

Georg Mayr-Duffner

Erste Schritte in Goobi workflow mit Goobi-to-go

Handbuch Repositorienmanagement, Hg. v. Blumesberger et al., 2024, S. 437–453
<https://doi.org/10.25364/978390337423223>



Dieses Werk ist lizenziert unter einer Creative Commons Namensnennung 4.0 International Lizenz,
ausgenommen von dieser Lizenz sind Abbildungen, Screenshots und Logos.

Georg Mayr-Duffner, Wirtschaftsuniversität Wien, Bibliothek, georg.mayr-duffner@wu.ac.at |
ORCID iD: 0000-0002-8935-175X

Zusammenfassung

Viele Institutionen, die in größerem Ausmaß Digitalisierungsprojekte durchführen, verwenden mittlerweile das Open-Source-Tool Goobi workflow. Kernstück dieser Software ist das Workflow-Management, anhand dessen das digitale Objekt auf seinem Weg durch das System geführt wird – vom ersten Schritt, wie dem Anlegen eines Metadatensatzes, bis zum letzten, beispielsweise der Ablage in einem externen LZA-System. Auf der flexiblen Gestaltung dieses Workflow-Managements beruht wohl der Erfolg der Software, die mittlerweile von Bibliotheken, Archiven, Museen und anderen Einrichtungen unterschiedlichster Größe aus vielen Ländern in Europa, Afrika, Amerika und Asien angewendet wird: Zum einen skaliert sie sehr gut mit der Größe der Einrichtungen, zum anderen lassen sich damit gleichzeitig völlig unterschiedliche Projekte realisieren. Für die Präsentation der Daten aus Goobi workflow bietet sich Goobi viewer an, ebenfalls eine Open-Source-Software. Beide Tools sind voneinander unabhängig und können jeweils auch mit anderen Anwendungen kombiniert werden. Um Goobi einfach auszuprobieren, wird von der Firma Intranda GmbH, die die Software entwickelt, Goobi-to-go zur Verfügung gestellt. Dabei handelt es sich um ein Bündel aus Goobi workflow und Goobi viewer, das in wenigen Schritten auf einem herkömmlichen Arbeitsplatzrechner installiert wird. Dieser Beitrag erklärt anhand dieses Pakets die grundlegenden Funktionen und Konfigurationen von Goobi workflow.

Schlagwörter: Goobi; Digitalisierungssoftware; Workflow-Management

Abstract

First Steps with Goobi Workflow Using Goobi-to-go

Many institutions that carry out digitisation projects on a larger scale now use the open source tool Goobi workflow. At the heart of this software is the workflow management, which guides the digital object on its way through the system, from the first step, such as the creation of a metadata record, to the last step, such as filing in an external long-term preservation system. Libraries, archives, museums and other institutions of all sizes from many countries in Europe, Africa, America, and Asia are using the software. Its success is based on the flexible design of the workflow management: Firstly, it scales very well with the size of the facilities; on the other hand, completely different projects can be realized at the same time. Goobi viewer, also an open-source software, is ideal for presenting the data from Goobi workflow. Both tools are independent of each other and each of them works well in combination with third party software. Intranda GmbH, the company that has developed the Goobi software, provides Goobi-to-go for the purpose of trying out

Goobi. Goobi-to-go bundles Goobi workflow and Goobi viewer in a single installable package for use on a conventional workstation computer in just a few steps. This package is used in this contribution in order to explain the basic functions and configurations of Goobi workflow.

Keywords: Goobi; digitisation software; workflow management

1. Überblick

Der Name Goobi steht für ein Software-Ökosystem aus dem Bereich der Digitalisierung. Im Mittelpunkt stehen dabei die beiden Applikationen Goobi workflow und Goobi viewer. Hinter diesen Produkten steht das Göttinger Unternehmen Intranda, das die Software pflegt und weiterentwickelt. Goobi wird als Open-Source-Software entwickelt und nur wenige, besonders aufwändige Erweiterungen vertreibt Intranda unter einer kommerziellen Lizenz. Goobi workflow und Goobi viewer, sowie die Plugins, deren Entwicklung bereits bezahlt ist, stehen unter der GNU General Public License, version 2.0¹. So können die täglichen Entwicklungsschritte im Quelltext auf Github² nachvollzogen werden und die Community kann Korrekturen und Verbesserungen direkt als Code einbringen.

1.1. Dokumentation

Open Source bedeutet aber im Idealfall nicht nur, dass der Code frei verfügbar ist, sondern auch, dass die Dokumentation, also Handbücher, Anleitungen und Schulungsunterlagen offen verfügbar sind. Intranda geht auch hier den Open-Source-Weg konsequent und pflegt diese Unterlagen wie den Code auf Github und stellt sie ohne Hürden auf <https://docs.goobi.io> zur Verfügung. Diese Dokumente stehen unter einer CC-BY-NC-ND-Lizenz^{3,4}. Die Handbücher richten sich in eigenen Unterbereichen an die verschiedenen Zielgruppen und bemühen sich, eine der jeweiligen Zielgruppe entsprechend niederschwellige Darstellungsweise zu wählen. So bietet der Anwenderbereich von Goobi workflow eine einfache bebilderte Schritt-für-Schritt-Erklärung der Anwendungsoberfläche, während im Bereich Management

1 <https://www.gnu.org/licenses/old-licenses/gpl-2.0.en.html>

2 <https://github.com/intranda>

3 Siehe „Urheberrechte“ unter <https://docs.goobi.io/goobi-workflow-de/> und unter <https://docs.goobi.io/goobi-viewer-de/>

4 <http://creativecommons.org/licenses/by-nc-nd/4.0/>

davon ausgegangen wird, dass gewisse Kompetenzen, wie das Zurechtfinden in einem Dateisystem, vorhanden sind.

1.2. Goobi workflow

Steffen Hankiewicz, Geschäftsführer der Intranda GmbH und als Entwickler maßgeblich für Goobi workflow verantwortlich, bezeichnet Goobi workflow als „Workflow Tracking Tool“, welches seit 2004 entwickelt wird.⁵ In dieser Webanwendung werden Regeln festgelegt, nach denen die Abläufe in einem Digitalisierungs(teil)projekt stattfinden. Dafür wird der Weg eines Objekts von der Registrierung im System bis zum Abschluss des Digitalisierungsworkflows in kleine Schritte zerlegt, die einzelnen Aufgaben entsprechen, die einer Rolle zugeordnet werden können. Bei den Rollen kann es sich um Benutzerrollen, die mit bestimmten Berechtigungen verknüpft sind, handeln, aber auch um ein (externes) Script oder Computerprogramm im Fall eines automatisierten Schrittes. Im Durchlauf eines Objekts durch die Digitalisierungsroutine wird dann Schritt für Schritt der Workflow durchlaufen. Die/der Benutzer:in sieht dabei nur jeweils jene Aufgaben, die auch ihrer/seiner Rolle entsprechen. Die flexible Gestaltung mit Plugins erlaubt es, für die unterschiedlichsten Anforderungen passende Funktionalitäten innerhalb der Software bereitzustellen, beginnend beim Datenimport, über Seitenlayouterkennung und Qualitätskontrolle bis hin zum Export in ein Langzeitarchivierungssystem. Ebenso können die eingebauten Statistikfunktionen, die die Überwachung der Projektfortschritte erlauben, mittels Plugins erweitert werden.

1.3. Goobi viewer

Mit dem Goobi viewer entwickelt Intranda eine flexible Präsentationsplattform für digitale Objekte und Sammlungen. Viewer und workflow passen sehr gut zusammen, weil sie im selben Unternehmen aufeinander abgestimmt entwickelt werden, können aber jeweils auch mit anderen Softwarekomponenten von Drittanbietern kombiniert werden. Metadaten und Volltexte von Objekten, die an den Goobi viewer übergeben werden, werden von einem Apache Solr-Indexer indexiert und somit durchsuchbar gemacht. Um den Auftritt der digitalen Sammlungen mit verschiedenen Seitentypen systematisch gestalten zu können, enthält der Goobi viewer auch ein Content-Management-System. Neben diversen Exportoptionen bietet der viewer auch eine REST-Schnittstelle und eine IIIF-Presentation-API. Weitere besondere Features sind für den Bereich Crowdsourcing zu finden. Zum einen

⁵ Hankiewicz, S. (2018), S. 77.

sind sogenannte „Kampagnen“ eingebaut, mit denen der Öffentlichkeit die Möglichkeit gegeben wird, an einer bestimmten Aufgabe für einen begrenzten Zeitraum mitzuarbeiten. Zum anderen gibt es ein dezidiertes Crowdsourcing-Modul (kostenpflichtig unter einer kommerziellen Lizenz), über welches die Öffentlichkeit an der Transkription und an der inhaltlichen Erschließung der Digitalisate mitwirken kann.

1.4. Anwendungsbeispiele

Die oben beschriebene Flexibilität erlaubt es, die beiden Softwarekomponenten in unterschiedlichen Umgebungen einzusetzen. Goobi viewer bietet keine Benutzerschnittstelle für die Aufnahme neuer Objekte. Stattdessen erwartet es, von einem zweiten System mit Objekten und Metadaten versorgt zu werden, was aufgrund der Verwendung standardisierter Formate verschiedenste Kombinationen erlaubt. Denkbar wäre beispielsweise, Goobi viewer an ein Archivinformationssystem anzuhängen, um digitales Archivgut verfügbar zu machen. Goobi workflow andererseits wird gerne verwendet, um komplexe Ingestvorgänge – häufig im Zusammenhang mit Digitalisierungsaufgaben – zu organisieren und die Endprodukte an ein Repositorium zu übergeben⁶. Viele Institutionen entscheiden sich aber dafür, beide Komponenten gemeinsam einzusetzen. Je nach Konfiguration entsteht so eine Digitale Bibliothek, wie beispielsweise bei der digitalen Landesbibliothek Oberösterreich⁷ und den Digitalen Sammlungen der Herzogin Anna Amalia Bibliothek⁸ oder ein Repositorium wie im Fall der E-Medien der Arbeiterkammern und des ÖGB-Verlags⁹. In allen drei Fällen finden die Verwaltung von Metadaten und die Sammlungsorganisation in Goobi workflow statt, während der viewer für die Darstellung und den öffentlichen Zugang zuständig ist. Letzteres inkludiert auch Schnittstellen für den automatisierten Zugriff wie OAI-PMH.

⁶ Vgl. Frost, H. et al. (2019)

⁷ <https://digi.landesbibliothek.at/>

⁸ <https://haab-digital.klassik-stiftung.de>

⁹ <https://emedien.arbeiterkammer.at/>

2. Goobi-to-go

2.1. Was ist Goobi-to-go?

Üblicherweise werden Goobi workflow und Goobi viewer auf zwei getrennten Servern installiert. Diese Aufstellung ist auch sinnvoll, um die workflow-Applikation, die häufig nur innerhalb der Institution zugänglich sein soll, von der allgemein zugänglichen Präsentationssoftware zu trennen. Mit Goobi-to-go¹⁰ bietet Intranda eine sehr einfach zugängliche Möglichkeit, Goobi workflow und Goobi viewer zusammen zu testen.

Setup und Wartung eigener Server für das Testen bedeuten einen großen Aufwand, der mitunter in keinem Verhältnis zum Aufwand der Tests steht. Dieses Dilemma löst Goobi-to-go, indem es eine vollständige lauffähige Anwendungsumgebung mitbringt. Es enthält Goobi viewer und Goobi workflow sowie alle für den Betrieb notwendigen Programme in der jeweils erforderlichen Version, insbesondere eine Java-Laufzeitumgebung, ein H2-Datenbankmanagementsystem und einen Apache Tomcat-Server. Das Softwarepaket ist damit unabhängig von eventuell am Rechner installierter Software und kann nach dem Download und Entpacken direkt gestartet werden. Auf dem Rechner wird nichts geändert, außer dass Speicherplatz belegt wird. Wird es nicht mehr benötigt, wird der Goobi-to-go-Ordner einfach gelöscht, ohne dass Spuren davon zurückbleiben. Da eine Goobi-to-go-Anwendung keine Einstellungen und Abhängigkeiten außerhalb des eigenen Ordners kennt, können auf einem Rechner auch mehrere „Goobi-to-gos“ mit unterschiedlichen Konfigurationen oder in unterschiedlichen Versionen liegen, was das Testen unterschiedlicher Szenarien erheblich vereinfacht.

2.2. Goobi-to-go starten

Um Goobi-to-go zu starten, muss das Softwarepaket von der Seite <https://goobi.io/goobi-to-go/> heruntergeladen werden. Aktuell werden Pakete für Windows, macOS und Linux angeboten¹¹. Es handelt sich dabei um eine Zip-Datei im Umfang von gut 1,1 GB. Nach dem Download wird die Datei entzipped. Der dabei extrahierte Ordner g2g muss jedenfalls an einem Ort abgelegt werden, von dem aus der/die User:in berechtigt ist, Programme auszuführen. In diesem Ordner befindet sich ein Starter, der auf Linux GoobiToGo, auf Windows GoobiToGo.exe und auf

¹⁰ <https://goobi.io/goobi-to-go>

¹¹ Dieser Beitrag basiert auf dem Goobi-to-go Snapshot 20211220. Der darin enthaltene Goobi workflow trägt die Versionsnummer 21.11.5, der Goobi viewer trägt die Versionsnummer 21.12.

macOS GoobiToGo.command heißt. Nach einem Doppelklick auf den Starter¹² wird die Anwendung gestartet. Dies bedeutet, dass ein Server¹³ auf dem Rechner gestartet wird, der die Webapplikationen viewer und workflow bereitstellt und diese beiden mit einer Datenbank versorgt. Sobald der Startvorgang abgeschlossen ist, wird das Startfenster geschlossen, dafür erscheint ein Goobi-Symbol in der Task- bzw. Menüleiste. Über dieses Symbol lassen sich viewer und workflow öffnen, sowie der Server herunterfahren. Seit Ende September 2021 gibt es hier zudem Zugriff auf Hilfe- und Infoseiten, die ebenfalls im heruntergeladenen Paket integriert sind.

2.3. Goobi workflow erkunden

Wenn der Goobi-to-go-Server gestartet ist, kann Goobi workflow mit Klick auf den entsprechenden Menüeintrag im Goobi-to-go-Menü geöffnet werden. Alternativ kann im Webbrower auch die Adresse der Anwendung direkt eingegeben werden. Diese ist vorkonfiguriert als <https://localhost:8888/goobi>.¹⁴ Im Browser öffnet sich die Anmeldeseite von Goobi workflow. Hier kann man sich mit einem der vorgefertigten Benutzerkonten anmelden, die auf der Seite <https://goobi.io/goobi-to-go> dokumentiert sind. Diese Benutzerkonten unterscheiden sich in ihren Rollen und Berechtigungen.

Loggt man sich mit dem Adminstrator:innenaccount „goobi“ ein, landet man auf einer Dashboardseite, die einen Überblick über die laufenden Projekte und Schnellzugriff auf wichtige Funktionen wie Vorgangssuche und Produktionsvorlagen bietet. Zudem wird der Newsfeed von Intranda eingeblendet. In der permanenten Menüleiste finden sich Zugriffsmöglichkeiten auf die einzelnen Funktionalitäten der Anwendung, für die der Account berechtigt ist:

¹² Die Anwendung kann auch von der Kommandozeile aus gestartet werden. Dies ist auf <https://goobi.io/goobi-to-go/> für die unterschiedlichen Betriebssysteme dokumentiert.

¹³ Ein Server bezeichnet in diesem Fall ein Programm, das anderen Programmen Dienste zur Verfügung stellt. Dieser Server stellt einerseits den Webapplikationen ein Datenbanksystem zur Verfügung, andererseits den Browsern auf dem Rechner die Applikationen Goobi workflow und Goobi viewer selbst.

Obwohl es sich um einen Webserver handelt, bedeutet das nicht automatisch, dass von fremden Rechnern aus auf die Goobi-to-go-Anwendungen zugegriffen werden kann. Damit das möglich ist, müssen bewusst die entsprechenden Ports in der Firewall, die jedes moderne Betriebssystem von Haus aus mitbringt, geöffnet werden. Standardmäßig sind diese zu.

¹⁴ Die Portadressen der einzelnen Komponenten können in der Konfigurationsdatei g2g/config/g2g_config.xml angepasst werden, dies kann nötig sein, wenn es einen Konflikt mit anderen Anwendungen auf dem Rechner gibt, die dieselben Ports verwenden. Nach der Portänderung muss Goobi-to-go neu gestartet werden.

- „Meine Aufgaben“ öffnet die Aufgabenliste mit allen Aufgaben, die den Benutzerrollen des/der eingelogten User:in zugewiesen sind und entweder offen sind (also noch nicht von einem/einer Benutzer:in übernommen wurden) oder bereits von der/dem gerade eingeloggten Benutzer:in übernommen aber noch nicht abgeschlossen wurden.¹⁵
- „Workflow“ öffnet ein Dropdown-Menü, das Zugriff auf alle Funktionen rund um die Vorgänge der einzelnen Objekte erlaubt. Insbesondere lassen sich hier die Vorgänge auflisten oder suchen, zum anderen werden unter „Produktionsvorlagen“ die Workflows konfiguriert und neue Vorgänge anhand einer Produktionsvorlage angelegt.
- Unter „Administration“ finden sich alle Konfigurationsmöglichkeiten für Goobi workflow, sofern es dafür ein grafisches Interface gibt.
- „Controlling“ erlaubt Zugriff auf die Statistiken.
- Das Goobi-Logo und der Menüpunkt „Dashboard“ führen beide zum Dashboard zurück.

2.4. Das zentrale Objekt – der „Vorgang“

Das Tracking der Digitalisierung eines Objekts in Goobi workflow findet in Form eines „Vorgangs“ statt. Dies ist das zentrale Objekt, das unter einem eindeutigen Vorgangstitel und einer ID alle zugehörigen Daten, wie Metadaten, Masterdateien und Derivate speichert. In Goobi wird der Vorgang üblicherweise mit dem Vorgangstitel identifiziert, der meist automatisch aus Teilen der Autor:innennamen, des Titels und einem Identifier aus den Metadaten gebildet wird¹⁶. Dadurch erhält es einen sprechenden Teil, der die Identifizierung durch Bearbeiter:innen erleichtert. Will man sich den angelegten Vorgang im lokalen Speicher des PCs ansehen, benötigt man die Vorgangs-ID, die beim Anlegen des Vorgangs von Goobi workflow hochgezählt wird. Um sie sichtbar zu machen, kann in der Liste der Vorgänge die entsprechende Spalte angewählt werden. Im Speicher sind die Vorgänge unter g2g/workspace/workflow/metadata/ zu finden. Der Vorgangsordner ist mit der Vorgangs-ID benannt und enthält eine METS-Datei meta.xml mit den Metadaten des Vorgangs. Daneben gibt es je nach Bedarf Ordner für diverse Objekte wie bspw. Bilder (images) oder OCR-Daten (ocr). Verweist der Vorgang auf eine andere Hierarchiestufe, beispielsweise einen Zeitschriftenband, wenn im Vorgang ein Zeitschriftenheft digitalisiert wird, wird für dieses sogenannte „Anker-Objekt“ eine eigene METS-Datei meta_anchor.xml abgelegt.

¹⁵ In Goobi-to-go haben die beiden Administratorkonten in den Voreinstellungen keine Aufgabenbereiche zugeordnet, weshalb die Listen hier leer sind.

¹⁶ In Goobi-to-go sind bereits drei Mustervorgänge in zwei Projekten angelegt.

2.5. Einen Vorgang durchspielen

In Goobi-to-go sind bereits zwei Muster-Produktionsvorlagen angelegt, anhand derer die Workflows direkt ausprobiert werden können. Im Menü unter „Workflow > Produktionsvorlagen“ werden der „Manuscript_Workflow“ für das „Manuscript_Project“ und der „Sample_Workflow“ für das „Archive_Project“ aufgelistet. Mit Klick auf das Stiftsymbol gelangt man in die Bearbeitungsansicht der Vorlage und sieht im Block „Abfolge der Aufgaben“, welche Schritte vorgesehen sind. Bereits in dieser Übersicht lässt sich erkennen, ob ein Schritt durch eine:n Bearbeitenden auszuführen ist oder automatisch abgearbeitet wird. In diesem Beispiel ist Schritt 5 „Image processing“ ein automatischer Schritt, was durch das Zahnradsymbol gekennzeichnet wird. Zudem kommt ein Plugin zum Einsatz, ersichtlich aus dem Puzzleteil-Symbol in der Spalte „Aktionen“.

Weitere Details zu dem jeweiligen Schritt werden sichtbar durch Klick auf den Pfeil in der Spalte „Titel“. Hier interessiert uns nun besonders die Zeile „Zugewiesene Rechte“: Dabei handelt es sich um die Benutzergruppe, die nötig ist, um den Schritt durchführen zu dürfen. Wollen wir einen Vorgang durchspielen, haben wir zwei Möglichkeiten: Wir können ein Benutzerkonto allen benötigten Benutzergruppen zuweisen oder wir wechseln für jeden Schritt das Konto. In letzterem Fall ist es möglicherweise sinnvoll, mit zwei Browsern oder einem normalen und einem Inkognitofenster desselben Browsers parallel zu arbeiten. So kann man in einem Fenster ständig mit dem Admin-Account eingeloggt sein und beobachten, was passiert, während man im anderen mit den jeweils benötigten Funktionskonten die Schritte abarbeitet. Welches Konto welchen Gruppen zugewiesen ist, kann in der Liste unter dem Menüpunkt „Administration > Benutzer“ kontrolliert werden.

Über den aktuellen Zustand eines Schritts informiert der Status. Dieser wird in einer Art Fortschrittsbalken farblich gekennzeichnet:

- Gesperrt (rot): Der Schritt kann noch nicht ausgeführt werden, da der vorhergehende noch nicht abgeschlossen wurde.
- Offen (orange): Dieser Schritt ist zur Bearbeitung offen und befindet sich in der Aufgabenliste aller berechtigten Nutzer:innen.
- In Bearbeitung (gelb): diese Aufgabe wurde von einem/einer Benutzer:in übernommen oder wird gerade von einem Skript automatisch abgearbeitet.
- Abgeschlossen (grün): Diese Aufgabe wurde abgeschlossen.
- Fehler (rot schraffiert): Bei der Bearbeitung der Aufgabe ist ein Fehler aufgetreten. Bis zur Behebung des Fehlers ist der Vorgang angehalten. Eine Möglichkeit, die Goobi dazu bietet, ist, eine Korrekturmeldung an jene Benutzer:innengruppe zu senden, die den Fehler beheben kann (beispielsweise

durch erneutes Scannen). Für die Fehleranalyse hilfreich sind die Einträge im Vorgangslog, das bei den einzelnen Aufgaben und auf der Übersichtsseite des Vorgangs angezeigt wird.

Die Schritte und die benötigte Benutzergruppe für den „Manuscript_Workflow“ sind:

Nr.	Titel	Benutzer:innengruppe	Konto
1	Data Import	Project Management	testprojectmanagement
2	Get manuscript from book depot	Book managing officers	testbookmanager
3	Scanning	Scanning officers	testscanning
4	Quality Control	Quality control officers	testqc
5	Image processing	(automatischer Schritt)	—
6	Metadata enrichment	Metadata officers	testmetadata
7	Export to viewer	(automatischer Schritt)	—
8	Bring manuscript back to book depot	Book managing officers	testbookmanager

Die zweite Produktionsvorlage „Sample_Workflow“ unterscheidet sich hier nur dadurch, dass die Tätigkeit des Book managing officers (Werk ausheben bzw. zurückstellen) ausgelassen werden¹⁷.

Schritt 1 – Data Import

Die vorliegende Konfiguration geht davon aus, dass ein Werk digitalisiert wird, das in einem Bibliothekskatalog verzeichnet ist. Vorkonfiguriert sind der Katalog der Library of Congress und der K10Plus-Verbundkatalog. Mit einem Konto, das der Benutzer:innengruppe „Project Management“ zugeordnet ist, oder auch mit einem der Admin-Konten kann ein neuer Vorgang angelegt werden, indem am Dashboard im Produktionsvorlagen-Widget bei der gewünschten Produktionsvorlage auf den blauen Button mit dem Dokumentensymbol (Titel: „Einen Vorgang auf Basis dieser Produktionsvorlage anlegen“) geklickt wird (nicht auf den weißen Pfeil!).

Daraufhin öffnet sich ein Formular, das eine „Suche im Opac“ anhand einer ID oder eines Barcodes erlaubt. Für diese Demonstration wird die ID 150899661 aus dem K10Plus-Katalog in das entsprechende Feld eingegeben und mit „Übernehmen“ bestätigt. Goobi holt sich die Daten über die SRU-Schnittstelle und befüllt damit das

¹⁷ Hier wäre es wünschenswert, mehr bzw. vor allem unterschiedlichere Workflows vorinstalliert mitzubekommen, mit denen eine größere Bandbreite der Funktionalitäten demonstriert werden kann.

Formular. Händisch zu befüllen ist das Feld „Autoren“¹⁸, damit auch der Vorgangstitel diese Komponente enthält. Grundsätzlich müssen alle mit * gekennzeichneten Felder befüllt werden, insbesondere muss eine Digitale Sammlung ausgewählt werden, allerdings wird der Vorgangstitel, wenn er freigelassen wird, automatisch aus den Feldern „ATS“ und „Identifier digital“ erzeugt, wobei „ATS“ wiederum aus den Inhalten der Felder „Autoren“ und „Titel“ erzeugt wird (ATS = Autor-Titel-Schlüssel). Die Angabe der geschätzten Seitenzahl wird für die Berechnung des Projektfortschritts in den Statistiken benötigt. Beim Abspeichern werden allenfalls die Felder ATS und Vorgangstitel erzeugt und der Vorgang im Speicher abgelegt. Zum Abschluss des Schritts kann noch ein Laufzettel im PDF-Format erzeugt werden.

Schritt 2 – Get manuscript from book depot

Dieser Schritt ist dazu gedacht, dass ein/e Bearbeiter:in ein Werk aushebt. Der Account testbookmanager zeigt nur zwei Menüpunkte an – neben dem Dashboard „Meine Aufgaben“. Dort ist der neu erstellte Vorgang mit der Aufgabe „Get manuscript from book depot“ zu finden. Die Aufgabe wird mit einem Klick auf den blauen Button mit dem Häkchensymbol (Titel „Die Bearbeitung dieser Aufgabe übernehmen“) übernommen. Im nächsten Fenster werden die Aufgabendetails angezeigt. Hier können im Widget „Vorgangslog“ allgemeine Kommentare und Meldungen verfasst werden. Im Widget „Mögliche Aktionen“ kann die erfolgreiche Erledigung mit „Die Bearbeitung der Aufgabe abschließen“ quittiert werden, insbesondere kann an dieser Stelle aber auch eine Fehlermeldung an einen der vorhergehenden Schritte abgesetzt werden. Die Aufgabe scheint dann in der Aufgabenliste jener Personen auf, die der Benutzer:innengruppe des gewählten Schritts angehören, oder die Bearbeitung der Aufgabe kann abgebrochen werden, wodurch sie wieder in der Aufgabenliste der für diesen Schritt zuständigen Bearbeiter:innen aufscheint.

Die hier beschriebenen Funktionen gelten bei jeder Übernahme einer Aufgabe. Im Folgenden wird nicht mehr darauf eingegangen.

¹⁸ Warum an dieser Stelle keine Automatisierung stattfindet, ist schleierhaft und seitens Intranda vielleicht zu überdenken. Unter Zuhilfenahme des Opac SRU Plugins (<https://github.com/intranda/goobi-plugin-opac-sru>) ist es jedenfalls möglich, die Übersetzung der OPAC-Daten granular zu steuern und auch hier die Daten automatisch zu übernehmen.

Schritt 3 – Scanning

In diesem Schritt werden die gescannten Bilddateien hochgeladen. Im Widget „Dateien“ wird mit Klick auf „Dateien auswählen“ ein Dateibrowser geöffnet. Hier können nun die zu übertragenden Bilder ausgewählt werden¹⁹, im Reiter „Übersicht“ kann das Ergebnis der Übertragung kontrolliert werden. Das Widget „Eigenschaften“ erlaubt die Angabe von Details zum Scavorgang – voreingestellt ist die Frage nach dem Öffnungswinkel des zu scannenden Objekts und nach der Schreibrichtung (RTL = right to left, wie beispielsweise Hebräisch oder Arabisch).

Schritt 4 – Quality Control

In diesem Schritt soll geprüft werden, ob die gescannten und hochgeladenen Bilder den Qualitätskriterien entsprechen. In diesem Schritt kommt das Plugin „Bildkontrolle“ zum Einsatz. Bei der Übernahme der Aufgabe sieht der/die Bearbeiter:in daher im Widget „Mögliche Aktionen“ einen zusätzlichen Button mit Puzzlesteinsymbol mit der Beschriftung „Plugin: Bildkontrolle“. Das Plugin zeigt eine Galerie der gescannten Bilder an und bietet grundlegende Bildkontrollfunktionen (Zoom und Rotation). Mit „Ergebnisse speichern und abschließen“ wird die Aufgabe abgeschlossen. Sie kann hier auch wieder abgebrochen werden und, wie bei Schritt 2 beschrieben, dem Scanning mit einer Korrekturmeldung zurückgegeben werden.

Schritt 5 – Image Processing

In diesem automatischen Schritt werden von den Masterdateien die Derivate für die weitere Verwendung erstellt, beispielsweise werden hier aus TIFF-Dateien JPG-Dateien generiert.

Schritt 6 – Metadata enrichment

In diesem Schritt werden die Metadaten im METS-Editor eingegeben. Bei der Übernahme der Aufgabe gibt es unter „Mögliche Aktionen“ den Button „Metadaten bearbeiten“, mit dem der Editor geöffnet wird. Das Editorfenster ist dreigeteilt: links wird der Strukturabaum dargestellt, wo das aktuell zu bearbeitende Strukturelement ausgewählt wird. Rechts werden die Digitalisate angezeigt. In der Mitte befindet sich der eigentliche Editor, wo verschiedene Arten von Metadaten bearbeitet werden können. Vier Bearbeitungsmodi kennt der viewer, die über Reiter zugänglich sind:

¹⁹ In Goobi-Installationen ist es sinnvoll, dass Scan-Mitarbeiter:innen ihr Arbeitsverzeichnis am Server auf ihrem Arbeitsplatzrechner als Netzlaufwerk einbinden, um größere Datenmengen leichter zu übertragen.

Im ersten Reiter wird die Paginierung des Digitalisats festgelegt. Die Nummerierung kann Seite für Seite festgelegt oder automatisch durchgezählt werden, wobei auch ungezählte Seiten, fingeierte Paginierung und Doppelseiten berücksichtigt werden können.

Im zweiten Reiter, „Strukturdaten“, können Strukturen innerhalb des Werks identifiziert werden. Dabei kann es sich um physische Strukturen wie Buchdeckel oder Farbkeil handeln, aber auch um inhaltliche wie Kapitel, Verzeichnisse oder Abbildungen. Die Lokalisierung ist dabei nicht auf die Seitenangabe beschränkt, sondern es kann der genaue Bildausschnitt im Digitalisat identifiziert werden. Zusätzlich kann jedem Strukturelement auch ein Satz Metadaten mitgegeben werden. In welchem Umfang, hängt davon ab, welche Metadaten für das jeweilige Element im Regelsatz (s. Abschnitt Konfiguration/Der Regelsatz, \$\$\$S. 11\$\$\$) definiert sind.

Der dritte Reiter, „Metadaten“, dient zur Eingabe der bibliographischen Metadaten. Dieser ist im Prinzip vorausgefüllt mit den Daten aus dem Katalog, allerdings wird hier der Name nicht mit übernommen und muss gesondert eingegeben werden. Welche Metadaten eingegeben werden können, wird durch den Regelsatz für den Dokumententyp wie auch für jedes Strukturelement bestimmt. Eine Verknüpfung mit Normdaten ist vorgesehen, implementiert sind in Goobi-to-go die GND, VIAF, KulturNav und Geonames. Wie ein Feld dargestellt wird, ob als Freitextfeld oder mit kontrolliertem Vokabular, wird über die Datei g2g/workspace/workflow/config/goobi_metadataDisplayRules.xml kontrolliert.

Der vierte Reiter bietet die Möglichkeit, ein Bild von einer beliebigen Stelle durch ein neues auszutauschen, beispielsweise, um es durch einen Scan mit einer besseren Qualität zu ersetzen.

Der Metadateneditor bietet ein eigenes Einstellungsmenü, mit dem die Ansicht nach Bedarf geändert werden kann. Außerdem können die eingegebenen Daten vor dem Speichern validiert werden. In diesem Fall wird Feld für Feld die Eingabe mit einem allenfalls im Regelsatz vorhandenen Validierungscode abgeglichen.

Mit Klick auf das Haussymbol ganz rechts oben und „Speichern und zurück“ kehrt man zur Aufgabenübersicht zurück, wo man die Bearbeitung wie gewohnt abschließen kann.

Schritt 7 – Export in viewer

Wieder ein automatischer Schritt – die Daten werden an den viewer übergeben, wo sie indexiert werden und danach sofort angezeigt werden können.

Schritt 8 – Bring manuscript back to book depot

Mit dieser Aufgabe wird im Vorgang dokumentiert, dass das Werk wieder zurückgestellt wird.

2.6. Konfiguration

Hier werden wichtige Konfigurationen und Konfigurationsdateien kurz beschrieben. Die Konfigurationsdateien sind durchgängig XML-Dateien, die relativ einfach angepasst werden können. Die bereits erwähnte Flexibilität bedingt allerdings auch eine gewisse Komplexität. Mit der in Goobi-to-go bereits vorliegenden Konfiguration können Standardfälle sicher weitestgehend, allenfalls mit wenigen Änderungen, bearbeitet werden. Im Echtbetrieb finden sich dann aber fast immer Gegebenheiten, die eine tiefergehende Beschäftigung mit der Konfiguration verlangen. Hinzu kommt die Vielzahl an Plugins, die den Funktionsumfang stark erweitern, aber ebenfalls korrekt konfiguriert werden müssen. In Einzelfällen sind dazu auch Kenntnisse in XSLT nötig.

2.6.1. Der Regelsatz

Die zentrale Konfiguration für die Metadaten, die in einem Vorgang zur Verfügung stehen, findet im Regelsatz statt. Hier wird festgelegt:

- welche Metadatentypen (Titel, Körperschaft, etc.) existieren, sowie ihre Übersetzung in die gewünschten Anzeigesprachen. Darüber hinaus können hier für jeden Typ Validierungsregeln in Form von regulären Ausdrücken definiert werden.
- Metadatengruppen, die verwendet werden sollen – beispielsweise für ein strukturiertes Titelelement.
- die Strukturtypen, die verwendet werden können. Dabei handelt es sich sowohl um Dokumenttypen (Buch, Artikel,) als auch um Strukturtypen, die einen Teil eines Werks beschreiben (Buchdeckel vorne, Deckblatt, Abbildung). Wie bei den Metadatengruppen wird hier definiert, welche der Metadatentypen im jeweiligen Strukturtyp zulässig sind, ob sie obligatorisch oder optional und ob sie wiederholbar sind.
- Formatierungsregeln für die Umsetzung der zuvor festgelegten Regeln in einem Metadatenformat. Der vordefinierte Regelsatz in Goobi-to-go enthält Formatierungsregeln für die Umsetzung in PicaPlus, MARC, LIDO und METS/MODS.

Die Regelsätze werden in g2g/workspace/workflow/rulesets abgelegt²⁰. Ein neuer Regelsatz muss Goobi unter „Administration > Regelsätze“ bekanntgegeben werden. Hier wird dann auch festgelegt, ob die Anzeigenreihenfolge der Metadatenfelder der Reihenfolge im Regelsatz folgen soll – ansonsten ist sie alphabetisch. In der hier beschriebenen Version ist kein Editor für die Regelsätze eingebaut. Dementsprechend erfolgt die Bearbeitung in einem einfachen Texteditor außerhalb von Goobi²¹.

2.6.2. Projekte

Unter „Administration > Projekte“ werden Digitalisierungsprojekte definiert. Hier werden zum einen organisatorische Details angegeben (Anzahl der Vorgänge, Anzahl der Seiten usw.), um den Status bzw. die Zielerreichung von Digitalisierungsprojekten statistisch analysieren zu können, aber auch technische Details festgelegt (Speicher- und Exportformat, Verzeichnispfade) und Parameter für die METS-Metadaten.

2.6.3. Produktionsvorlage

Unter „Workflow > Produktionsvorlagen“ können neue Produktionsvorlagen angelegt werden. Hier wird in einem ersten Schritt festgelegt, welchen Laufzettel und welchen Regelsatz ein Vorgang verwenden soll, sowie zu welchem Projekt die Vorgänge gehören. Im zweiten Schritt werden die Aufgaben angelegt und einzeln konfiguriert.

2.6.4. Digitale Sammlungen

Vorgänge werden einer oder mehreren Sammlungen zugeordnet. Dies geschieht bereits beim Anlegen des Vorgangs und kann später im Metadateneditor auch nachträglich adaptiert werden. Die Sammlungen werden in der Datei g2g/workspace/workflow/config/goobi_digitalCollections definiert. Dort kann auch die Zuordnung einer oder mehrere Sammlungen zu einem Projekt vorgenommen werden.

20 Die Dokumentation der Regelsätze ist zu finden unter <https://docs.goobi.io/ugh-de/3>

21 Ein Editor für Regelsätze und Konfigurationsdateien kann mittlerweile mit Plugins in Goobi workflow installiert werden (https://docs.goobi.io/goobi-workflow-plugins-de/administration/intranda_administration_config_file_editor und https://docs.goobi.io/goobi-workflow-plugins-de/administration/intranda_administration_ruleset_editor).

2.6.5. Metadatenanzeige

In der Datei g2g/workspace/workflow/config/goobi_metadataDisplayRules.xml kann projektabhängig definiert werden, wie die Metadaten im Metadateneditor dargestellt werden sollen, wie beispielsweise Auswahlbox (Dropdown) oder Mehrfachauswahl aus einer festgelegten Anzahl an Werten oder auch die Verknüpfung mit einer Liste aus dem Vokabularmanager. Der Vokabularmanager ist in „Administration > Vokabularverwaltung“ zu finden.

2.6.6. OPAC-Anbindung

Der Metadatenimport aus Bibliothekskatalogen wird über die Datei g2g/workspace/workflow/config/goobi_opac.xml gesteuert. Es werden die Dokumenttypen festgelegt, die aus den Katalogen abgerufen werden können, und für jeden Katalog die Verbindungsdetails und die Suchfelder, in denen ein Datensatz per ID oder Barcode gesucht werden kann.

2.6.7. Anlegemaske

Das Aussehen und Verhalten der Anlegemaske für die Vorgänge wird über die Datei g2g/workspace/workflow/config/goobi_projects.xml gesteuert. Das betrifft sowohl die Auswahl der verfügbaren Metadatenfelder, als auch automatische Abläufe wie die Generierung des Vorgangstitels. Außerdem können der abzurufende Katalog hier definiert und die für die Generierung der Tiff-Header benötigten Felder aus den Metadaten vorbefüllt werden.

3. Schlussbemerkung

Dieser Artikel bietet nur einen Überblick über die grundlegenden Funktionalitäten von Goobi workflow und die wichtigsten Konfigurationen. Für eine tiefere Beschäftigung ist der Blick in die Dokumentation zu empfehlen. Außerdem gibt es eine sehr aktive User-Community, die auf <https://community.goobi.io> erreicht werden kann. Dort bekommt man nicht nur hilfreiche Tipps, sondern trifft mitunter auch auf Denkanstöße für ganz neue Verwendungsmöglichkeiten dieser wandlungsfähigen Software.

Schließlich sei angemerkt, dass Goobi-to-go sich nicht nur dafür eignet, die Programme kennenzulernen. Da es sich um die vollständigen Goobi-Programme handelt, können auch komplexe Szenarien getestet werden. Es ist damit leicht, für unterschiedliche Testzwecke neue Goobi-Instanzen einzurichten. Diese Möglichkeit

einer einfach verfügbaren Testinfrastruktur wird nicht zuletzt durch ein Plugin²² unterstützt, mit dem Konfiguration und Datenbank von einem existierenden Goobi workflow in ein neues exportiert werden können. Somit kann in einigen wenigen Schritten das eigene Produktionssystem in einer Goobi-to-go-Instanz repliziert werden. Die Tests finden dann mit den aktuellen Einstellungen und den eigenen Vorgängen und Daten statt.

Bibliografie

- Frost, H.; Mangiafico, P. (2019): Integrating Digitization Workflow with the Stanford Digital Repository. Stanford University Libraries. https://repo.samvera.org/concern/parent/cdbeada1-0946-462c-a288-6c7647315bfd/file_sets/22df08d1-97bb-4908-911c-6a5f5f01ffb4 (abgerufen am 28.03.2022).
- Hankiewicz, S. (2018): Goobi entwickeln. Eine Open-Source Software zur Verwaltung von Workflows in Digitalisierungsprojekten. In: Neuböck, Gregor (Hg.): Digitalisierung in Bibliotheken. Viel mehr als nur Bücher scannen. Berlin: De Gruyter Saur. (Bibliotheks- und Informationspraxis 63), S. 77–87.

Georg Mayr-Duffner ist seit 2014 Systembibliothekar an der Wirtschaftsuniversität Wien. Seit 2016 arbeitet er an der technischen Umsetzung der Digitalen Sammlungen der WU-Bibliothek mit Goobi workflow und Goobi viewer.

²² https://docs.goobi.io/goobi-workflow-plugins-de/administration/intranda_administration_goobi2goobi