

Workflow zum Projekt „Aufbau eines Quellencorpus für die seit den 1990er Jahren entstehende Literaturgattung „Netzliteratur““

Status: Draft

Mit * markierte Prozesse sind noch weiter auszuführen

Beschreibung des Workflows

Begleitender Text zu den Graphen 4.a. bis 4.i. Beschrieben werden Rechteeinholung, Katalogisierung, Dokumentation, Analyse, Erfassung und Archivierung von Werken der Gattung Netzliteratur.

4.a. Rechteeinholung

Nach der Auswahl liegen die bibliographischen Daten wie Autor und Titel des Werks vor.

Diese werden in die Datenbank I eingetragen und eine Anfrage an den Autor wird gestellt.

Es werden zwei Versuche unternommen, den Autor/die Autoren bezüglich einer Rechteerteilung zu kontaktieren. Ein dritter Versuch soll dabei bezüglich der Formulierung dringlicher gestaltet werden.

Werden die Rechte nicht erteilt, so wird das Werk nicht vom DLA Marbach archiviert und der Grund dafür in Datenbank I vermerkt.

Im Falle einer Rechteerteilung werden die Rechte in der Datenbank I vermerkt.

4.b. Katalogisierung

Die bibliographischen Daten werden in Kallías übertragen und eine Titelaufnahme erstellt.

Handelt es sich um ein monographisches Werk, kann die Titelaufnahme direkt in den SWB hochgeladen werden. Ist es periodisch, so muss zunächst eine Titelaufnahme in der ZDB erstellt werden und danach in den SWB hochgeladen werden.

In beiden Fällen wird beim Upload eine ppn bzw. eine ZDB-ID für den jeweiligen Titel vergeben.

Diese ppn bzw ZDB-ID wird in SWBContent eingegeben und damit eine Titelaufnahme in SWBContent angelegt.

Dabei wird ein URN für die Frontpage in SWBContent vergeben.

Zuletzt werden URL und URN der SWBContent-Titelaufnahme in Kallías vermerkt.

4.c. Wiki I

An Hand der in der Katalogisierung erfassten bibliographischen Daten wird ein Wiki-Eintrag angelegt und mit Grundinformationen befüllt. Dies geschieht durch Ausfüllen eines Templates*.

Anschließend wird der Eintrag kategorisiert*.

4.d.i. Vorabanalyse

In einem ersten Schritt wird überprüft, ob das Werk online verfügbar ist.

Ist das Werk online verfügbar, wird an Hand einer Beurteilungsmatrix* ermittelt, ob das Werk gespiegelt werden kann.

Die Beurteilung wird im Wiki dokumentiert.

Ist das Werk spiegelungsfähig, wird mit dem Crawl begonnen.

Ist das Werk nicht spiegelungsfähig, wird geprüft, ob alle Komponenten des Werks wie Quellcode oder Datenbank beschafft werden können.

Ist dies nicht der Fall, wird mit einem Screencast begonnen.

Können die Komponenten beschafft werden, wird der Prozess wie ‚nicht online verfügbar‘ fortgesetzt.

Ist das Werk ‚nicht online verfügbar‘, werden alle Komponenten des Werks wie Quellcode oder Datenbank beschafft.

Das Trägermedium der Daten wird in Datenbank II beschrieben.

Es wird ein Image des Datenträgers/eine Kopie der Daten angefertigt.

War die Erstellung eines Abbildes nicht erfolgreich, wird geprüft, ob eine erneute Medienbeschaffung durchgeführt wird.

Soll keine neue Medienbeschaffung durchgeführt werden, wird entschieden, ob das Trägermedium ein Sammlungsgut darstellt. Ist dies der Fall, wird es an die Abteilung Erwerbung übergeben. Ist dies nicht der Fall, wird das Medium zurückgesendet oder vernichtet.

Der Vorgang wird im Abschnitt Wiki II dokumentiert.

Soll eine erneute Beschaffung durchgeführt werden, tritt der Prozess erneut in die Beschaffung ein.

War die Erstellung eines Abbildes erfolgreich, wird eine mögliche und authentische Abspielumgebung auf Grundlage historisch-technischer Daten beschrieben.

Anschließend werden die Daten in Datenbank II beschrieben.

Nach diesem Schritt wird entschieden, ob das Trägermedium ein Sammlungsgut darstellt. Ist dies der Fall, wird es an die Abteilung Erwerbung übergeben. Ist dies nicht der Fall, wird das Medium zurückgesendet oder vernichtet.

4.d.ii. Sichtung Liveweb / Spiegelungsfähigkeit

Der Start-URL wird ermittelt.

Dieser dient dem Tool wget als Grundlage, alle zum Werk gehörenden Links zu spideren. Alle gefundenen Links werden in eine Textdatei gespeichert.

Dem Tool arrr.py wird diese Datei als Eingabe übergeben. arrr.py sendet jeden URL an den Webdienst archiveready.com, der die Spiegelungsfähigkeit einer Ressource beurteilt.

Die Rückgabe des Dienstes wird durch arrr.py formatiert in eine Textdatei geschrieben, den Beurteilungsreport.

Das Ergebnis muss einer Plausibilitätsprüfung unterzogen werden.

4.e.i. Crawl

Wird ein Werk als spiegelungsfähig beurteilt, so wird zuerst die URI übergeben, sodass der Crawler konfiguriert werden kann. Dies geschieht an Hand der Erkenntnisse aus der Beurteilung im Liveweb.

Der Crawl wird gestartet.

Ist dieser abgeschlossen, erhält man eine Datei im warc-Format, die in die Wayback Machine geladen werden kann.

So kann eine Qualitätsprüfung vorgenommen werden.

Fällt diese negativ aus, so muss entschieden werden, ob ein erneuter Versuch unternommen wird, das Werk zu spiegeln.

Soll ein erneuter Versuch unternommen werden, so muss die Konfigurationsdatei des Crawls angepasst und der Crawl neu gestartet werden.

Soll kein erneuter Versuch unternommen werden, so muss zum Punkt 4.d. Vorabanalyse zurückgekehrt und das Werk neu beurteilt werden.

Zuvor muss entschieden werden, ob die produzierte warc-Datei dennoch archiviert wird.

Bei einer positiv verlaufenen Qualitätsprüfung wird die warc-Datei transferiert.

Anschließend wird der Quellcode des Werks analysiert und in einem geeigneten Format* beschrieben. Außerdem wird die erforderliche Umgebung analysiert und ebenfalls in einem geeigneten Format* beschrieben.

Die warc-Datei, die Beschreibung der warc-Datei, der Quellcode, die Beschreibung des Quellcodes sowie die Beschreibung der Umgebung werden in einem temporären Verzeichnis gespeichert.

4.e.ii. Screencast

Zu Beginn werden an Hand der erstellten Beurteilung aus 4.d. die Einstellungen des darstellenden Desktops bestimmt.

Diese Werte werden im Wiki dokumentiert.

Ein Testlauf ohne Aufnahme dient der Erstellung eines Drehbuchs für die spätere Aufnahme. An Hand des Drehbuchs wird später die Aufnahme durchgeführt.

Das Drehbuch wird im Wiki dokumentiert.

Die Screencastsoftware wird konfiguriert.

Die Konfiguration der Screencastsoftware wird im Wiki dokumentiert.

Die Aufnahme wird dem Drehbuch folgend durchgeführt.

Nach Beendigung wird die Qualität der Aufnahme überprüft.

Wird die Qualität der Aufnahme für ausreichend befunden, wird die Videodatei in einem geeigneten Format* in einem Projektverzeichnis gespeichert. Eine technische Beschreibung der

Videodatei wird ebenfalls in diesem Verzeichnis abgelegt.

Wird die Qualität der Aufnahme für nicht ausreichend befunden, wird der Prozess erneut mit einer entsprechend angepassten Konfiguration der Screencast-Software begonnen.

4.e.iii. Offline

An Hand der Beschreibung der Umgebung aus der Voranalyse und der Werkdateien wird eine Abspielumgebung aufgebaut.

Das Werk wird in die Umgebung kopiert.

Das Werk wird wiedergegeben.

Es wird eine Qualitätskontrolle mittels einer Beurteilungsmatrix* durchgeführt.

Wird die Wiedergabe für ausreichend empfunden, wird die Umgebung in einem geeigneten Format dokumentiert.

Anschließend wird ein Screencast durchgeführt. Hierfür wird die Screencast-Software konfiguriert und anschließend die Aufnahme durchgeführt und die Videodatei erstellt.

Umgebung und Test werden im Wiki dokumentiert.

Die Beschreibung der Umgebung, die Videodatei und die Werkdateien werden in einem temporären Verzeichnis gespeichert.

Wird die Wiedergabe für nicht ausreichend empfunden, wird eine Fehleridentifikation durchgeführt.

Anschließend wird entschieden, ob eine Modifikation der Umgebung durchgeführt wird.

Wird eine Modifikation durchgeführt, so beginnt der Prozess mit einer angepassten Beschreibung der Umgebung.

Wird keine Modifikation durchgeführt, so werden die Ursachen für die gescheiterte Wiedergabe im Wiki dokumentiert. Die Ursachen werden in einem geeigneten Format dokumentiert. Diese Beschreibung sowie die Werkdateien werden in einem temporären Verzeichnis gespeichert.

4.f. Wiki II

Nachdem der Archivierungsprozess abgeschlossen ist und die nötigen Informationen im Wiki vermerkt sind, werden die Autoren und Experten informiert.

Diese können für einen noch zu bestimmenden Zeitraum die Einträge bearbeiten und Informationen hinzufügen.

Nach diesem Zeitraum werden die Einträge abgeschlossen und in ihrer vorliegenden Form in Kallias (und somit in den SWB) und in die Metadaten zum Werk übernommen.

4.g. BagIt

Alle Dateien, die zuvor temporär gespeichert wurden, werden in einem BagIt-Container strukturiert abgelegt*.

Aus dem BagIt-Container wird ein tar-Archiv erstellt.

Das Archiv wird über SWBContent an das BSZ übertragen. *

Dem Archiv wird ein URN zugeordnet.

Der Vorgang wird in Datenbank I vermerkt, der URN wird dem Werk in Datenbank I zugeordnet.

Der Vorgang wird abgeschlossen.

4.h. BagIt Zugriff

Der Zugriff und die Darstellung der Inhalte erfolgt durch den Dienstleister BSZ.

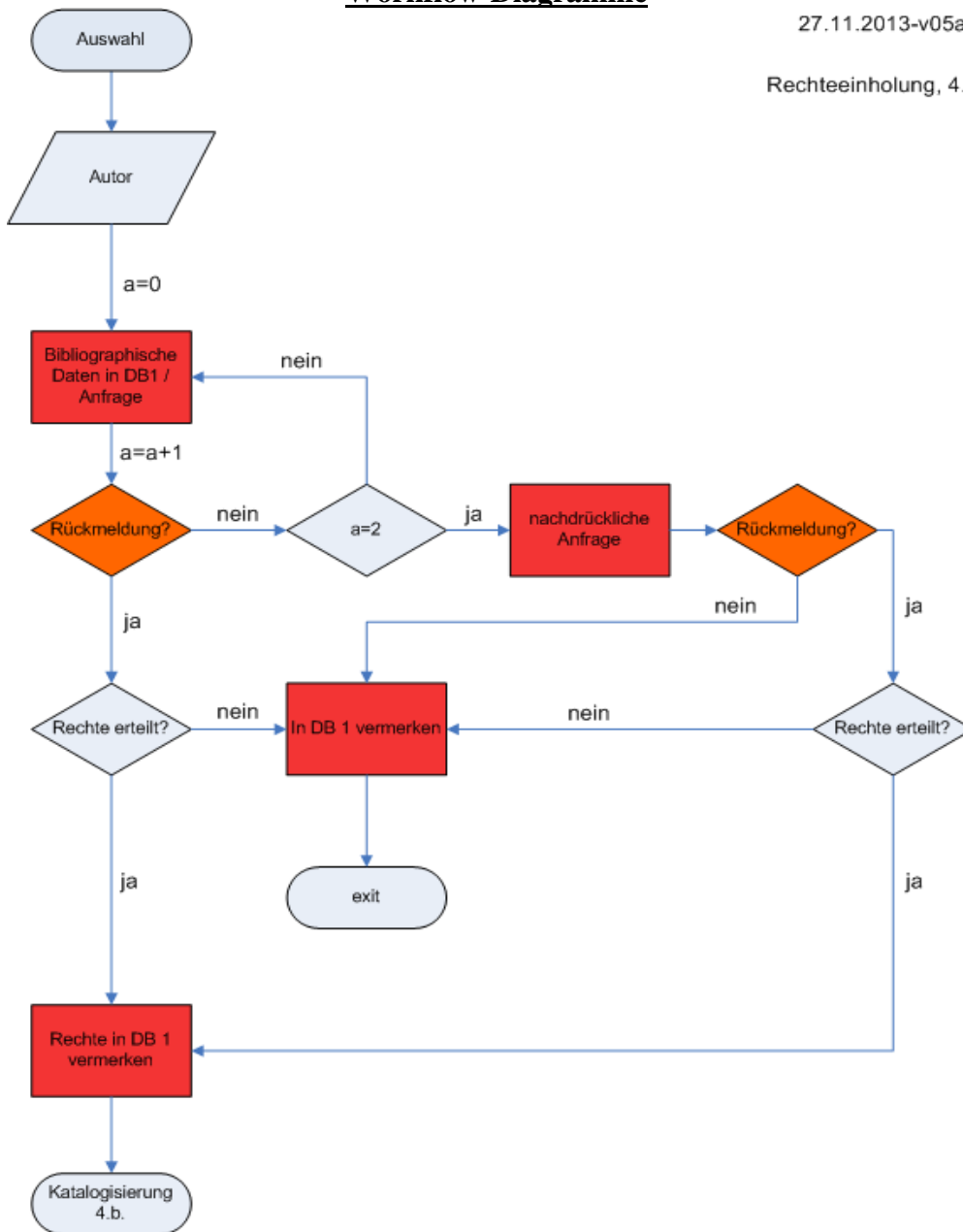
4.i. Emulation

-offen-

Workflow-Diagramme

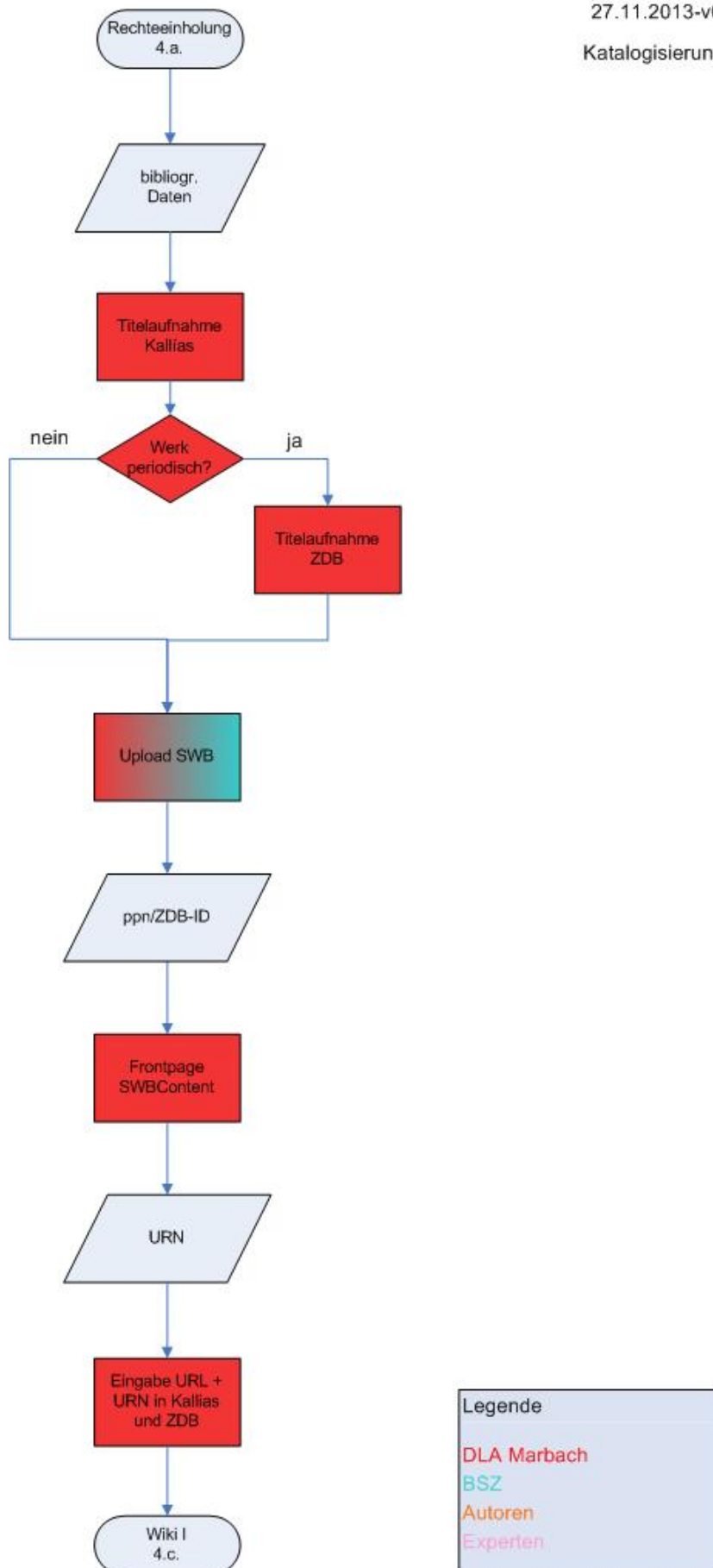
27.11.2013-v05a

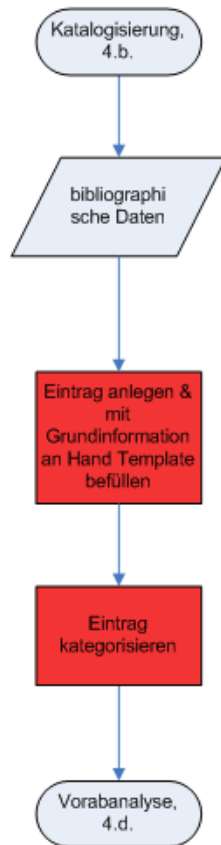
Rechteeinholung, 4.a.

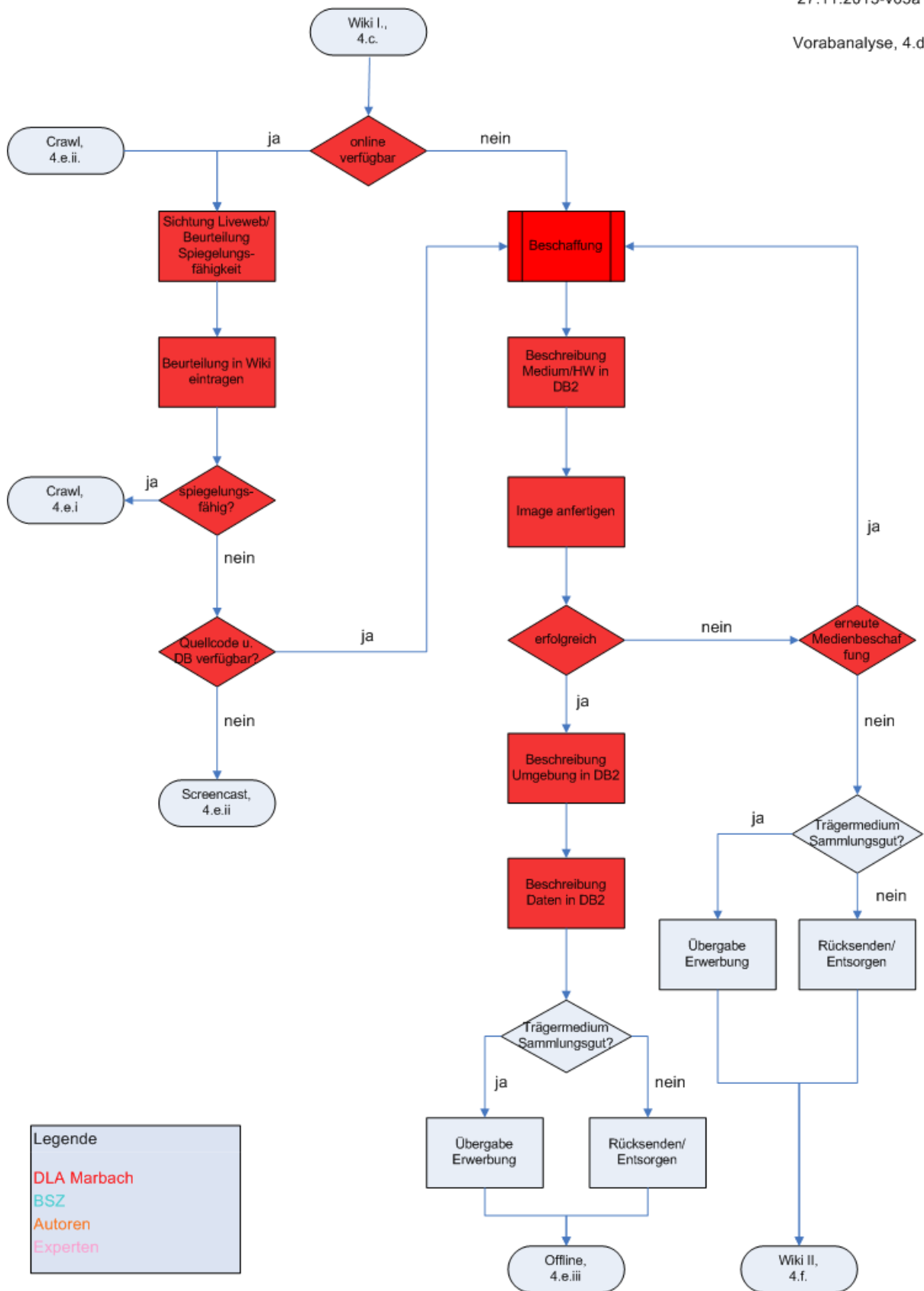


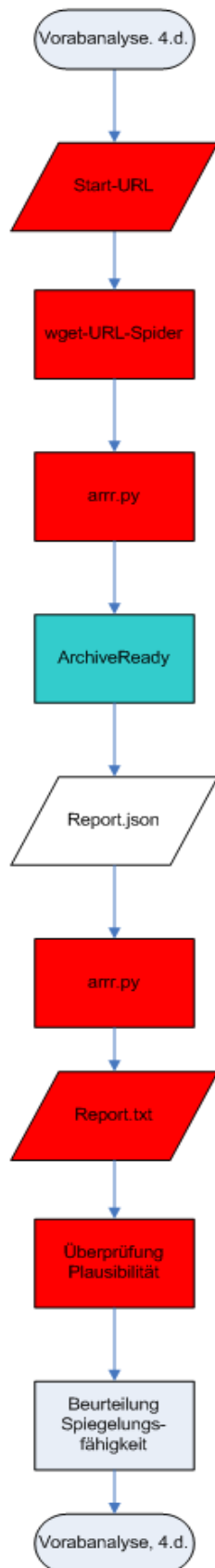
Legende

- DLA Marbach
- BSZ
- Autoren
- Experten

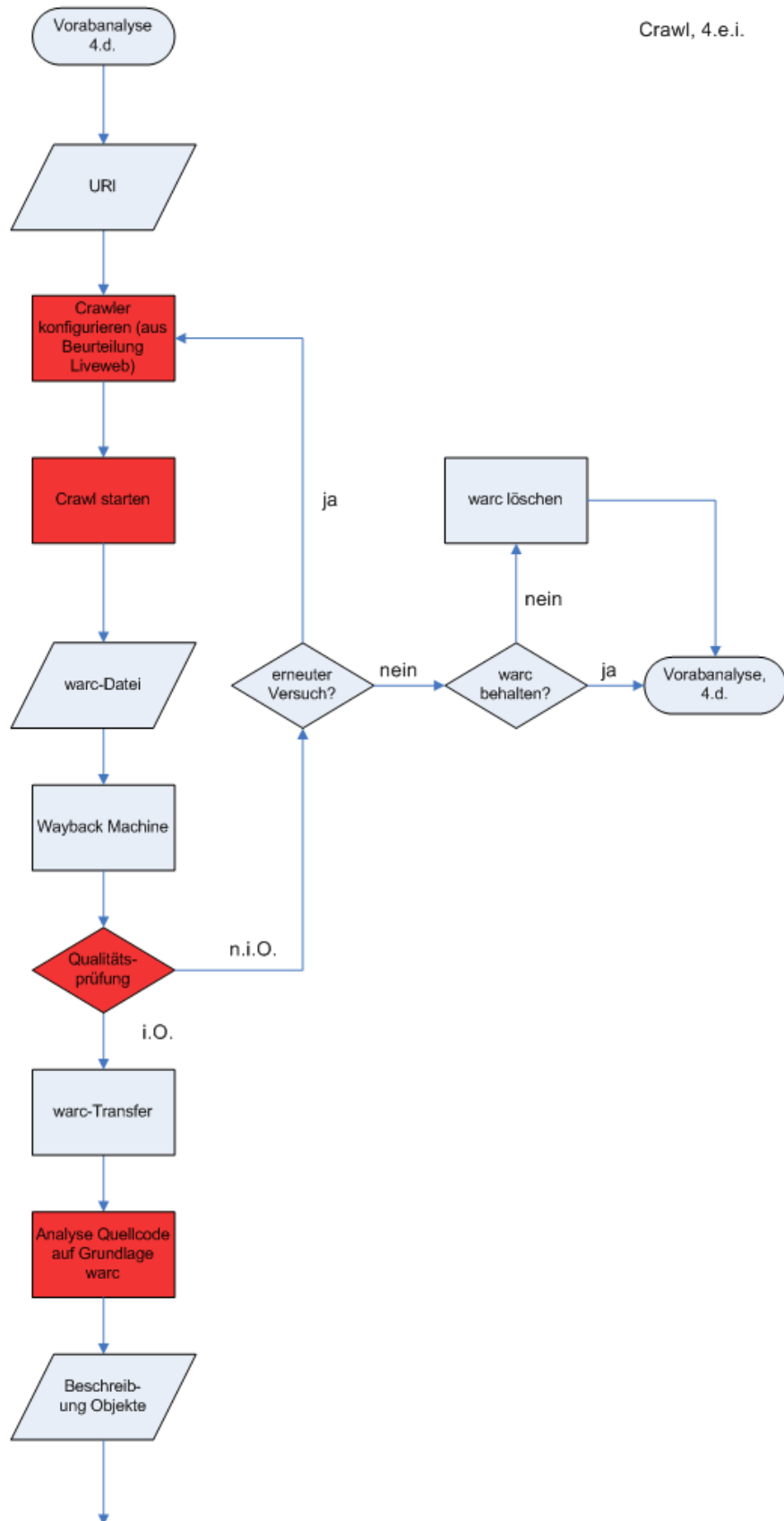






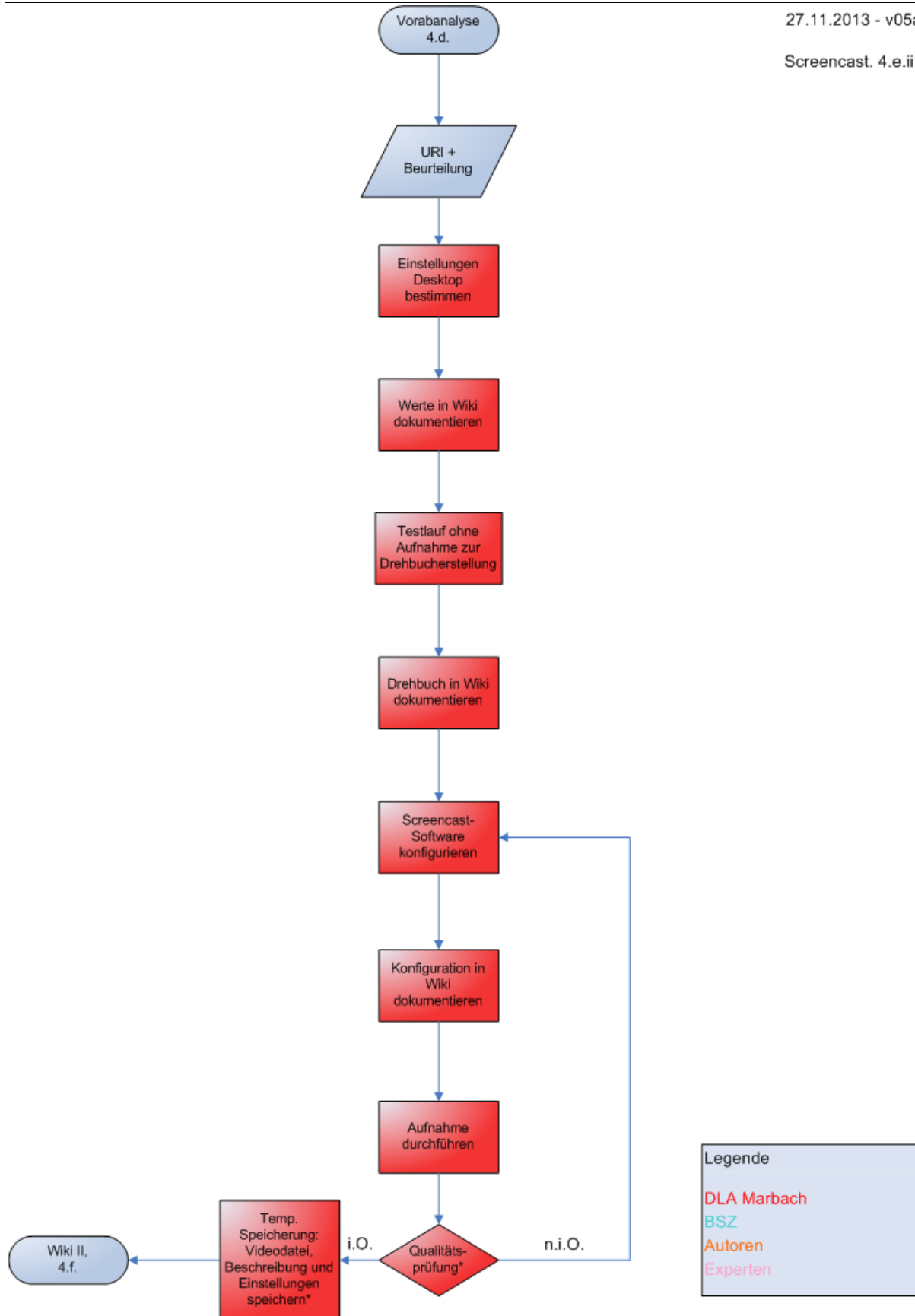


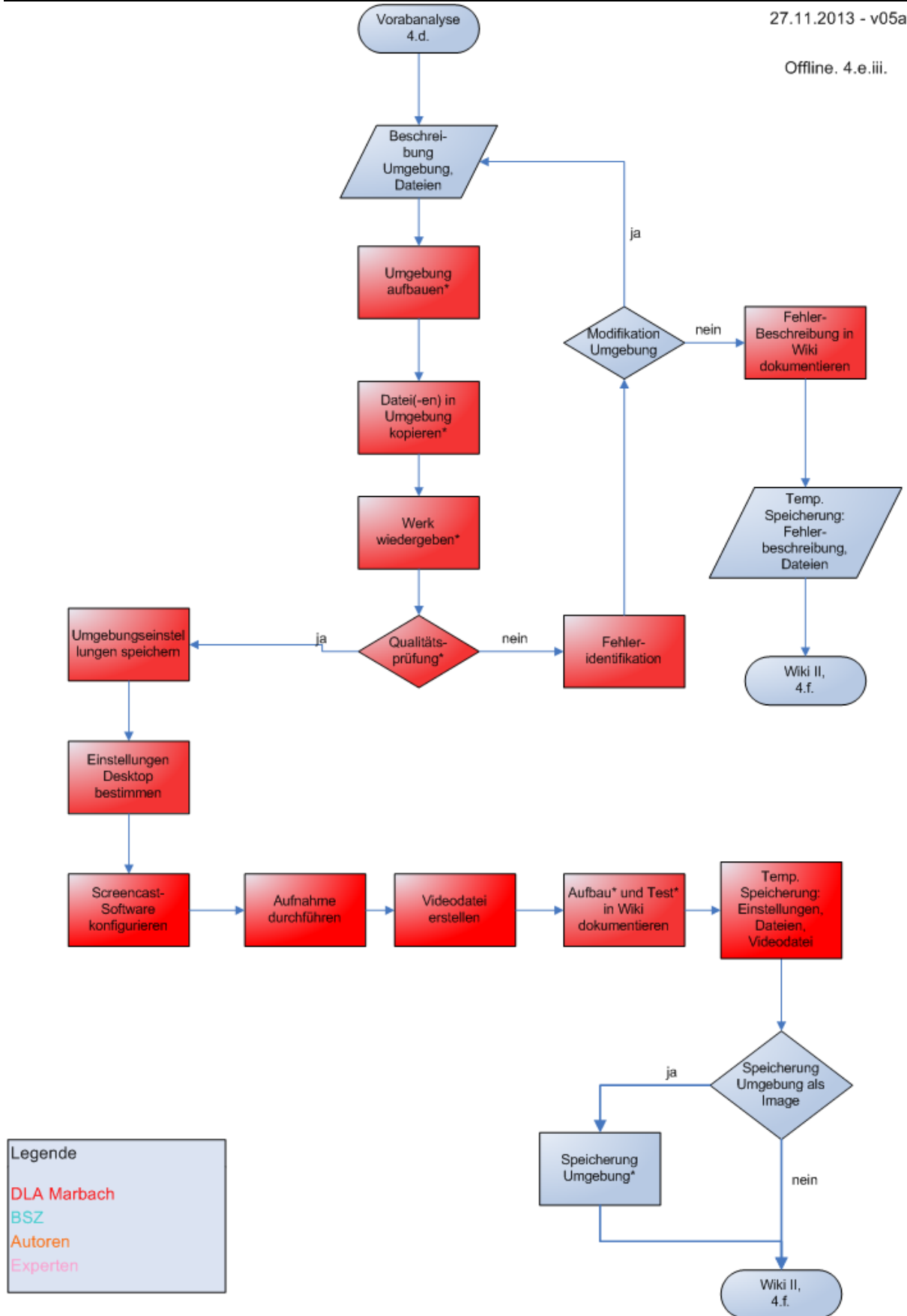
Legende
DLA Marbach
BSZ
Autoren
Experten
ArchiveReady





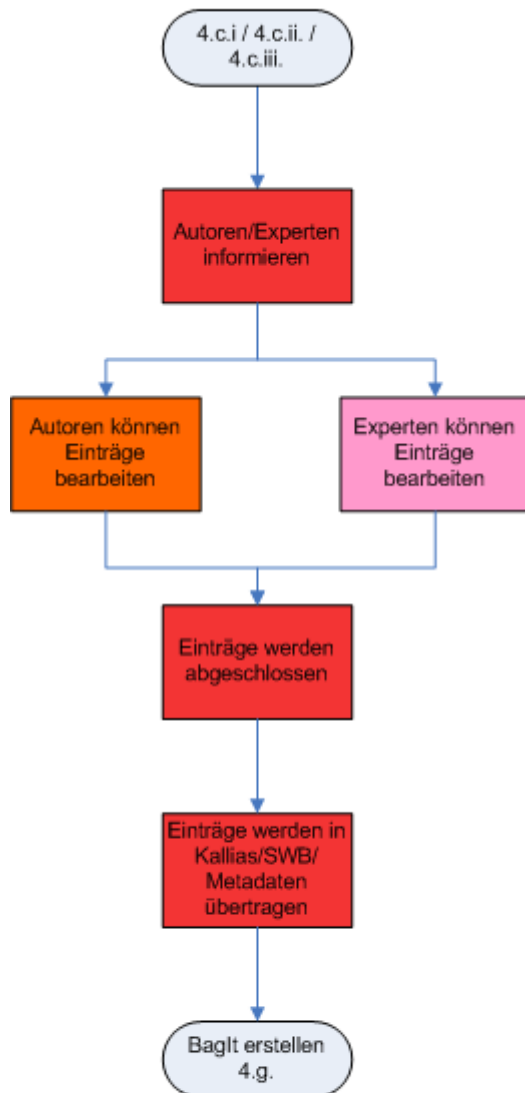
Legende
DLA Marbach
BSZ
Autoren
Experten

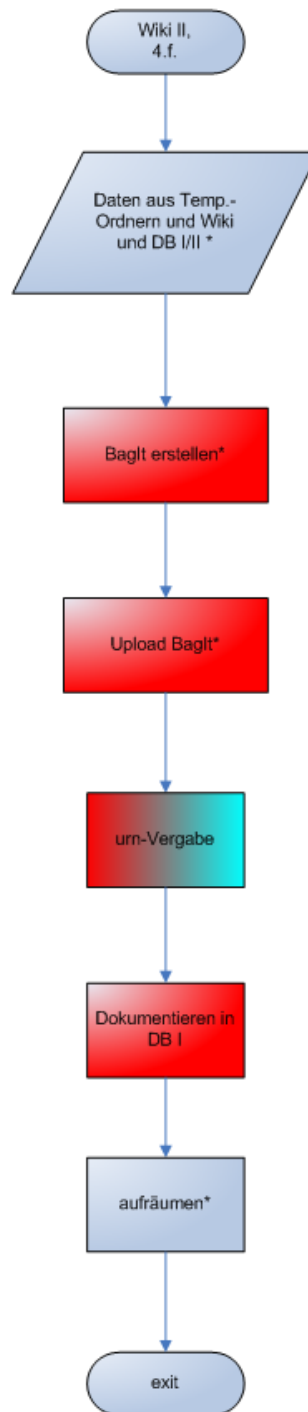




Legende

- DLA Marbach
- BSZ
- Autoren
- Experten





Der Zugriff auf Bags und die Darstellung der darin enthaltenen Daten erfolgt durch den Dienstleister BSZ.

- in Bearbeitung -