.w3.org/> (besucht am 10.07.2012).

Zentrales Verzeichnis Digitalisierter Drucke. *Zentralen Verzeichnisses Digitalisierter Drucke*. URL: <http://www.zvdd.de/> (besucht am 26.07.2012).

## Software

Ancestry.com. *Ancestry World Archives Project*. URL:

<http://community.ancestry.co.uk/awap> (besucht am 16.06.2012).

Apache Software Foundation. *Apache OpenOffice*. URL:

<http://www.openoffice.org/de/> (besucht am 16.06.2012).

Atlas of Living Australia, Gaia Resources. *FieldData*. URL:

<http://www.ala.org.au/get-involved/citizen-science/fielddata-software/> (besucht am 16.06.2012).

Brumfield, Ben W. *FromThePage*. URL: <http://beta.fromthepage.com/> (besucht am 16.06.2012).

California Digital Library. *XTF*. URL: <http://xtf.cdlib.org/> (besucht am 16.06.2012).

Center for History and New Media. *Scripto*. URL: <http://scripto.org/> (besucht am 16.06.2012).

- Citizen Cyberscience Center und Open Knowledge Foundation. *PyBossa*. URL:  
<http://pybossa.com/> (besucht am 16.06.2012).
- Crowder, Ben. *Unbindery*. URL:  
<http://bencrowder.net/blog/category/unbindery/> (besucht am 16.06.2012).
- Drupal Community. *Drupal*. URL: <http://www.drupal.de/> (besucht am 16.06.2012).
- Family Search. *Family Search Indexing*. URL:  
<https://indexing.familysearch.org/newuser/nuhome.jsf?3.9.6> (besucht am 16.06.2012).
- Hofstra University, John Bryant et al. *TextLab*. URL:  
<http://mel.hofstra.edu/textlab.html> (besucht am 16.06.2012).
- Huygens Institute for the History of the Netherlands. *eLaborate*. URL:  
<https://www.elaborate.huygens.knaw.nl/> (besucht am 17.06.2012).
- Impact Center of Competence. *Alto Editor*. URL: <http://www.digitisation.eu/>  
(besucht am 16.06.2012).
- inspyde. *Wordpress*. URL: <http://wpde.org/> (besucht am 16.06.2012).
- Integrating Digital Papyrology. *Son of Suda On-line*. URL:  
<http://papyri.github.com/documentation/README.html> (besucht am 16.06.2012).
- Monasterium.net und Historisch-Kulturwissenschaftliche Informationsverarbeitung der Universität Köln. *MOM-CA*. URL: <http://www.mom-wiki.uni-koeln.de/> (besucht am 17.06.2012).
- *VdU-Editor*. URL: <http://www.mom-wiki.uni-koeln.de/> (besucht am 16.06.2012).
- New York Public Library. *What's on the Menu?* URL: <http://menus.nypl.org/>  
(besucht am 16.06.2012).
- Pape, Sebastian und Stefan Achler. *TEICHI*. URL: <http://www.teichi.org/> (besucht am 16.06.2012).
- Roy Rosenzweig Center for History and New Media. *Omeka*. URL: <http://omeka.org/>  
(besucht am 16.06.2012).
- Saint Louis University, Center for Digital Theology. *T-PEN*. 2012. URL:  
<http://digital-editor.blogspot.de/> (besucht am 16.06.2012).
- Staatsbibliothek zu Berlin, Humboldt-Universität zu Berlin und 3pc GmbH Neue Kommunikation. *Refine!*
- Stadarchief Leuven und Universität Köln (Historisch-Kulturwissenschaftliche Informationsverarbeitung). *Itineranova-Editor*. URL:  
<http://www.mom-wiki.uni-koeln.de/> (besucht am 16.06.2012).

- Syncro Soft. *oXygen XML-Editor*. URL: <http://www.oxygenxml.com/> (besucht am 16.06.2012).
- TextGrid Forschungsverbund. *Textgrid*. URL: <http://www.textgrid.de/> (besucht am 17.06.2012).
- The University of Iowa Libraries. *Civil War Diaries and Letters Transcription Project*. URL: <http://digital.lib.uiowa.edu/cwd/transcripts.html> (besucht am 16.06.2012).
- ThomasV. *Proofread Page*. URL: [http://www.mediawiki.org/wiki/Extension:Proofread\\_Page](http://www.mediawiki.org/wiki/Extension:Proofread_Page) (besucht am 03.07.2012).
- United States Geological Survey. *North American Bird Phenology Program*. URL: <http://www.pwrc.usgs.gov/bpp/> (besucht am 16.06.2012).
- United States National Archives. *National Archives Transcription Pilot Project*. URL: <http://transcribe.archives.gov/> (besucht am 16.06.2012).
- University of London Computer Centre, UCL Bentham Project. *Bentham Transcription Desk*. 2012. URL: <http://www.ucl.ac.uk/transcribe-bentham/> (besucht am 16.06.2012).
- Wikimedia Foundation. *MediaWiki*. URL: [http://www.mediawiki.org/wiki/Main\\_Page](http://www.mediawiki.org/wiki/Main_Page) (besucht am 03.07.2012).
- *Wikisource*. URL: [http://en.wikisource.org/wiki/Main\\_Page](http://en.wikisource.org/wiki/Main_Page) (besucht am 16.06.2012).
- Zend Technologies. *Zend Framework*. URL: <http://framework.zend.com/> (besucht am 21.07.2012).
- Zooniverse. *Scribe*. URL: <https://github.com/zooniverse/Scribe> (besucht am 16.06.2012).

## E-Mail-Listen, E-Mails

- Archivschule Marburg. *Archivliste*. [archivliste@lists.uni-marburg.de](mailto:archivliste@lists.uni-marburg.de). E-Mail-Liste.
- Association of European Research Libraries. *Discussion list of European Manuscript Librarians*. E-Mail-Liste.
- Berlin-Brandenburgische Akademie der Wissenschaften. *Digital Humanities Liste Berlin*. [dhberlin@bbaw.de](mailto:dhberlin@bbaw.de). E-Mail-Liste.
- Brumfield, Ben W. *Re: FromThePage Questions*. E-Mail. 18. Juli 2012.
- *Re: Master Thesis*. E-Mail. 19. Juli 2012.

Causer, Tim. *RE: Questions on Bentham Transcription Desk*. E-Mail. 31. Juli 2012.

Text Encoding Initiative. *Text Encoding Initiative Public Discussion List*.

tei-l@listserv.brown.edu. E-Mail-Liste.

Thomas, Christian. *Re: Fragen zu Refine!* E-Mail. 24. Juli 2012.

# 10 Anhang

## 10.1 Umfrage: Textantworten

### 10.1.1 Art der Quellen

#### Aktuelle Editionsprojekte

| Um was für eine Art von Quellen handelt es sich bei Ihrer digitalen Edition?  | What kind of source material do you publish?   |
|---|--|
| Gemeinfreie deutschsprachige Quellentexte, Bsp. handschriftlicher Quellen finden sich im Handschriftenverzeichnis auf Wikisource ( <a href="http://de.wikisource.org/wiki/Handschriftenverzeichnis">http://de.wikisource.org/wiki/Handschriftenverzeichnis</a> )  | Early Modern English   |
| autograph   | Modern books   |
| Archivgut des 12. bis 21. Jh.   | Correspondence and a diary.  |
| Alles mögliche. Sitzungsprotokolle, MA-Urkunden, antike Münzen und Texte, Briefe  | Transcriptions of medieval manuscript materials  |
| vorwiegend Briefe, auch sonstige Manuskripte und Bildmaterial   | Manuscripts and reports  |
| Zivilstandsregister, Kirchenbuchduplikate   | Printed books 1475-1800; modern (mostly 19th-cent) editions of Middle English works; [other]         |
| Erweiterter Nachlass des Medizinnobelpreisträgers Emil von Behring (1854-1917) + Behringwerkkorrespondenz (1904-1917)   | Archival materials and born digital work   |
| Kirchenbuchduplikate und Zivilstandsregister  | Historical documents, literary documents, journals, teaching materials, dictionaries, research data. |
| handschriftliche autobiografische Aufzeichnungen aus dem 16./17. Jahrhundert  | Primarily letters, novels, short stories, and essays: edited (not only transcribed).                 |
| Die Urkunden der Stadt Minden ab 1232, späterhin auch Amtsbücher (Bürgerbücher / Schosslisten u.ä.) sollen transkribiert / registriert und in der Archivierungssoftware AUGI-AS erfasst werden; gleichzeitig bietet sich die Präsentation im Netz an, beispielsweise der Urkunden in der Westfälischen Urkundendatenbank. | 17th Century English Literary Manuscripts  |
| Deutsche Drucke aus dem 17. Jahrhundert, gebunden und ungebunden  | manuscripts  |

Aktuelle Projekte: Art der Quellen

(Tabelle wird fortgesetzt)



| Um was für eine Art von Quellen handelt es sich bei Ihrer digitalen Edition?                               | What kind of source material do you publish?  |
|--|---|
| Bücher, Zeitschriften  | Correspondence  |
| A) Handschrift Anfang 19Jh. B) PDF-Scans von maschinenschriftlichen Berichten                              | printed books, manuscripts, born-digital editions   |
| Wir betreuen verschiedenste Projekte   | Historical manuscripts from 18th and 19th centuries   |
| Handschriften, Drucke, Archivalien   | Early modern printed texts by women   |
| hist. Drucke   | Various: works by Victorian women writers, works by authors from Indiana, proceedings from the Indiana legislature from the 19th century...                         |
| Traktat aus dem französischen 18. Jahrhundert.   | All sorts. We work with researchers and research projects to transform and publish their digital editions.  |
| Amts-Tagebuch von 1808   | Journals, published writings, letters   |
| Briefe   | historical theatre reviews; parliamentary debates; historical literary periodicals  |
| Tagebücher, Briefen, Künstlernoteizen, mittelalterliche Urkunden   | archival documents  |
| Texte  | diplomatic correspondences; medieval charters   |
| 18. Jahrhundert, franz.  | not sure what you want me to say here. Are you talking about just digital publishing or also traditional publishing? This is cambridge university we do many things |
| Handschriften aus dem 17 - 18 Jh/Schlesien   | medieval source material  |
| Mittelalterliche Handschriften   | Diaries and letters   |
| Notizbücher aus dem 19. Jh.  | Archival materials, such as correspondence  |
| Briefe   | digital scholarly editions, newsletters   |
| Briefe Handexemplare von Büchern   | medieval Latin manuscripts  |
| Wissenschaftshistorische Handschriften und politische Korrespondenz, spätes 16. Jhdt. bis frühes 20. Jhdt. | old printed books, manuscripts, incunabula  |
| Transkription - nicht Edition - von Briefen aus dem Nachlass Franz Brümmer, geplant sind weitere Vorhaben. | manuscripts   |
| erst geplant! handschriftliche Quellen aus 350 Jahren; Fotografien, etc.                                   | Manuscript and printing : arts and sciences, mainly literature books, deeds and minutes. (early modern period)  |
| Urkunden   | Family diaries and letters.   |
| Historische Rechtstexte  | Family diaries and letters  |
| mittelalterliche Quellen (Urkunden) Fotos (Zeitgeschichte) Zeugnisse/Dokumente aller Art (Zeitgeschichte)  | Stendhal's manuscripts, 19th century  |
| Korrespondenzen und reichsständische Protokolle des 17. und 18. Jahrhunderts                               | I think about the collaborative curation of EEBO TCP texts, i.e. texts that were commercially transcribed but need more scholarly attention                         |
| Broschüren   | Genealogical materials  |

Aktuelle Projekte: Art der Quellen

(Tabelle wird fortgesetzt)

| Um was für eine Art von Quellen handelt es sich bei Ihrer digitalen Edition? | What kind of source material do you publish?   |
|--|--|
| mittelalterliche Urkunden  | Genealogical records, for example church records, legal records  |
| Akten  | Classical manuscripts, medieval charters/registries and (literary) texts, modern (mainly literary) manuscript. |
| maschinenschriftliche Unterlagen   | manuscripts  |
|  | mss, early printed books   |

Aktuelle Projekte: Art der Quellen

## Geplante Editionsprojekte

| Um was für eine Art von Quellen handelt es sich bei Ihrer möglichen Edition? | What kind of sources do you want to publish in your future digital edition? |
|--|---|
| Ratsprotokolle, Kriegstagebücher   | archival material such as protokolls, reports, documentet history           |
| Akten, Briefe  | medieval sources  |
| Tagebücher von Forschungsreisen  | Unpublished manuscripts and correspondence.                                 |
| Urkunden   |   |
| kundenseitige Quellen, i.d.R. 20. Jh.  |   |
| Urkunden, Akten  |   |
| Briefkopierbücher  |   |
| Diarium  |   |
| Handschriften, Nachlässe, Bücher, Karten ...                                 |   |
| Akten  |   |
| Archivalien der frühen Neuzeit: Verwaltung, Wirtschaftsschriftgut            |   |
| Stadtbücher, Chroniken (Handschriften)                                       |   |
| Karteikarten   |   |
| Briefe Handexemplare   |   |
| Protokollbände (v.a. Magistrat)  |   |
| Korrespondenz, Briefe, Aufsätze  |   |

Geplante Projekte: Art der Quellen

## 10.1.2 Eingesetzte Editoren

### Aktuelle Projekte

| Mit welchen Editionswerkzeugen / Editoren arbeiten Sie?            | Which tools / editors do you use?  |
|--|--|
| OxygenXML  | various; at the moment, wikibooks is our main platform   |
| noch in Vorbereitung   | Amaya Bluefish   |
| Verkartung in Excel  | oXygen XML Editor  |
| Datenbank Allegro HANS   | T-PEN for a transcription tool and editing environment. Oxygen, etc. for XML manipulation. SharedCanvas for display.   |
| noch nicht zu beantworten  | notetab pro  |
| a) Word b) Adobe Professional Oxygen XML-Editor                    | To do what? We outsource transcription, edit with text editors (e.g. TextPad, EditPad, EmEditor), parse with XML-lint and NSGMLS, process with Perl and XSLT (often via Windows batch files or Unix shell scripts), display in browsers. |
| OxygenXML  | oXygen, Voyant, Drupal, Ruby on Rails  |
| Hauptsächlich oXygen, daneben editMOM                              | Oxygen, TEI, Saxon, eXist, Image Markup Tool, Subversion, GIMP, Inkscape.  |
| bevorzugt XML-Editoren (oXygen), zukünftig womöglich auch TextGrid | oXygen, TEI-XML, XSLT+CSS2, WordPerfect, InDesign. It depends upon the intended output; WordPerfect + InDesign are for printed books, the others for Web output. We have also created a simultaneous Web + print release.                |
| oXygen   | TEI, XSLT, CSS   |
| Oxygen, XML Copy Editor.   | Omeka Scripto Oxygen   |
| LAGIS Hessen   | oXygen, Scripto, Omeka; CodeLobster  |
| Textgrid, Smultron, Oxygen   | BBEdit, OxyGen   |
| Oxygen, MOM-CA, OpenOffice   | oXgen and XMetaL   |
| Textgrid   | Emacs, Oxygen, BBEEdit; we will shortly begin using a web interface as well for public contributors  |
| collatio   | XTF for publishing text, Oxygen for encoding text, Mets Navigator for publishing images.   |
| word   | All sorts. oXygen is our preferred XML editor. We do a lot of bespoke XSLT. Drupal is our preferred CMS (which we sometimes import TEI into). JQuery is our preferred javascript library. Solr, etc. also useful.                        |
| Oxygen, TextGridLab  | OCR software (Abby FineReader); basic text editors (TextWrangler, TextMate); specialized text editors (Oxygen XML editor)  |
| Refine!  | currently none   |
| Oxygen   | Oxygen as XML editor; Diple (our own tool) for XSL transformation and publication  |
| noch nicht entschieden   | different projects use different tools. Oxygen, Word, excel, bespoke Gui editors, bespoke web based slightly less gui editors  |

Aktuelle Projekte: Eingesetzte Editoren

*(Tabelle wird fortgesetzt)*

| Mit welchen Editionswerkzeugen / Editoren arbeiten Sie?       | Which tools / editors do you use?  |
|---|--|
| jEdit /oXygen   | TEI, XML, Oxygen   |
| traditionelle Editions methen der historischen Wissenschaften | XML/ TEI   |
| eigene Datenbank, WordPerfect                                 | oxygen   |
| TextGrid 1.0 und 2.0 Laboratory                               | EditTEI, online input form developed by IRHT ( <a href="http://form-tei.irht.cnrs.fr/">http://form-tei.irht.cnrs.fr/</a> )   |
|   | FromThePage  |
|   | FromThePage  |
|   | XML editors (Morphon and Oxygen) using a dedicated DTD (CLELIA) transposing to TEI (to share) and to AID (to publish with InDesign)  |
|   | oXygen. There are no suitable collaborative curation tools yet, and their development is under discussion by Project Bamboo  |
|   | Flickr, Excel, CONTENTdm   |
|   | MsWord, Mediawiki  |
|   | eLaborate ( <a href="https://www.elaborate.huygens.knaw.nl/">https://www.elaborate.huygens.knaw.nl/</a> ), CTE, oXygen, various off the shelf text editors, and various custom made scripts/software. The following questions I will answer with respect to the first (eLaborate). |
|   | oxygen   |
|   | oxygen   |
|   | Transcription tools, such as Transcript and Wikipedia  |

Aktuelle Projekte: Eingesetzte Editoren

## 10.1.3 Konfrontation des Bearbeiters mit Quelltext

### Aktuelle Projekte

---

**Halten Sie eine Transkriptionsfunktion ohne Konfrontation des Bearbeiters mit dem Quelltext (z. B. TEI, HTML) für sinnvoll?**

---

Ich denke seit vielen Jahren über diese Frage nach und bin noch immer nicht zu einem eindeutigen Schluss gekommen

komplexere, verschachtelte Sachverhalte erfordern genaue Kontrolle und gegebenenfalls händische Eingriffe. Wer das dann nicht gewohnt ist, kann leicht fehlgehen.

**Do you think it is sensible to expose the user to source code (e. g. TEI, HTML) in the transcription module?**

---

Yes and no. In principle there is no objection to users getting to representational code as TEI/HTML. Some of our senior researchers think that working with XML is an analytic form of engaging with text structure, others think it's an error prone nuisance and want an editor that hides away XML specifics and let's them model text in more natural/intuitive ways. (NB 1: I would call the software code of eLaborate the source code, not the TEI/XML/HTML representation of an edition; NB 2: For us the TEI/XML/HTML surface is just a representational state of the text that can be modeled in various ways).

Some users are uncomfortable using mark-up (in our case wiki-links and basic HTML) and transcribe to plaintext. Even this is valuable, as other users follow behind to proofread and add mark-up.

It entirely depends on who are the users we are working with. A scientist who has to concentrate on the scientific validation of the transcription and generally doesn't get used to source code should not to be troubled with the source code. If the users are used to it, they may be exposed to the source code if it may clearly help them.

BTW - in the previous question - T-PEN allows individual transcription, group transcription, and crowd-sourcing. So - not either/or. We use Oxygen exclusively to do markup, and we handle version control using Subversion. I don't believe in tools that purport to hide the XML from the user; they're confusing and a big waste of time. Nobody finds XML difficult.

This depends on the user and the type of encoding being done. I think it is in any case good to have this option, but also to have the option to conceal the code (similar to the way a wiki offers Rich Text view and markup view)

In our case, transcribers also encode, so of course it makes sense for us. In the case of crowdsourcing I don't think it makes sense. Yes, training in TEI to feel comfortable enough to use TEI tags can be done in a day. If researchers spend years learning dead languages, spending a day or two being comfortable with markup is hardly a burden.

---

Aktuelle Projekte: Konfrontation des Benutzers mit Quelltext

## Geplante Editionsprojekte

Wäre für das geplante Editionsprojekt eine Transkriptionsfunktion ohne Konfrontation des Bearbeiters mit dem Quelltext (z. B. TEI, HTML) sinnvoll?

Do you think it would be sensible to expose the user to source code (e. g. TEI, HTML) in the transcription module?

Frage ist mir unklar. Wenn die Qualität so gut ist, braucht's keine Überprüfung am Quelltext??

You have to use source code like HTML to make a good digital edition, but for the user it can be very irritating to type codes all the time. So you have to simplify the method. For example a word between () activates the HTML-code for italic.

?? wahrscheinlich nein

Es sollte ihm freigestellt sein, den Quelltext sichtbar zu machen.

Geplante Projekte: Konfrontation des Benutzers mit Quelltext

## 10.1.4 Anreize zur Mitarbeit an Transkriptionsprojekten

### Aktuelle Projekte

| Um welche Anreize zur Mitarbeit handelt es sich bei Ihrem Transkriptionsprojekt?   | Which incentives for participation do you create in your transcription project?                         |
|--|---|
| Nennung als Urheber  | We list contributors on project sites.  |
| Nennung als offizieller Mitarbeiter oder Textbearbeiter; Praktikantenstelle in einem reizvollen Digital Humanities Projekt | List of transcribers by number of edits, number of new transcriptions, and number of new indexes added. |
| Statistik z. B. über gemeldete/behobene Inkonsistenzen in den TEI-XML-Files Transkriptionen)                               | A leader board of active transcribers, a progress bar, and frequent personal feedback                   |
| Dabei sein!  |   |

Aktuelle Projekte: Anreize zur Mitarbeit

### Geplante Projekte

| Welche Anreize zur Mitarbeit sollten in Ihrem geplanten Transkriptionsprojekt gegeben werden?  |
|--|
| Publikationen z. B. in den Reihen unseres Archivs  |
| Forschungsdiskussion (Fragen und Kommunikation) zu den Editionen und weiterreichenden Themen; Einblicke in die Arbeit in digitalen Projekten; Erweiterung und Verbesserung |
| Anerkennung in passender Weise, Namensnennung, Würdigung durch Einrichtung (Blumenstrauß?)   |

Geplante Projekte: Anreize zur Mitarbeit

## 10.1.5 Technische Hilfsmittel zur Qualitätssicherung

### Aktuelle Editionsprojekte

| Welche technischen Hilfsmittel werden zur Qualitätssicherung eingesetzt?                           | Which tools do you employ for quality assurance?  |
|--|---|
| Durchsicht durch einen Zweitkorrektor  | Primarily XSLT and human proofing.  |
| noch nicht zu beantworten  | Scripto   |
| keine  | Validation, proofreading, expert internal review  |
| keine  | We have lead editors who check transcriptions - no special tools used.  |
| Kontrollierte Vokabularien   | Custom XSLT   |
| Schemata, Kontrolllisten für die Verwendung einzelner Elemente, Attribute                          | Manual proofreading. (Other staff proofread the work of the person who transcribed.)  |
| <a href="http://www.deutschestextarchiv.de/dtaq/about/">www.deutschestextarchiv.de/dtaq/about/</a> | Nothing special. Manual verification with an automatic help on Oxygen Editor (regexp search and Xpath query for formatted or tagged data).  |
| Oxygen, Vergleich von Transkrip  | Wiki version control enables correction in multiple passes. Comments allow users to report problems or ask questions about difficult reading.   |
| Feedback an Moderatoren  | Guidelines for transcription (project specific), peer review at transcription level, and often we have a 'quartermaster' per edition project: a person that is first line 'knowledge base' for anyone related to the project on what standards and practices were decided on for that specific project. |
| keine  |   |
| Indices, Normdaten   |   |
| keine. muss ja gelesen werden  |   |

Aktuelle Projekte: Technische Hilfsmittel zur Qualitätssicherung



## Geplante Editionsprojekte

| Welche technischen Hilfsmittel könnte man zur Qualitätssicherung einsetzen?           | Which tools do you employ for quality assurance?  |
|---|---|
| Data-/Text Mining Werkzeuge   | Manual checking   |
| bislang sind uns für den Übergang von den Handschriften zum Digitalisat keine bekannt | - words are shown in red when a symbol for HTML-code is typed wrong. - moderators (humans) who check the transcriptions |
| keine Ahnung, Qualitätssicherung durch öffentliche Diskussion der Einträge            |   |
| Abgleich von Daten, Plausibilität , Namenverzeichnisse ???                            |   |

Geplante Projekte: Technische Hilfsmittel zur Qualitätssicherung

## 10.1.6 Funktion zur text- oder sachkritischen Kommentierung

### Aktuelle Editionsprojekte

| Besitzt Ihr Editionswerkzeug eine Funktion für die textkritische oder sachkritische Kommentierung der Texte?  | Does your tool come with a function for textual or thematic commentary?  |
|---|--|
| vorherige Frage ein bisschen unverständlich. Diese Frage: Da XML und TEI dies zulässt, ja.  | well, using TEI gives a plenty of possibilities for this ...   |
| unbedingt erforderlich!   | Users can create text place specific annotation, or general commentary linked to the text.   |
| Excel bietet aber die Möglichkeit in einer eigenen Spalte bemerkungen einzutragen.  | That's the desideratum, but the tool doesn't exist   |
| In Allegro HANS können längere Texte integriert werden, in meinem Fall Regesten für die einzelnen Dokumente, die ich erstelle                                   | Users may add notes to pages. There is an extensive toolset for indexing and commenting on the subjects mentioned within the text.   |
| Da es sich um einen reinen Editor handelt (an dieser Stelle keine Interaktion mit anderen) gibt es die natürlich, weil es sich um eine Kodierungsfrage handelt. | Not with Edit-TEI but with the online input form used for deeds and minutes since we want the transcribers scientists/students to be able to associate their analysis to the corresponding part of the transcription.  |
| nur für Admins  | It's not built-in, or at least we're not using a built-in tool. We add contextual encoding as we go.   |
|   | it depends if XML commentary can be considered as a function   |
|   | Handled via internal structures, which are user-shaped   |
|   | Sorry but I do not know.   |
|   | In the form of plain-text notes.   |
|   | Except for unusual cases, we do not transcribe: we pay others to transcribe for us, using whatever proprietary tools they like.  |
|   | Depends on what you mean. XML is flexible.   |
|   | I don't understand what this means. A text may come with a commentary; I don't see why a tool would have one.  |
|   | What is „function“, here? I recognize that your survey is interested in things like Scripto and T-PEN; we use WordPerfect to transcribe documents, so there is capacity for commentary but not in the way I think you ask.   |
|   | Our method of adding comments is not sophisticated   |
|   | Not at the moment, but this is something we may add in the future. (Note that we use several tools and that the community transcription part of our work is still being developed, so my answers to some of these questions are a bit hypothetical or cover several different situations.) |
|   | We don't have just one tool, we do bespoke tools for different projects. But we have had some that allow annotation of TEI files.  |

Geplante Projekte: Technische Hilfsmittel zur Qualitätssicherung

## Geplante Editionsprojekte

---

Besitzt Ihr Editionswerkzeug eine Funktion für die textkritische oder sachkritische Kommentierung der Texte?

---

Does your tool come with a function for textual or thematic commentary?

---

allerdings nur von fachwiss. Seite

I like a function where you can select a word and add commentary by rightclicking the mouse or clicking a button

---

Geplante Projekte: Technische Hilfsmittel zur Qualitätssicherung

## 10.1.7 Speicherung in TEI

### Aktuelle Projekte

| Halten Sie eine Speicherung der Editionstexte in einem XML-basierten Format wie TEI für unumgänglich?   | Do you think that there is no alternative to saving the transcriptions in an XML-based format like TEI?  |
|---|--|
| Einerseits ja, andererseits kann ich mir auch vorstellen, eine markup Auszeichnung zu verwenden. (Überlappende Strukturen)  | There are many alternatives, though if one wishes proper interoperability and shared use within the community, TEI makes most sense  |
| - keine Ahnung - dies kann unsere IT-Abteilung beantworten  | Not sure   |
| Standards sind sinnvoll und nützlich  | Plain text. Though we really don't care. The individual line-by-line transcriptions become the bodies of annotations (Open Annotation annotations) that target a region of interest on a canvas. We don't care how the transcriptions are stored - TEI, plain text, rtf, doesn't matter.   |
| Ja, weil das Community-bildend wirkt/wirken sollte, weil Interoperabilität gefördert wird. Nein, weil diese Eigenschaft potentiell auch in anderen Sprechern realisiert werden könnte (LMNL, nicht-hierarchische), TEI war nur eher da. ;-) | There are always alternatives. Our base format currently is still SGML; and we have certainly used database (flat and relational) as well in the past.   |
| zumindest nicht grundsätzlich, aber für unser Projekt ist TEI in der Tat unumgänglich.  | <p>There are always alternatives, but not necessarily good ones. Increasingly XML is becoming an archival format - what is actually used online is a database of some sort. Perhaps we will see database archiving standards that replace the need for XML.</p> <p>There are lots of alternatives. Some data is better stored in relational databases, for instance. But generally speaking, for digital editions, XML is the best choice.</p> <p>Indeed, at this stage we convert the edited texts from WordPerfect to TEI! Beware assumptions :)</p> <p>Obviously, there are alternatives, but I think there is no good reason not to be using TEI/XML, given the benefits of interoperability.</p> <p>I do think there are alternatives, but TEI is a good standard to use.</p> <p>For some of our community data capture, we save the data first to a database, and then generate XML later on. As long as the right fields are captured, there is nothing special about XML (unless one is working with mixed content).</p> <p>That's an odd way to phrase the question, including the negative. Of course there are alternatives to XML-based formats like TEI. Ideally transcriptions could be created using some simple format (even Word documents) and then converted automatically to TEI. The TEI could then be the master record format (perhaps that's actually what you are asking). I've been working with TEI for over ten years, and the longer I work with it the more I'm convinced that TEI encoding should be automated as much as possible, with initial work being done in other formats.</p> <p>Yes and No. There are alternatives, but they usually aren't as rich. We convert from DOCX to TEI regularly but then improve the TEI through semi-automatic means (up-conversion/up-scaling through programmatic XSLT). TEI is by and large the best format for storing a transcription if you really care about the editorial recognition of textual phenomena.</p> |

Aktuelle Projekte: Speicherung in TEI

(Tabelle wird fortgesetzt)

---

**Halten Sie eine Speicherung der Editionstexte in einem XML-basierten Format wie TEI für unumgänglich?**

---

**Do you think that there is no alternative to saving the transcriptions in an XML-based format like TEI?**

---

Question is unclear. It is difficult to imagine an alternative standard, given the extensive work done on TEI.

No, I'd be fine with using another format. I'm not a coding expert, so I'm using this format primarily because our IT advisers recommended it.

Database MySQL. Yet, TEI seems to me the best solution for textual transcriptions if there is structuration part in the project, not only a full-text and metadata research objective.

Lots of alternatives exist. FromThePage uses wiki-links backed by a RDBMS. Wikisource also doesn't use XML, and attributed-range systems like eComma may become useful for annotating transcriptions.

Neither yes or no. TEI is an excellent means to share and redistribute, however, in France, there is for the moment, no framework to centralize and preserve digital data. When dealing with unsharable date (for legal reasons in particular), TEI can become an impediment to transcription / transcribers.

That's not a a yes/no question. IN practice, TEI is the only game in town. But others are imaginable. People have talked about modeling a simple version TEI in html5

Don't understand the question

There are lots of alternatives (ePub, XHTML, JSON) and quite frankly I don't think TEI-XML has any chance to become an industry standard like ePub. At best TEI-XML could be a useful modeling tool for text, but I sincerely doubt it will be the start and end all text representational solution. Foremost also as TEI has inbuilt and principle inadequacies for certain text structures. Which can all be solved but the solutions are highly non-xml in nature. It boils down to the fact that text is not a hierarchy and can therefore never be truly represented by only XML.

---

Aktuelle Projekte: Speicherung in TEI

## Geplante Editionsprojekte

---

Halten Sie eine Speicherung der Editionstexte möglicher Editionsprojekte in einem XML-basierten Format wie TEI für unumgänglich?

Do you think that there is no alternative to saving the transcriptions in an XML-based format like TEI?

kenne Tei nicht. sollt aber XML basiert sein.

kann ich nicht beantworten, müsste mit unsererm IT-Spezialisten besprochen werden

kann ich nicht einschätzen

Das wäre der derzeitige Standard, vermutlich werden weitere mit der Ausbreitung digitaler Editionen entwickelt.

?? Was gibt es den noch für Möglichkeiten?

---

Geplante Projekte: Speicherung in TEI

## 10.1.8 Erstellung von Indizes

### Aktuelle Projekte

| Ist die Erstellung von Indizes (z. B. für Orts- oder Personennamen) Teil Ihres Editionsprojektes? | Does your digital edition include indexes (e. g. for names of persons and places)?   |
|---|--|
| Verkartung von Registern  | sometimes  |
| Sowohl die Indices als auch die bereits erwähnten Regesten  | <p>Most of our projects include only lightly encoded transcriptions of print sources, with bibliographic headers. Our markup does not extend to word-level interpretive identification of names, places, etc.</p> <p>People, places, vessels, sources, and many others.</p> <p>It's a desideratum.</p> <p>tagging, anyway</p> <p>We include at the moment a personography. We hope in the future to include a bibliography as well.</p> <p>Yes, if you have marked these, it is trivial to produce sensible indexes, so why wouldn't you?</p> <p>We will include one for names of people and names of organizations/associations/businesses.</p> <p>Entities are indexed by tagging in TEI and we are working on a project of geolocation and association of people to places. But we don't publish indexes in the usual way.</p> <p>We are able to view all mentions of a subject and read those pages in context. Extensive functionality exists for automating this indexing. Linguistic annotation of texts would generate an index nominum, while not providing full named entity extraction but this is on our wish-list</p> |

Aktuelle Projekte: Erstellung von Indizes

### Geplante Editionsprojekte

| Sollte ein für Sie angepasstes Editionswerkzeug eine Funktion zur Erstellung von Indizes (z. B. für Orts- oder Personennamen) beinhalten? | Should your digital edition include indexes (e. g. for names of persons and places)?       |
|---|--|
| einschl. Volltextindizierung auf Dateiebene   | It should be possible to select words in the transcription and add them to a certain index |
| unbedingt!  |  |

Geplante Projekte: Erstellung von Indizes

## 10.1.9 Weitere Funktionen zur Textnavigation

### Aktuelle Projekte

| Bietet Ihre Edition weitere Funktionen zur Textnavigation?   | Does your edition provide further functions for text navigation?  |
|--|---|
| Ich beziehe mich jetzt in den Antworten auf die digitalisierten Sitzungsprotokolle der Akademie. Zeitliche Navigation. | Curation/exploration are seen as going hand in hand in our project, so the integration of curation and exploration tools is a key element.  |
| Dank der Regesten ist eine Stichwort- und kombinierte Suche der TEXTINHALTE möglich                                    | No navigation functions besides table of content, indexes, and full-text searches.  |
| Volltextrecherche  | Navigation by occurency, by structural element.   |
| Verstehe die Frage nicht   | We'll link to related documents within the edition, so that people can see the response to letter A or otherwise follow a conversation thread.  |
| ? Meinen Sie in Textstrukturen, wie Kapiteln, etc?   | Named entities are linked to new search queries across the corpus   |
| bzw. ist geplant   | very minor, though  |
|  | I do not understand what this question is asking  |
|  | Our local interface provides only search (full-text and bibliographic; fielded; proximity; boolean); and a navigable auto-extracted outline or table of contents. But other sites are free to add their own navigation features as they like. |
|  | Searches, browse trees, schedules, notes, references, links back from indexes to other mentions in text.  |
|  | Not sure what this means (further to what?)   |
|  | I don't know what „further functions“ really means, but we have searching and browsing functions.   |
|  | Yes, depending on project, a variety of methods of navigation or jQuery show/hide certain aspects.  |

Aktuelle Projekte: Weitere Funktionen zur Textnavigation

### Geplante Projekte

| Sollte ein für Ihre Bedürfnisse angepasstes Editionswerkzeug weitere Funktionen zur Textnavigation bieten?                                       | Should your digital edition provide further functions for text navigation?      |
|--|---|
| für Sprachwissenschaftler zum Beispiel: Veränderungen der Schreibweise eines Wortes nachvollziehen in einer Reihe von zeitlich unterschiedlichen | Subject navigation, e. g. contracts, legislation processes, correspondence etc. |
| Quellen; Suchfunktion nach Wörtern;  | the edition should be presented with a scan of the source                       |

Geplante Projekte: Weitere Funktionen zur Textnavigation



## 10.1.10 Verweise auf Normdaten oder Erschließungsinformationen

### Aktuelle Projekte

| Um welche Daten handelt es sich (z. B. eigene Kataloge, OPAC, Kalliope, GND o.ä.) bei Ihren Verweisen? | Which data do you integrate (e.g. your own catalogues, OPAC, Library of Congress Authorities and the like) with your digital edition?   |
|--|---|
| GND, PND ist doch quasi out. OPAC, Kataloge  | own catalogues  |
| eigene und fremdec   | own catalog   |
| Ausführliche Erläuterung in einer eigenen Datei (PDF)  | We generate TEI-like headers from MARC that we (mostly) receive from external suppliers and modify globally in house. Occasionally we write our own headers, or even write our own MARC. Some projects (but not all) are linked to from local OPAC. |
| OPAc, PND, GKD, Wikipedia, Pagel, NDB etc.   | Dictionary of National Biography (UK), Canadian Dictionary of National Biography, mythindex.com, mythologia.fr, and many others.  |
| GND (ehemalige PND), VD17  | Our own; we are also a physical archive.  |
| eigenen Katalog  | LC Authorities  |
| eigene Daten, PND  | Library of congress authorities, TGN, local controlled vocabularies.  |
| alles Genannte, plus geonames/TGN, IconClass, G(!)ND, VIAF   | It depends on the project, linking to dbpedia, LoC authorities, LGPN, DNB, our own authority files, etc.  |
| OPAC (der bestandhaltenden Institutionen, d.h. i.d.R. Bibliotheken), PND/GND, GND                      |   |
| Personennormdaten, geografische Normdaten  | Again hypothetical> The English Short Title Catalogue is an obvious control tool for TEI transcriptions of items in it  |
| Kalliope, PND  | CONTENTdm records   |
| PND, OPAC, Google Maps usw.  | mostly the OPAC data  |

Aktuelle Projekte: Verweise auf Normdaten und Erschließungsinformationen

### Geplante Editonsprojekte

| Der Verweis auf welche Datenbestände wäre für Ihr mögliches Projekt sinnvoll (z. B. eigene Kataloge, OPAC, Kalliope, PND o.ä.)? | Which data would you integrate (e.g. your own catalogues, OPAC, Library of Congress Authorities and the like) with your digital edition? |
|---|--|
| eigene Kataloge   | catalogue, Grotefend, Capelli, dictionaries  |
| PND, Geodaten, Wikipedia, Kataloge...   | LoC Authority File VIAF  |
| Normdaten der DNB, Kalliope, eigene Kataloge  |  |
| Archivportale wie z. B. archive-nrw.de  |  |

Geplante Projekte: Verweise auf Normdaten und Erschließungsinformationen

(Tabelle wird fortgesetzt)

| Der Verweis auf welche Datenbestände wäre für Ihr mögliches Projekt sinnvoll (z. B. eigene Kataloge, OPAC, Kalliope, PND o.ä.)? | Which data would you integrate (e. g. your own catalogues, OPAC, Library of Congress Authorities and the like) with your digital edition? |
|---|---|
|---|---|

PND, eigene Kataloge

Findbuch, PND.

Das hängt von sinnvollen Referenzkatalogen ab, PND wäre ein Minimum.

Stadtbuchdatenbank; Links; Slub-Opac; online-edition CDRS

eigene Datenbestände

eigene Katalog, Kalliope, PND

Geplante Projekte: Verweise auf Normdaten und Erschließungsinformationen

## 10.1.11 Datenlieferung an größere Portale

### Aktuelle Projekte

| An welche Portale liefern Sie Daten Ihrer digitalen Edition?      | Which web portals do you provide with content from your digital edition?  |
|---|---|
| naja, ist in Planung. Europeana, Deutsches Textarchiv, ddb        | wikimedia products  |
| Archive NRW.de  | Depends on what you mean by 'major'. All data is hosted on our institutional web site; much is also distributed by ProQuest; by the Philologic site at the University of Chicago; by the JISC (UK) Historic Books Portal. Data is also locally hosted at member institutions (e.g. Helsinki). First steps have been taken toward distribution via ebook portals, e.g. the Apple itunes store. |
| TextGrid Repository   |   |
| TEL, Europeana, Manuscripta Mediaevalia, zukünftig Kalliope, CERL |   |
| europeana   | OAI PMH metadata harvesters.  |
| TextGridRep   | Archivegrid   |
|   | We have participated in EU projects which provide content to Europeana.   |
|   | Europeana, Gallica, Isidore   |

Aktuelle Projekte: Datenlieferung an Portale

### Geplante Editonsprojekte

| An welche größeren Portale sollten in Ihrem möglichen Editionsprojekt Daten geliefert werden? | Which web portals do you plan to provide with content from your digital edition? |
|---|--|
| Europeana, DDB  | Wikipedia, homepage Guelders Archives  |
| z. B. DDB, Westf. Geschichte  |  |
| z. B. Europeana   |  |
| Google, usw.  |  |
| Da wir uns erst im Stadium der Vorüberlegungen befinden, noch unklar. Museumsportale          |  |

Geplante Projekte: Datenlieferung an Portale

## 10.1.12 Weitere Kommentare zu Anforderungen an Editionswerkzeuge

### Aktuelle Projekte

| Haben Sie weitere Kommentare zu Ihrem Editionsprojekt und Ihren Anforderungen an Editionswerkzeuge?  | Do you have further comments on your digital edition and requirements for transcription tools?  |
|--|---|
| Meines Wissens wird bisher wenig mit Regesten gearbeitet, bei einer Vorstellung 2011 beim Workshop in Karlsruhe (KIT) stieß unser Projekt daher auf großes Interesse | Like everyone, we are interested in pursuing collective 'crowd-sourced' transcription, revision, and correction; and perhaps even crowd-sourced markup. (And for another project—collaborative editing of a reference work, in a Wiki manner.) But we have hitherto depended mainly on contract keying and conversion firms, depending on them to find, develop, and deploy the appropriate transcription tools.  |
| OCR in Adobe Professional nahezu unbrauchbar.  | Over the years, there have been dozens of projects attempting to create online collaborative editing tools. These are completely useless IMHO, even when tailored to a specific project. Teach people XML, teach people Oxygen, and get on with transcribing and marking up text. XML is easy, Oxygen is excellent, Subversion and GitHub will handle versioning, and presentation/rendering should be written for a specific project by people who are skilled in doing it.  |
| Als problematisch empfinde ich weniger die Transkription selbst als die Visualisierung und Publikation des transkribierten Textes.                                   | The answers to the survey apply to several projects in varying degrees of advancement. Scripto is the crowdsourcing tool of choice for the moment. Transcription tools that permit multiple users to transcribe individual pieces of source material, so that these can be compared (in the manner of the NYPL menu project) would be very useful in some circumstances. We are also very interested in plug-ins to XML tools like oXygen (e.g. linking to name authority lists).   |
| nein   | We'd like to incorporate tools for crowdsourcing, particularly for transcriptions of our archival material, but we haven't figured out the best tool to use yet.<br><br>no<br><br>I am looking into designing a web interface for basic transcriptions and corrections of OCR-ed scans.<br><br>no<br><br>I have written about the subject on my blog for the last five years. <a href="http://manuscripttranscription.blogspot.com">http://manuscripttranscription.blogspot.com</a><br><br>Transcription tools have to be adapted to the knowledge owners and although there has been a tremendous amount of work done to characterize textual data (TEI), little has been done to allow digital illiterate researchers to transcribe in humanly acceptable conditions no, except to stress that I am writing about something that does not yet exist<br><br>We currently have people transcribe in word straight from the digital photos. We would like to put the images online and link the transcription to the images. |

---

Aktuelle Projekte: Weitere Kommentare zu Anforderungen an Editionswerkzeuge

(Tabelle wird fortgesetzt)

Haben Sie weitere Kommentare zu Ihrem Editionsprojekt und Ihren Anforderungen an Editionswerkzeuge?

Do you have further comments on your digital edition and requirements for transcription tools?

Transcriptiontools should be highly interoperable. We are too often creating silos for digital editions. Ideally transcription (and publication) tools should just be 'windows' on a generalized and extensible text model (not the same as TEI), so that each tool implementing the generalized parts of the model could work with any text basically. This 'frees' any digital text from the specific tool that it has been made with. This is important for exchange, research and sustainability of digital texts.

Aktuelle Projekte: Weitere Kommentare

## Geplante Projekte

Haben Sie weitere Kommentare zu Ihrem möglichen Editionsprojekt und den damit verbundenen Anforderungen an Editionswerkzeuge?

Do you have further comments on your possible project and requirements for a transcription tool?

nein

it is very important to have a transcription tool for handwritten sources, because students aren't taught anymore to read handwritten (medieval) sources

Bislang erscheinen die Editionsverfahren noch zu aufwändig, daher sehen wir derzeit noch von Digitaleditionen ab, insofern wären Werkzeuge mit überschaubaren Nutzeroberflächen, die nicht zu viel Spezialwissen voraussetzen hilfreich.

Überblick über Editinoswerkzeuge wäre gut!

Geplante Projekte: Weitere Kommentare

## 10.2 Umfrage: Multiple-Choice-Antworten

### Ergebnisse deutsche und englische Sprachversion

| In which language do you want to answer the survey? In welcher Sprache möchten Sie die Umfrage durchführen?                                   |                          | Beantwortung in Prozent |        | Anzahl Beantwortungen |    |          |
|---|--------------------------|-------------------------|--------|-----------------------|----|----------|
| Englisch  |                          | 30,6 %                  |        | 64                    |    |          |
| Deutsch   |                          | 69,4 %                  |        | 145                   |    |          |
| Arbeiten Sie in Ihrer Institution an Projekten zur Erstellung digitaler Editionen?  |                          | Beantwortung in Prozent |        | Anzahl Beantwortungen |    |          |
|   |                          | DE                      | EN     | DE                    | EN | $\Sigma$ |
| Ja  |                          | 54,2 %                  | 91,8 % | 65                    | 56 | 66,9 %   |
| Nein  |                          | 45,8 %                  | 8,2 %  | 55                    | 5  | 33,1 %   |
| Unterstützt Ihr Editionswerkzeug die Erstellung nutzergenerierter Transkriptionen?  |                          | Beantwortung in Prozent |        | Anzahl Beantwortungen |    |          |
|   |                          | DE                      | EN     | DE                    | EN | $\Sigma$ |
| Aktuelle Projekte   | Ja                       | 58,3 %                  | 60 %   | 14                    | 21 | 50,7 %   |
|   | Nein                     | 41,7 %                  | 40 %   | 10                    | 14 | 49,3 %   |
| Geplante Projekte   | Ja                       | 88,9 %                  | 100 %  | 16                    | 3  | 90,5 %   |
|   | Nein                     | 11,1 %                  | 0 %    | 2                     | 0  | 9,5 %    |
| Ist die Transkription auf einen beschränkten Nutzerkreis (z. B. von Wissenschaftlern) oder eine breitere Community (Crowdsourcing) ausgelegt? |                          | Beantwortung in Prozent |        | Anzahl Beantwortungen |    |          |
|   |                          | DE                      | EN     | DE                    | EN | $\Sigma$ |
| Aktuelle Projekte   | Beschränkter Nutzerkreis | 47,1 %                  | 83,3 % | 8                     | 15 | 65,7 %   |
|   | Breitere Community       | 52,9 %                  | 16,7 % | 9                     | 3  | 34,3 %   |
| Geplante Projekte   | Beschränkter Nutzerkreis | 68,8 %                  | 33,3 % | 11                    | 1  | 63,2 %   |
|   | Breitere Community       | 31,3 %                  | 66,7 % | 5                     | 2  | 36,9 %   |

Multiple Choice: Ergebnisse Deutsch und Englisch

(Tabelle wird fortgesetzt)

| Halten Sie eine Transkriptionsfunktion ohne Konfrontation des Bearbeiters mit dem Quelltext (z. B. TEI, HTML) für sinnvoll? |      | Beantwortung in Prozent |        | Anzahl Beantwortungen |    |          |
|---|------|-------------------------|--------|-----------------------|----|----------|
|   |      | DE                      | EN     | DE                    | EN | $\Sigma$ |
| Aktuelle Projekte   | Ja   | 62,5 %                  | 76,5 % | 10                    | 13 | 69,7 %   |
|   | Nein | 37,5 %                  | 23,5 % | 6                     | 4  | 30,3 %   |
| Geplante Projekte   | Ja   | 57,1 %                  | 66,7 % | 8                     | 2  | 58,8 %   |
|   | Nein | 42,9 %                  | 33,3 % | 6                     | 1  | 41,2 %   |
| Werden in Ihrem Transkriptionsprojekt Anreize zur Mitarbeit gegeben (z. B. Veröffentlichung der aktivsten Bearbeiter)?      |      | Beantwortung in Prozent |        | Anzahl Beantwortungen |    |          |
|   |      | DE                      | EN     | DE                    | EN | $\Sigma$ |
| Aktuelle Projekte   | Ja   | 33,3 %                  | 18,8 % | 5                     | 3  | 25,8 %   |
|   | Nein | 66,7 %                  | 81,3 % | 10                    | 13 | 74,2 %   |
| Geplante Projekte   | Ja   | 57,1 %                  | 33,3 % | 8                     | 1  | 52,9 %   |
|   | Nein | 42,9 %                  | 66,7 % | 6                     | 2  | 47,1 %   |
| Wird in Ihrem Editionsprojekt eine Qualitätssicherung (Redaktion der Transkriptionen) vorgenommen?                          |      | Beantwortung in Prozent |        | Anzahl Beantwortungen |    |          |
|   |      | DE                      | EN     | DE                    | EN | $\Sigma$ |
| Aktuelle Projekte   | Ja   | 93,3 %                  | 75,0 % | 14                    | 12 | 83,9 %   |
|   | Nein | 6,7 %                   | 25,0 % | 1                     | 4  | 16,1 %   |
| Geplante Projekte   | Ja   | 100 %                   | 100 %  | 12                    | 3  | 100,0 %  |
|   | Nein | 0,0 %                   | 0,0 %  | 0                     | 0  | 0,0 %    |
| Besitzt Ihr Editionswerkzeug eine Funktion für die textkritische oder sachkritische Kommentierung der Texte?                |      | Beantwortung in Prozent |        | Anzahl Beantwortungen |    |          |
|   |      | DE                      | EN     | DE                    | EN | $\Sigma$ |
| Aktuelle Projekte   | Ja   | 61,1 %                  | 55,6 % | 11                    | 15 | 57,8 %   |
|   | Nein | 38,9 %                  | 44,4 % | 7                     | 12 | 42,2 %   |

Multiple Choice: Ergebnisse Deutsch und Englisch

(Tabelle wird fortgesetzt)

|  |      |                                |         |                              |    |          |
|--|------|--------------------------------|---------|------------------------------|----|----------|
| Geplante Projekte  | Ja   | 92,9 %                         | 100 %   | 13                           | 3  | 94,1 %   |
|  | Nein | 7,1 %                          | 0,0 %   | 1                            | 0  | 5,9 %    |
| <hr/>  |      |                                |         |                              |    |          |
| <b>Halten Sie eine Speicherung der Editionstexte in einem XML-basierten Format wie TEI für unumgänglich?</b> |      | <b>Beantwortung in Prozent</b> |         | <b>Anzahl Beantwortungen</b> |    |          |
|  |      | DE                             | EN      | DE                           | EN | $\Sigma$ |
| Aktuelle Projekte  | Ja   | 60,0 %                         | 15,4 %  | 12                           | 4  | 34,8 %   |
|  | Nein | 40,0 %                         | 84,6 %  | 8                            | 22 | 65,2 %   |
| <hr/>  |      |                                |         |                              |    |          |
| Geplante Projekte  | Ja   | 75,0 %                         | 0,0 %   | 6                            | 0  | 66,7 %   |
|  | Nein | 25,0 %                         | 100,0 % | 2                            | 1  | 33,3 %   |
| <hr/>  |      |                                |         |                              |    |          |
| <b>Verweisen Sie in Ihrer digitalen Edition auf Normdaten oder Erschließungsinformationen?</b>               |      | <b>Beantwortung in Prozent</b> |         | <b>Anzahl Beantwortungen</b> |    |          |
|  |      | DE                             | EN      | DE                           | EN | $\Sigma$ |
| Aktuelle Projekte  | Ja   | 94,4 %                         | 60,7 %  | 17                           | 17 | 73,9 %   |
|  | Nein | 5,6 %                          | 39,3 %  | 1                            | 11 | 26,1 %   |
| <hr/>  |      |                                |         |                              |    |          |
| Geplante Projekte  | Ja   | 100,0 %                        | 100,0 % | 12                           | 2  | 100,0 %  |
|  | Nein | 0,0 %                          | 0,0 %   | 0                            | 0  | 0,0 %    |
| <hr/>  |      |                                |         |                              |    |          |
| <b>Liefern Sie Daten an größere Portale (z. B. Europeana, Deutsche Digitale Bibliothek o.ä.)?</b>            |      | <b>Beantwortung in Prozent</b> |         | <b>Anzahl Beantwortungen</b> |    |          |
|  |      | DE                             | EN      | DE                           | EN | $\Sigma$ |
| Aktuelle Projekte  | Ja   | 47,1 %                         | 28,6 %  | 8                            | 8  | 35,6 %   |
|  | Nein | 52,9 %                         | 71,4 %  | 9                            | 20 | 64,4 %   |
| <hr/>  |      |                                |         |                              |    |          |
| Geplante Projekte  | Ja   | 50,0 %                         | 100,0 % | 7                            | 2  | 56,3 %   |
|  | Nein | 50,0 %                         | 0,0 %   | 7                            | 0  | 43,7 %   |
| <hr/>  |      |                                |         |                              |    |          |
| <b>Ist die Erstellung von Indizes (z. B. für Orts- oder Personennamen) Teil Ihres Editionsprojektes?</b>     |      | <b>Beantwortung in Prozent</b> |         | <b>Anzahl Beantwortungen</b> |    |          |
|  |      | DE                             | EN      | DE                           | EN | $\Sigma$ |
| Aktuelle Projekte  | Ja   | 77,8 %                         | 71,4 %  | 14                           | 20 | 73,9 %   |

Multiple Choice: Ergebnisse Deutsch und Englisch

(Tabelle wird fortgesetzt)



|                   |   |                                |                              |    |    |          |
|-------------------|---|--------------------------------|------------------------------|----|----|----------|
|                   | Nein  | 22,2 %                         | 28,6 %                       | 4  | 8  | 26,1 %   |
| Geplante Projekte | Ja  | 100,0 %                        | 100,0 %                      | 14 | 2  | 100,0 %  |
|                   | Nein  | 0 %                            | 0,0 %                        | 0  | 0  | 0,0 %    |
|                   |   |                                |                              |    |    |          |
|                   | <b>Bietet Ihre Edition weitere Funktionen zur Textnavigation?</b>   | <b>Beantwortung in Prozent</b> | <b>Anzahl Beantwortungen</b> |    |    |          |
|                   |   | DE                             | EN                           | DE | EN | $\Sigma$ |
| Aktuelle Projekte | Ja  | 80,0 %                         | 71,4 %                       | 12 | 20 | 74,4 %   |
|                   | Nein  | 20,0 %                         | 28,6 %                       | 3  | 8  | 25,6 %   |
| Geplante Projekte | Ja  | 84,6 %                         | 100,0 %                      | 11 | 2  | 86,7 %   |
|                   | Nein  | 15,4 %                         | 0,0 %                        | 2  | 0  | 13,3 %   |
|                   |   |                                |                              |    |    |          |
|                   | <b>Wären Sie an einem weiteren Austausch zum Thema Editingswerkzeuge interessiert und möchten Ihre Kontaktdaten hinterlassen?</b> | <b>Beantwortung in Prozent</b> | <b>Anzahl Beantwortungen</b> |    |    |          |
|                   |   | DE                             | EN                           | DE | EN | $\Sigma$ |
| Aktuelle Projekte | Ja  | 76,5 %                         | 57,1 %                       | 13 | 16 | 64,4 %   |
|                   | Nein  | 23,5 %                         | 42,9 %                       | 4  | 12 | 35,6 %   |
| Geplante Projekte | Ja  | 80,0 %                         | 100,0 %                      | 8  | 2  | 83,3 %   |
|                   | Nein  | 20,0 %                         | 0,0 %                        | 2  | 0  | 16,7 %   |

Tabelle 10.24: Multiple Choice: Ergebnisse Deutsch und Englisch

## 10.3 Liste der Transkriptionswerkzeuge

Die hier (siehe Tabelle 10.25 Seite 147) dargestellte Auflistung von Transkriptionsswerkzeugen ist ein Auzug aus der von Ben W. Brumfield veröffentlichten Tabelle<sup>229</sup>. Sie wird durch einige in der ursprünglichen Aufstellung nicht aufgeführte Programme erweitert.

<sup>229</sup>Vgl. Ben W. Brumfield. *Transcription Tool List*. Blog-Eintrag. 11. Apr. 2012. URL: <http://manuscripttranscription.blogspot.de/2012/04/crowdsourced-transcription-tool-list.html> (besucht am 14.05.2012).

| Werkzeug-<br>name                  | Verant-<br>wortliche<br>Organisa-<br>tion                                | Platt-<br>form            | Lizenz  | Geho-<br>stet? | Text-<br>typ                            | TEI? | CMS-Integration   | Besondere Funktio-<br>nen   | Projekt- und<br>Quellcode-URL  | Anwen-<br>dungsbel-<br>spiele   |
|------------------------------------|--|---------------------------|---------|----------------|---|------|---|---|--|---|
| Bentham<br>Transcript-<br>ion Desk | University<br>of London<br>Computer<br>Centre; UCL<br>Bentham<br>Project | Media-<br>Wiki            | GPL 2.0 | Ja             |   | Ja   |   | Volle Unterstützung<br>für Textauszeichnung<br>mit TEI; Werkzeug-<br>eiste zur Verwen-<br>dung von TEI Tags | Projekt:<br><a href="http://www.ucl.ac.uk/transcribe-bentham/">http://www.ucl.ac.uk/transcribe-bentham/</a><br>Quellcode:<br><a href="http://code.google.com/p/tb-transcription-desk/">http://code.google.com/p/tb-transcription-desk/</a>                         | Transcribe<br>Bentham;<br>Public Re-<br>cord Office<br>of Victoria<br>(Melbourne);<br>Transcripti-<br>on Pilot<br><a href="http://prov.versity.edu.au/">http://prov.versity.edu.au/</a> |
| T-Pen                              | St. Louis<br>University<br>Center for<br>Digital<br>Theology             | Java /<br>Java-<br>script | EPL 2.0 | Ja             | zeilen-<br>basiert,<br>Mittelal-<br>ter | Ja   | Benutzer können Ex-<br>portkanäle erstellen<br>mit deren Hilfe Tran-<br>skriptionen direkt in<br>eine CMS-Datenbank<br>wie Drupal exportiert<br>werden können | Verknüpfung von<br>Transkriptionszeilen<br>im Text mit dem Scan   | Projekt:<br><a href="http://digital-e-ditor.blogspot.de/">http://digital-e-ditor.blogspot.de/</a>  | <a href="http://t-pen.org/TPEN/">http://t-pen.org/TPEN/</a>   |
| TextLab                            | John<br>Bryant, et<br>al, Hofsta-<br>University                          |                           |         | Nein           | Un-<br>strukturiert                     | Ja   |   | Direkte Annotation<br>von TEI add/del Tags<br>zu Scans  | Projekt:<br><a href="http://mel.hofstra.edu/textlab.html">http://mel.hofstra.edu/textlab.html</a>  | Melville<br>Electronic<br>Library   |
| Islandora<br>TEI Editor            | UPEI (?)   | Drupal /<br>Fedora        | GPL 3.0 | Ja             | Un-<br>strukturiert                     | Ja   | Fedora  | TEI-Auszeichnung für<br>in Fedora gehostete<br>Dokumente  | Projekt:<br><a href="http://wiki.tei-c.org/index.php/IslandoraTEIEditor">http://wiki.tei-c.org/index.php/IslandoraTEIEditor</a><br>Quellcode:<br><a href="https://github.com/Islandora/islandora_tei_editor">https://github.com/Islandora/islandora_tei_editor</a> | Public<br>Records<br>Office,<br>Victoria<br><a href="http://prov.versity.edu.au/">http://prov.versity.edu.au/</a>   |
| Son of Suda<br>On-Line             | Integrating<br>Digital<br>Papyrology                                     | Ruby on<br>Rails          | GPL 3.0 | Nein           | Un-<br>strukturiert                     | Ja   | Keine   | Git Back-End  | Projekt: <a href="http://papyri.github.com/documentation/README.html">http://papyri.github.com/documentation/README.html</a><br>Quellcode:<br><a href="https://github.com/papyri/sosol">https://github.com/papyri/sosol</a>  | papyri.info   |

Transkriptionswerkzeuge

(Tabelle wird fortgesetzt)

| Werkzeug-<br>name | Verant-<br>wortliche<br>Organisa-<br>tion                                  | Platt-<br>form                    | Lizenz         | Geho-<br>stet? | Text-<br>typ                     | TEI? | CMS-Integration             | Besondere Funktio-<br>nen                             | Projekt- und<br>Quellcode-URL  | Anwen-<br>dungsbel-<br>spiele   |
|-------------------|--|-----------------------------------|----------------|----------------|----------------------------------|------|-----------------------------|---|--|---|
| Wikisource        | Wikimedia  | Media-<br>Wiki                    | GPL 2.0        | Ja             | Un-<br>strukturiert              | Nein | Archive.org                 | Workflow<br>management                                | Projekt:<br><a href="http://en.wikisource.org/wiki/Main_Page">http://en.wikisource.org/wiki/Main_Page</a><br>Quellcode:<br><a href="http://www.mediawiki.org/wiki/MediaWiki">http://www.mediawiki.org/wiki/MediaWiki</a> | NARA<br>Citizen<br>Archivist<br>Dashboard   |
| FromThePage       | Ben Brumfield  | Ruby on<br>Rails                  | AGPL<br>3.0    | Ja             | Un-<br>strukturiert              | Nein | Archive.org                 | Semantische Auszeichnung für Indexierung / Annotation | Projekt:<br><a href="http://beta.fronthepage.com/">http://beta.fronthepage.com/</a><br>Quellcode:<br><a href="https://github.com/benbrum/fronthepage/wiki">https://github.com/benbrum/fronthepage/wiki</a>               | San Diego<br>Natural<br>History<br>Museum:<br>Laurence<br>M. Klauber<br>Field Notes<br>Southwestern University:<br>Zenas<br>Matthews<br>Diary               |
| Scripto           | Center for<br>History and<br>New Media<br>at George<br>Mason<br>University | PHP<br>library,<br>Media-<br>Wiki | GPL 3.0        | Nein           | Un-<br>strukturiert,<br>Wikitext | Nein | Omeka, Wordpress,<br>Drupal | Kann theoretisch in<br>jedes CMS integriert<br>werden | Projekt:<br><a href="http://scripto.org">http://scripto.org</a><br>Quellcode:<br><a href="https://github.com/chmm/Scripto">https://github.com/chmm/Scripto</a>   | Papers of<br>the War<br>Department,<br>1784-1800  |
| Scribe            | Zooniverse   | jQuery /<br>Ruby on<br>Rails      | MIT-<br>Lizenz | Auf<br>Antrag  | Strukturierte<br>Daten           | Nein |                             | Triple-Keying; Daten<br>werden mit Scans<br>verlinkt  | Quellcode:<br><a href="https://github.com/zooniverse/Scripto">https://github.com/zooniverse/Scripto</a>  | „What's the<br>Score“ an<br>der Bodleian<br>Library<br>(frühere<br>Versionen<br>auf <a href="http://www.oldweat-her.org/">http://www.oldweat-her.org/</a> ) |
| PyBOSSA           | Citizen<br>Cyber-<br>science<br>Centre/<br>OKFN                            | Python<br>/ GDocs                 | AGPL<br>3.0    |                | tabellarisch                     | Nein |                             | Dateneingabe<br>GDDoc-Tabellen                        | Projekt:<br><a href="http://pybossa.com/">http://pybossa.com/</a><br>Quellcode:<br><a href="https://github.com/PyBossa/pybossa">https://github.com/PyBossa/pybossa</a>   |   |

Transkriptionswerkzeuge

(Tabelle wird fortgesetzt)

| Werkzeug-<br>name   | Verant-<br>wortliche<br>Organisa-<br>tion                     | Platt-<br>form                  | Lizenz                              | Geho-<br>stet? | Text-<br>typ                                     | TEI? | CMS-Integration | Besondere Funktio-<br>nen   | Projekt- und<br>Quellcode-URL  | Anwen-<br>dungsbei-<br>spiele   |
|---|---|---------------------------------|-------------------------------------|----------------|--|------|-----------------|---|--|---|
| FieldData   | Atlas of<br>Living Aus-<br>tralia/Gaia<br>Resources           | Java                            | Mozilla<br>Public<br>Licence<br>1.1 |                |  | Nein |                 |   | Projekt:<br><a href="http://www.ala.org.au/get-involved/">http://www.ala.org.au/get-involved/</a><br><a href="http://citizen-science/fielddata-software/">citizen-science/fielddata-software/</a><br>Quellcode:<br><a href="http://code.google.com/p/ala-citizenscience/">http://code.google.com/p/ala-citizenscience/</a><br>Projekt:<br><a href="http://www.digitisation.eu/">http://www.digitisation.eu/</a><br>Quellcode:<br><a href="https://github.com/impactcentre/alto-editor">https://github.com/impactcentre/alto-editor</a> | <a href="http://volunteeer.ala.org.au/project/index/42780">http://volunteeer.ala.org.au/project/index/42780</a> |
| Alto Editor   | IMPACT<br>Centre of<br>Competence                             | Javascript<br>/ Ruby            | Apache<br>2.0                       | Nein           |  | Nein | Nein            |   |  |   |
| OpenScribe  |   | Drupal                          | Perl                                |                | Un-<br>strukturiert                              |      | Drupal          |   | Quellcode:<br><a href="http://code.google.com/p/openscribe/">http://code.google.com/p/openscribe/</a>  |   |
| Ancestry<br>World<br>Archives<br>Project                    | <a href="http://www.ancestry.com">http://www.ancestry.com</a> | installierter<br>.exe<br>Client | proprie-<br>tär                     |                | Struk-<br>turierte<br>Daten<br>(Genea-<br>logie) |      |                 | Schwierigkeitsbewer-<br>tung; kontextbasierte<br>Hilfe                        | Projekt:<br><a href="http://community.ancestry.co.uk/awap">http://community.ancestry.co.uk/awap</a>  | <a href="http://www.worldmemoryproject.org/">http://www.worldmemoryproject.org/</a>                             |
| Unbindery   | Ben Crowder   | PHP /<br>Javas-<br>cript        | MIT                                 | Ja             | Un-<br>strukturiert                              |      | Ist ein CMS     |   | Projekt:<br><a href="http://bencrowder.net/blog/category/unbindery/">http://bencrowder.net/blog/category/unbindery/</a><br>Quellcode:<br><a href="https://github.com/bencrowder/unbindery">https://github.com/bencrowder/unbindery</a>   | <a href="http://bencrowder.net/books/mtp/">http://bencrowder.net/books/mtp/</a>                                 |
| National<br>Archives<br>Transcrip-<br>tion Pilot<br>Project | U.S. Natio-<br>nal Archives                                   | Drupal                          |                                     |                | Un-<br>strukturiert                              |      | Drupal          | Schwierigkeitsbe-<br>wertung, Kommen-<br>tierung, Links zum<br>Online-Katalog | Projekt:<br><a href="http://transcribe.archives.gov/">http://transcribe.archives.gov/</a>  |   |

Transkriptionswerkzeuge

(Tabelle wird fortgesetzt)

| Werkzeug-<br>name   | Verant-<br>wortliche<br>Organisa-<br>tion                  | Platt-<br>form                                      | Lizenz          | Geho-<br>stet? | Text-<br>typ                                     | TEI? | CMS-Integration | Besondere Funktio-<br>nen  | Projekt- und<br>Quellecode-URL  | Anwen-<br>dungsbel-<br>spiele   |
|---|--|---|-----------------|----------------|--|------|-----------------|--|---|---|
| North Ame-<br>rican Bird<br>Phenology<br>Program                  | US Geologi-<br>cal Survey                                  |   |                 |                | Struk-<br>turierte<br>Daten                      | Nein |                 |  | Projekt:<br><a href="http://www.parc.usgs.gov/bpp/">http://www.parc.usgs.gov/bpp/</a>   | ?   |
| What's On<br>the Menu?  | New York<br>Public<br>Library                              |   |                 |                | Struk-<br>turierte<br>Daten                      | Nein |                 |  | Projekt:<br><a href="http://menus.nyp1.org/">http://menus.nyp1.org/</a>   |   |
| Family<br>Search<br>Indexing                                      | Family<br>Search   |   | proprie-<br>tär |                | Struk-<br>turierte<br>Daten<br>(Genea-<br>logie) | Nein |                 |  | Projekt:<br><a href="https://indexing.familysearch.org/neuser/nuhome.jsf?3.9.6">https://indexing.familysearch.org/neuser/nuhome.jsf?3.9.6</a> |   |
| Harold<br>„Doc“<br>Edgerton<br>Project                            | MIT?   |   |                 |                | Un-<br>strukturiert                              |      |                 |  | Projekt:<br><a href="http://edgerton-digital-collection.org/notebooks">http://edgerton-digital-collection.org/notebooks</a>                   |   |
| Civil War<br>Diaries and<br>Letters<br>Transcripti-<br>on Project | The Univer-<br>sity of Iowa<br>Libraries                   |   |                 |                | Un-<br>strukturiert                              |      |                 |  | Projekt:<br><a href="http://digital.lib.uiowa.edu/cwd/traascripts.html">http://digital.lib.uiowa.edu/cwd/traascripts.html</a>                 |   |
| Velchanden.nl   |  |   |                 |                | Struk-<br>turierte<br>Daten                      |      |                 |  |   | <a href="http://militieregisters.nl/">http://militieregisters.nl/</a> |
| Virtual<br>Trans-<br>cription<br>Laboratory                       | Poznan'<br>Supercom-<br>puting and<br>Networking<br>Center | Java En-<br>terprise<br>Edition<br>+ Tes-<br>seract |                 |                | Nein   |      |                 | Integriertes<br>Werkzeug<br>sche<br>Quellen;<br>zum Trainieren des<br>OCR-Werkzeugs;<br>Verlinkung von Tran-<br>skription und Scan | Projekt:<br><a href="http://vlt.synat.pcss.pl/">http://vlt.synat.pcss.pl/</a>   |   |
| Veridian  | DL Consul-<br>ting   |   | Proprie-<br>tär |                |  |      |                 | Strukturierte<br>Daten   |   |   |

Transkriptionswerkzeuge

(Tabelle wird fortgesetzt)

| Werkzeug-<br>name      | Verant-<br>wortliche<br>Organisa-<br>tion   | Platt-<br>form           | Lizenz  | Geho-<br>stet?   | Text-<br>typ   | TEI? | CMS-Integration | Besondere Funktio-<br>nen   | Projekt- und<br>Quellcode-URL   | Anwen-<br>dungsbe-<br>spiele  |
|------------------------|---|--------------------------|---------|--|--|------|-----------------|---|---|---|
| Itineranova-<br>Editor | Stadsarchief<br>Leuven/HKI<br>Cologne   | XRX /<br>Javas-<br>cript | GPL 3.0 |  |  | Ja   |                 |   | Projekt:<br>http://www.mon-w<br>iki.uni-koeln.de/<br>Quellcode:<br>https://subversi<br>on.rz.uni-koeln.d<br>e/trac/exist-A/b<br>rowser/trunk/my/<br>XRX/www/in                  | http://www.i<br>tineranova.b<br>e/  |
| VdU-Editor             |   | XRX /<br>Javas-<br>cript | GPL 3.0 | Un-<br>struk-<br>turiert<br>sowie<br>Struk-<br>turierte<br>Daten                       | Ja (für<br>jedes<br>XSD-<br>Schema<br>konfigu-<br>rierbar) |      |                 | Vollständiger XML-<br>Editor mit versteckter<br>XML-Syntax                      | Projekt:<br>http://www.mon-w<br>iki.uni-koeln.de/<br>Quellcode:<br>https://subversi<br>on.rz.uni-koeln.d<br>e/trac/exist-A/b<br>rowser/trunk/my/<br>XRX/www/core/app<br>/editor |   |
| Refinel                | Staatsbibliothek<br>zu Berlin,<br>Humboldt-<br>Universität<br>zu Berlin<br>und 3pc<br>GmbH Neue<br>Kommunika-<br>tion | FCKeditor                | GPL     | prinzipiell<br>durch<br>3pc als<br>kommer-<br>zieller<br>Dienst-<br>leister<br>möglich | Un-<br>struktu-<br>riert                                   | Ja   | Nein            | Automatisierte Erstel-<br>lung von TEI-Dateien;<br>Datenimport aus Kal-<br>hope | http://bruemm<br>er.staatsbib<br>liothek-berli<br>n.de/nlbruemm<br>er/<br>http://nicola<br>i.host6.3-poi<br>nt.de/  |   |
| eLaborate              | Huygens<br>Institute<br>for the His-<br>tory of the<br>Netherlands  |                          |         | Ja   |  |      |                 |   | Projekt:<br>https://www.elabo<br>rate.huygens.kna<br>w.nl/<br>https://www.w<br>iki.uni-koeln.d<br>e/trac/exist-A/b<br>rowser/trunk/my/<br>XRX/www/core/app<br>/editor           | http://bartho<br>lomeusengels<br>man.huygens.k<br>nav.nl/path<br>http://mart<br>ianus.huygens<br>instituut.kna<br>w.nl/path<br>http://www.w<br>aleweinendek<br>eye.huygens.k<br>nav.nl/path |

Transkriptionswerkzeuge

(Tabelle wird fortgesetzt)

| Werkzeug-<br>name | Verant-<br>wortliche<br>Organisa-<br>tion        | Platt-<br>form           | Lizenz | Geho-<br>stet? | Text-<br>typ        | TEI?    | CMS-Integration | Besondere Funktio-<br>nen  | Projekt- und<br>Quellcode-URL   | Anwen-<br>dungsbei-<br>spiele |
|-------------------|--|--------------------------|--------|----------------|---------------------|---------|-----------------|--|---|-------------------------------|
| OTTO              | Ruhr-<br>Universität<br>Bochum                   | PHP /<br>Javas-<br>cript |        |                | Un-<br>strukturiert | Geplant | Nein            |  | Projekt:<br>http://www.linguistics.ruhr-uni-bochum.de/~dipper/project_otto.html                                     |                               |
| TEXTUS            | Open Knowledge Foundation                        |                          | frei   |                |                     |         |                 |  | Projekt:<br>http://textusproject.org/   | http://openphilosophy.org/    |
| CWRC-<br>Writer   | Canadian<br>Writing<br>Research<br>Collaboratory | TinyMCE                  |        |                |                     | Ja      | Nein            | Visuelle Repräsentation von Markup; Editor zur Textauszeichnung im Browser | Projekt:<br>http://www.dh2012.uni-hamburg.de/conference/program/abstracts/cwrc-writer-an-1<br>n-browser-xml-editor/ |                               |

Tabelle 10.25: Liste der Transkriptionswerkzeuge

# Eidesstattliche Erklärung

Ich erkläre hiermit an Eides Statt, dass ich die vorliegende Arbeit selbständig und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe; die aus fremden Quellen (einschließlich elektronischer Quellen) direkt oder indirekt übernommenen Gedanken sind ausnahmslos als solche kenntlich gemacht.