

Digitale Nachlassrekonstruktion

Stefan Zweig

2016/17

Projektdefinition

Pflichtenheft

christopher.pollin@uni-graz.at

23.11.2016
<Version 6>

Inhaltsverzeichnis

Inhaltsverzeichnis

1 Projektbeschreibung

1.1 Projektdefinition

1.2 Projektziele

1.3 Auflistung des vorhandenen Datenmaterials

1.4 Zielgruppen

1.5 Work-Packages

1.5.1 Dokumentation

1.5.2 Datenmodellierung

1.5.3 Ingest

1.5.4 Portal

1.6 Technische Infrastruktur

2 Pflichtenheft

2.1 Funktionale Anforderungen

2.2 Nichtfunktionale Anforderungen

Definitionen und Abkürzungen

1 Projektbeschreibung

1.1 Projektdefinition

Der Nachlass des österreichischen Schriftstellers Stefan Zweig stellt ein wichtiges kulturelles Gut, sowohl für Literaturinteressierte, als auch für ein wissenschaftliches Publikum dar. Aus diesem Grund gilt es nicht nur das kulturelle Erbe – den Nachlass – zu bewahren, sondern auch zugänglich zu machen. Da Zweigs Nachlass verstreut ist und sich an unterschiedlichen Orten in unterschiedlichen Besitzverhältnissen befindet, besteht eine Notwendigkeit der Zusammenführung dieser Objekte im digitalen Raum in Form einer *digitalen Sammlung*. Aufbauend auf die digitale Sammlung soll eine *digitale Nachlassrekonstruktion* des Bestandes generiert werden. So entsteht ein strukturierter Bestand an digitalen Objekten, der im Sinne der digitalen Langzeitarchivierung repräsentiert wird und den NutzerInnen orts- und zeitunabhängig zugänglich ist. Das Projekt ist so angelegt, dass zu einem späteren Zeitpunkt Erweiterungen (z.B. digitalen Editionen etc.), sowie eine Migration des Projektes auf Server der Universität Salzburg möglich sind.

1.2 Projektziele

Folgende Projektziele gilt es in diesem Projekt zu erreichen. Jedes Projektziel entspricht einem Work-Package (siehe Kapitel 1.5) und kann als Milestone verstanden werden.

1. **Dokumentation**

Es existiert eine ausführliche Dokumentation des Projektes, um die Übertragung der Daten, der Workflows, der Ergebnisse der Implementierung zu einem späteren Zeitpunkt gewährleisten zu können.

2. **Datenmodellierung**

Die im Projekt generierten Daten aus dem Nachlass von Stefan Zweig und aus Sekundärquellen (siehe: 1.3) werden standardbasiert modelliert.

3. **Ingest**

Für den Ingest der digitalen Objekte (Import der Daten in die aufzubauende Plattform) existieren standardisierte Workflows.

4. **Portal**

Es soll ein Webportal existieren, das mittels unterschiedlicher Devices (Desktop-PC und Smartphone) barrierefrei erreichbar ist und den Zugang zur digitalen Repräsentation des Stefan Zweig Nachlasses ermöglicht.

5. Digitale Nachlassrekonstruktion

Die Nachlassrekonstruktion bietet Möglichkeiten des Information Retrieval und des Resource Discovery, um den erschlossenen Nachlass zugänglich zu machen. Neben einer facettierten Suchvariante kann das Nachlassmaterial - wenn dies im Projekt umgesetzt wird - unter Zugrundelegung einer Ontologie semantisch erschlossen werden.

1.3 Auflistung des vorhandenen Datenmaterials

Die gesamte Datenerfassung erfolgt an der Universität Salzburg über das Bibliothekssystem Aleph. Daten werden im Aleph-Format bzw. Excel-Sheets, die sich an diesem Format orientieren (gleiche Felder), an den Auftragnehmer übermittelt. Partnerinstitutionen übermitteln Katalogisate in einer mit der Universität Salzburg abgesprochenen Weise. Die an den Auftraggeber übermittelten Daten sind normalisiert und redaktionell überarbeitet. Der Auftraggeber verpflichtet sich, Datenformate und Modellierungsrichtlinien grundsätzlich mit dem Auftragnehmer abzusprechen.

Scans

Die Universitätsbibliothek Salzburg ist mit der Erstellung der Digitalisate der Objekte (Briefe, Handschriften, etc.) im Nachlass zuständig. Diese Scans werden an den Auftragnehmer als TIFF übermittelt.

Nachlasskatalog

Der Nachlasskatalog ist eine Katalogisierung der Objekte, die sich im Nachlass Stefan Zweigs befanden und die heute noch erhalten sind. Die Daten werden im Aleph-Format übergeben. In diesem Prozess gibt es weitere Partner (State University New York in Fredonia, USA), die ihre Daten als Excel-Sheets (orientiert am Aleph-Format) übergeben.

Bibliothekskatalog

Der Bibliothekskatalog ist ein Katalog, der die heute noch erhaltenen Bücher, die sich in Stefan Zweigs Bibliothek(en) befanden, umfasst. Für diese Daten gelten die gleichen Voraussetzungen wie für den Nachlasskatalog. Die Daten werden im Aleph-Format übergeben.

Lebenskalender

Die Daten werden als Excel-File übergeben. Der Lebenskalender beschreibt das Leben von Stefan Zweig chronologisch. Dieses Dokument wurde von einer privaten Person, ohne spezielle Fachkenntnisse, erstellt. Der Wert des Dokuments ist auf Grund seiner Ausführlichkeit und Genauigkeit von großem Interesse. Dieses Dokument bedarf einer redaktionellen Überarbeitung, einer Normalisierung, sowie einer Auswahl. Diese Datenquelle beinhaltet narrative Elemente und kann zur Rekonstruktion des Nachlasses herangezogen werden.

Kommentartexte

Beschreibende Fließtexte für Teile des Bestandes und einzelne Objekte.

Verzeichnis von Stefan Zweigs Autographensammlung

Es liegt ein PDF zur Autographensammlung von Stefan Zweig, bestehend aus Briefen, Partituren und handschriftlich Werken anderer berühmter Persönlichkeiten vor. Die Daten werden analog zu jenen des Nachlass- und Bibliothekskatalogs aufbereitet. Die Autographensammlung umfasst Daten zu Person, Titel, Jahr, Inhalt und ergänzenden Angaben. Verwertungsrechte für diese Datenquelle sind noch zu klären.

1.4 Zielgruppen

Im Projekt werden zwei unterschiedliche Haupt-Zielgruppen unterschieden, für die das Portal entsprechende Funktionalitäten bereitstellt. Eine Benutzerverwaltung ist zum gegenwärtigen Zeitpunkt nicht intendiert und auch ausdrücklich nicht Gegenstand des vorliegenden Vertrages zwischen Auftraggeber und Auftragnehmer.

General User

Darunter versteht man einen User mit einem akademischen oder fachlichem Hintergrund, der die digitale Sammlung als wissenschaftlichen Quellen- und Informationsfundus betrachtet und verwerten möchte. Er ist in der Lage mit Findmitteln umzugehen, kennt den Bestand und will ausgewählte Werke betrachten. Er verwendet adäquate Funktionalitäten des gezielten Suchens (Volltext, Inventar, Klassifikation, etc.).

Public User

Als Public User wird jene Gruppe von Usern verstanden, die als 'Laien' ihr Interesse am Schriftsteller und dessen Werke vertiefen möchte. Ihre Kenntnisse von Findmitteln sind gering. Diese Gruppe kennt einzelne Werke, hat aber keine Informationen darüber wie der Nachlass strukturiert ist, oder welche Inhalte dort

vorzufinden sind. Er braucht adäquate Funktionalitäten des Suchens und des *Resource Discovery* (Browsing, Themen, etc.)

1.5 Work-Packages

Vier Arbeitspakete definieren den Projektverlauf. Ihre Reihenfolge entspricht in etwa ihre gegenseitige Abhängigkeit. Die Modellierung der Daten und deren Ingest-Workflow sind Voraussetzung für eine funktionierende Darstellung im Web.

1.5.1 Dokumentation

Im Projekt wird besonderer Wert auf eine ausführliche Dokumentation gelegt. Folgende Aspekte werden dabei berücksichtigt:

a) Vorgaben

Die Dokumentation der Vorgaben sind alle Aspekte, die vor Projektbeginn festgelegt werden. Sie sind in Form dieses Pflichtenheftes definiert und beinhalten:

- Projektziele
- Workflows
- Work-Packages
- Funktionalitäten, Use-Cases
- Technische Vorgaben

b) Umsetzung

Die Dokumentation der Umsetzung ist eine Dokumentation des laufenden Prozesses:

- Framework/Plattform
- Softwarearchitektur
- Daten und Schnittstellen
- Quellcode

1.5.2 Datenmodellierung

Die Datenmodellierung beschreibt alle Arbeitsbereiche, die notwendig sind, um aus den Rohdaten standardisierte und strukturierte Daten zu gewinnen, die für den Ingest in die GAMS-Infrastruktur geeignet sind. Für die Modellierung werden vonseiten der Universität Salzburg Vorschläge erarbeitet, die bei der Projektrealisierung Berücksichtigung finden.

Digitalisate des Nachlasses

Masterfiles werden als TIFFs übergeben und als JPEGs in GAMS gehostet. Die Bereitstellung der Scans orientiert sich am "Seitenparadigma". Das Metadatenformat für die Strukturinformation der Archivobjekte muss noch festgelegt werden.

Kommentartexte

Für die Erfassung der Kommentartexte wird ein Workflow implementiert, der es ermöglicht Office-Dokumente automatisiert nach TEI zu transformieren.

Nachlasskatalog

Der Nachlasskatalog wird an der Universität Salzburg und an der State University New York in Fredonia erstellt und gepflegt. Der Auftragnehmer erhält diese Daten im Aleph-Export-Format mit MAB-Schema bzw. als Excel-Sheets (orientiert am Aleph-Format).

Bibliothekskatalog

Der Bibliothekskatalog wird als Excel-Sheet übergeben.

Lebenskalender

Der Lebenskalender wird in TEI modelliert und über zwei unterschiedliche Visualisierungsformen am Portal repräsentiert: (a) als Navigationsstruktur für den Nachlass (Zeitleiste, chronologische Darstellung des Kalenders) und (b) als narrative Struktur.

Verzeichnis von Stefan Zweigs Autographensammlung

Das Verzeichnis liegt derzeit in PDF-Format vor. Die Autographensammlung soll analog zu Nachlass- und Bibliothekskatalog gestaltet sein, die Daten werden dementsprechend an der Uni Salzburg vorbereitet.

Fotos

Für Fotografien aus dem Zweig Nachlass wird eine Bildgalerie angedacht.

1.5.3 Ingest

Dem Auftraggeber steht ein Client zur Verfügung mit dem die Seiteninhalte des Portals gepflegt und erweitert werden können.

1.5.4 Portal

Webdesign

Der Grafikentwurf des Portals wird vom Auftraggeber als Screenshots beigestellt. Besprochener Ablauf ist folgender: Vorerst wird, basierend auf einem GAMS-Standard-Template, entwickelt bis das Design im März 2017 fertig ist. Aus den Vorschlägen der Grafikerinnen wird ein Kompromiss gefertigt, der die Aspekte

der Langzeitarchivierung und der GAMS-Infrastruktur, Barrierefreiheit, Responsiveness etc. berücksichtigt. Unter dem Work-Packages Portal werden alle Tätigkeiten verstanden, die zur Implementierung der Webpräsenz der digitalen Sammlung mit all seinen im Pflichtenheft definierten Funktionalitäten. Diese Funktionalitäten sind im Pflichtenheft (siehe: Kapitel 2) definiert.

Ein **Mockup**:

<https://app.moqups.com/LiMaZa/V3y9v0Rdrh/view/page/ad64222d5>

1.6 Zeitplan

Ausgehend von einer Arbeitswoche im Ausmaß von **20 Stunden/Woche** werden folgende Zeiträume und Deadlines für Arbeitspakete festgelegt. Die Zeiträume werden als Richtwerte betrachtet. In Abhängigkeit von spezifischen Erfordernissen können die Zeiträume einzelner Arbeitspakete abweichen.

Projektlaufzeit: **15.12.2016-15.12.2017**
1 Jahr nach Vertragsbeginn

Vorbereitung (Pflichtenheft, Projekt einrichten)
Zeitraum: 01.12.2016-15.01.2017

Dokumentation
Zeitraum: 15.01.2017-15.12.2017

Datenmodellierung

a) ausgehend von Testdaten

Zeitraum: 15.01.2017-01.05.2017
Deadline: Erhalt Testdaten: **15.01.2017**

b) ausgehend von Daten

Zeitraum: 01.06.2017-01.09.2017
Deadline: Erhalt aller fertigen Daten: **01.06.2017**

Ingest

a) Einspielen Testdaten in Testumgebung

Zeitraum: 01.04.2017 - 01.05.2017

b) Einspielen Daten

Zeitraum: 01.9.2017 -01.10.2017

Portal

a) Implementierung Template + Testdaten

Zeitraum: 1.04.2017 - 01.07.2017

b) Implementierung Design + Daten

Zeitraum: 01.07.2017 - 01.10.2017

Digitale Nachlassrekonstruktion (Implementierung spezifischer Funktionalitäten)

Zeitraum: 01.07.2017-01.10.2017

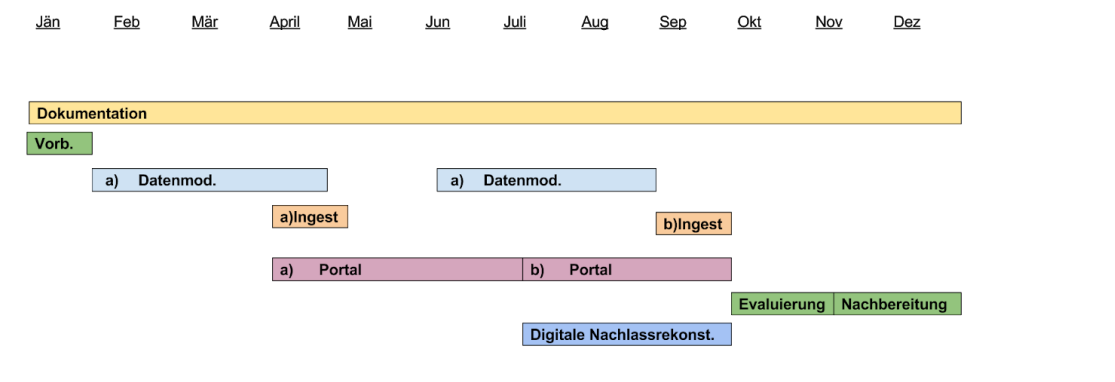
Evaluierung (Usability-Test)

Zeitraum: 01.10.2017-01.11.2017

Nachbereitung (Veröffentlichung, Projektabschluss)

Zeitraum: 01.11.2017-15.12.2017

Darstellung der Zeiträume der Arbeitspakete:



URL zur größeren Darstellung:

https://docs.google.com/drawings/d/1ihcbhmnQ6qw2jL0M29wRkxYBVhUn_S1YgNilIuK9aX8/pub?w=1485&h=600

1.7 Technische Infrastruktur

Die wichtigsten Punkte zur Infrastruktur sind extern dokumentiert (siehe <http://gams.uni-graz.at/doku>).

2 Pflichtenheft

Dieser Abschnitt definiert allgemein die Anforderungen, die im Projekt umgesetzt werden müssen. Dies beinhaltet die funktionalen Anforderungen und die nichtfunktionale Anforderungen.

2.1 Funktionale Anforderungen

Folgende Tabelle listet die funktionalen Anforderungen auf, sprich alle Funktionalitäten, die die Webseite aufweisen soll. In der Spalte 'Funktionalitäten' befindet sich die Bezeichnung und in der rechten Spalte die dazugehörige Beschreibung der Anforderung.

Funktionalität	Beschreibung
Webdarstellung - Static	Darstellung der Webinhalte, die im Portal angezeigt werden sollen bestehend aus: Header, Footer, Content-Bereich, Impressum, Kontakt. (szd-static.xsl)
Detailansicht Objekt	Darstellung einer Ansicht eines digitalen Objektes bestehend aus Metadaten, Exportmöglichkeiten und Viewer. (szd-object.xsl)
Volltextsuche	Eingabe eines Textstrings in ein Suchfenster. Ausgabe aller Objekte in denen der eingegebene Textstring vorkommt. (szd-search.xsl) Begrenzung der Suche auf Sammlungsbereiche.
Erweiterte Suche	Suche im Bestand mittels eines kontrollierten Vokabulars. Die Suchmöglichkeiten könnten beinhalten: nach Zeit; nach Person; nach Kategorien, Themen, Stichwörter; Facettierte-Suche; etc. Auto-Complete bei Eingabe. Begrenzung der Suche auf Sammlungsbereiche.

Resource Discovery (Entdecken)	Möglichkeit des Entdeckens und Browsings der Nutzerin ohne größere Kenntnis oder Ziel. Gleiche Anforderungen, wie im Punkt <i>Erweiterte Suche</i> , jedoch anderer Zugang und andere Visualisierung.
Exportmöglichkeit	Möglichkeiten Objekte in gewissen Formen von der Webseite herunterzuladen. PDF; Print-Version; TEI/XML; Metadaten;
Datenkorb	Digitale Objekte auswählen und in gewisser Form lokal für eine Session (Local Storage) zu speichern und zu exportieren. Beispielsweise in einer Excel-Liste oder einem PDF.
Viewer	Ansicht der Digitalisate (Scans) in einem DFG-Viewer.
Lebenskalender (Erschließung)	HTML-Darstellung des Lebenskalenders in einer chronologisch, scrollbaren Form. Aufklappbare Jahre mit Möglichkeit tieferer Erschließung. Tendenziell für General User.
Lebenskalender (Narratives Element)	Lebenskalender in Form eines TimeLine.js (oder ähnliche Applikation). Tendenziell für Public User.

2.2 Nichtfunktionale Anforderungen

Folgende nichtfunktionale Anforderungen gelten allgemein.

- **Zuverlässigkeit**
 - Der Aufruf der Seite mittels Webbrowser sollte gewährleistet sein. Eine Erreichbarkeit kann nicht zu 100% gewährleistet sein (Wartungsarbeiten), aber sollte sich diesem Wert annähern.
- **Performance**
 - Niedrige bzw. vertretbare Latenz beim Laden von Inhalten.
- **Portierbarkeit und Responsiveness**
 - Aufruf der Webseite muss für unterschiedliche Devices (Desktop oder Smartphone) gewährleistet sein.

- Dazu werden beide Systeme basierend auf dem Template Bootstrap entwickelt, dessen grundlegendes Prinzip die Responsiveness darstellt.
- **Usability**
 - Ein User muss ohne Vorkenntnisse zielsicher die Funktionalitäten umsetzen können, die im System möglich sind.
- **Erweiterbarkeit**
 - Das Projekt soll darauf ausgelegt sein, dass nach dem Erreichen der in diesem Projektplan definierten Ziele noch weiter daran gearbeitet werden kann. Aus diesem Grund wird Wert auf eine ausführliche Dokumentation gelegt.
- **Mehrsprachigkeit**
 - Inhalt wird in deutscher und in englischer Sprache angezeigt.
- **Accessibility**
 - Soweit möglich sollen alle gängigen Standards der Accessibility eingehalten werden, um einen barrierefreien Zugang zur Website zu ermöglichen.

Definitionen und Abkürzungen

Cirilo

Client für den Ingest von Daten in die GAMS.
<http://gams.uni-graz.at/docs#cirilo>

DC: Dublin Core

<http://dublincore.org>

DFG-Viewer

<http://www.dfg-viewer.de>

Digitale Nachlassrekonstruktion

Möglichkeit auf Basis der vorhanden und dargestellten Daten, sowie Funktionalitäten den Nachlass im Web rekonstruieren zu können.

Digitale Sammlung

Eine digitale Sammlung besteht aus digitalen Objekten, den standardisierten Metadaten, einem User Interface und ermöglicht auf Grund der Einbettung digitaler Objekte in Ordnungsstrukturen den Zugang, das Suchen und Entdecken, sowie die Nutzung digitalisierte Objekte.

Digitales Objekt

Ein digitales Objekt wird als ein „object of a set of bit sequences“ definiert, die von einer Maschine verarbeitet werden können. Digitale Objekte bestehen aus Teilaspekten, die durch ein physikalisches, logisches und konzeptionelles Objekt beschrieben werden, da digitale Objekte in unterschiedlichen Repräsentationen auftreten können.

Erweiterte Suche

Sucht nicht nach Textstrings sondern nach semantischen Wertigkeiten und nutzt dabei Normdaten und kontrollierte Vokabularien.

FEDORA Commons

<http://fedora-commons.org>

Der GAMS zugrunde liegender Repositorium-Software.

<https://wiki.duraspace.org/display/FEDORA/All+Documentation>

GAMS-Infrastruktur

Beschreibt das Zusammenspiel des OAIS-konformen Asset Management am ZIM aus Server, Software, Applikationen, Institution und Workflows.

JPEG

<https://de.wikipedia.org/wiki/JPEG>

Metadaten

Metadaten sind strukturierte Information über ein bestimmtes digitales Objekt. So sind zum Beispiel die Informationen Autor, Titel und ISBN Metadaten eines Buches. Die Metadaten kann man wiederum in unterschiedliche Ausprägungen aufteilen. So unterscheidet man beschreibende, strukturelle, technische und administrative Metadaten.

METS: Metadata Encoding and Transmission Standard

<http://www.loc.gov/standards/mets>

Ontologie

Eine Ontologie beschreibt ein Datenmodell zur Darstellung eines Sets miteinander vernetzter Konzepte innerhalb einer Domäne, mit der sich komplexere Sachverhalte und Schlüsse ziehen lassen können. Ontologien stellen einen Teilbereich des Semantic Webs dar. Im Unterschied zu einer hierarchischen Struktur (Thesaurus) können auf Basis des Netzwerkes von Information und ihrer definierten Regeln konsistente Schlüsse gezogen werden.

Persistent Identifier PID

Eine eindeutige Identifikation eines digitalen Objekts.

Bsp.: <http://gams.uni-graz.at/o:ufbas.1563>

Resource Discovery

Versteht alle Funktionalitäten des Entdeckens und Stöberns auf der Seite, die einem User, vor allem dem Public User, ermöglichen sollen den Bestand kennen zu lernen und Retrieval Möglichkeiten zu eröffnen.

SKOS: Simple Knowledge Organization System

<http://www.w3.org/2004/02/skos>

TEI: Text Encoding Initiative

<http://tei-c.org>

TIFF

https://de.wikipedia.org/wiki/Tagged_Image_File_Format

Volltextsuche

Die Volltextsuche basiert darauf, dass 'nur' Textstrings gesucht werden.

Der Rückgabewert umfasst alle Objekte, die diesen Textstring enthalten, entweder in den Metadaten, oder im Fließtext. Gibt nicht die Position im Dokument an, sondern nur die Dokument bzw. PID