

ARCHIVING2019

14-17 May 2019 • Lisbon, Portugal

SC3C: End-to-End Digitization Workflow: Goobi to Go for Newbies

14 May 2019

Steffen Hankiewicz and Jan Vonde, intranda GmbH

14:15 – 18:30

These course notes are made available by the author(s) as supporting materials for the Short Course class noted above. Course notes are distributed only to registered participants and are not to be used as a stand-alone publication. All rights are reserved by the author, the copyright holder of these materials. These notes—or any part thereof—may not be reproduced nor distributed in any form without the written permission of the author(s).

This short course is sponsored by the Society for Imaging Science and Technology.

Table of content

1	Introduction	3
1.1	What is Goobi?	3
1.1.1	What is Goobi workflow?	3
1.1.2	What is the Goobi viewer?	3
1.2	What is Goobi to go?	4
2	Preparing and installing Goobi to go.....	4
2.1	Installing Goobi to go for macOS.....	4
2.2	Installing Goobi to go for Microsoft Windows	5
2.3	Installing Goobi to go for Linux	5
3	Hands-on: Goobi workflow.....	7
3.1	Create new objects.....	8
3.2	Reading in media/images.....	9
3.3	Quality control	10
3.4	Editing metadata	11
3.5	Export to Goobi viewer	12
3.6	Other workflow steps	13
4	Hands-on: Goobi viewer	14
4.1	Working with the Goobi viewer as a user	14
4.1.1	Working with the GUI.....	14
4.1.2	Interfaces and formats.....	16
4.2	Working with the Goobi viewer as an administrator	16
4.2.1	Configuring access restrictions	17
4.2.2	Creating new pages via CMS	17
4.2.3	Configuring how metadata are displayed.....	19
5	Processing legacy data.....	20
5.1	Importing data from Excel files.....	20
5.2	Automated catalogue requests	21
5.3	Other automated workflow steps and export	22
5.4	Metadata enrichment and subsequent re-export	22
6	Next steps.....	23
6.1	Documentation	23
6.2	Community and Forum	23
6.3	User meetings	23
6.4	Contact	23

1 Introduction

1.1 What is Goobi?

Goobi is an open-source software suite. It can be used to manage and present digitisation projects. There are two main components - [Goobi workflow](#) and [Goobi viewer](#). There is also an extensive ecosystem of additional tools and plugins.

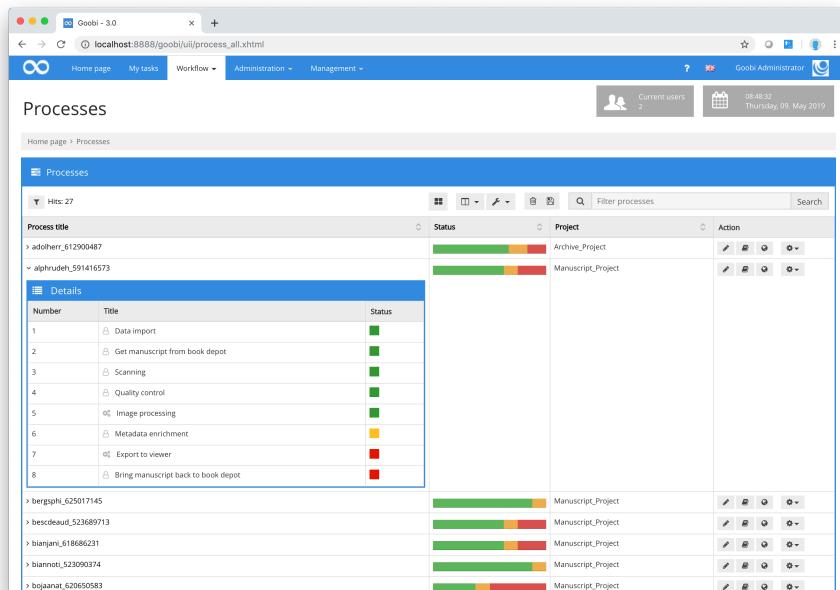
Goobi is currently used by over 80 cultural institutions (predominantly libraries, museums and archives) in 17 countries.



1.1.1 What is Goobi workflow?

[Goobi workflow](#) is the main workflow management component. It can be used to manage digitisation projects. Its strength lies in the flexibility it offers thanks to a wide range of plugin interfaces. As a web-based application, it can be operated independently of the user's operating system and location.

In terms of process management, [Goobi workflow](#) accepts various data, manages how those data are processed and then exports the desired results. The user can define any number of manual or automated processing steps. To help prevent potential errors, configurable validation procedures can be used at an early stage in the process to check subsets of results (e.g. files and metadata). Statistical analyses can be performed throughout the digitisation workflow giving the user a detailed picture of the project's current status. The export function supports numerous formats and targets.



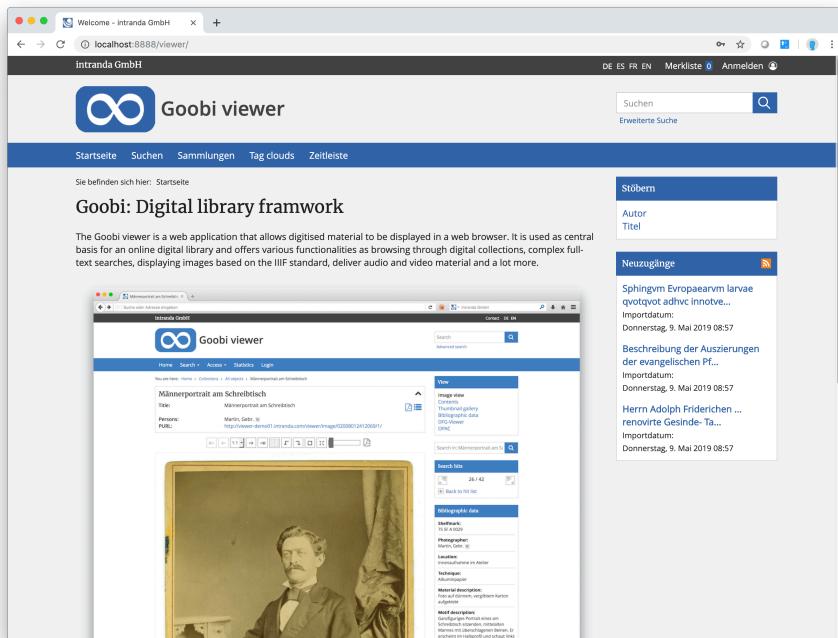
The screenshot shows the Goobi Workflow application running in a web browser. The main window displays a table of processes. The columns include Process title, Status, Project, and Action. A sidebar on the left shows a tree view of process categories. A modal dialog box is open, titled 'Details', showing a list of steps with titles like 'Data import', 'Get manuscript from book depot', 'Scanning', etc., each with a status indicator (green, yellow, red).

1.1.2 What is the Goobi viewer?

The function of the [Goobi viewer](#) is to present digital works and objects. As well as simply displaying objects, the [Goobi viewer](#) allows users to compile works into digital collections. It supports various searches and facetting within the entire data collection.

The GUI is available in several languages and can be fully adapted visually to meet individual needs. Content can be curated and individual pages created through an integrated content management system.

As well as displaying objects on the internet for end users, the [Goobi viewer](#) supports a large number and range of interfaces.



1.2 What is Goobi to go?

[Goobi to go](#) is a complete Goobi work environment. It consists of the core components [Goobi workflow](#) and [Goobi viewer](#) together with a number of selected plugins and some example data. Once you have downloaded Goobi to go as a zip file, you can launch it directly without prior installation in macOS, Microsoft Windows and Linux by simply double-clicking.

[Goobi to go](#) makes it very easy for users to evaluate individual components or try out and run through modifications without having to use a live system.



2 Preparing and installing Goobi to go

In order to install [Goobi to go](#), you must have [Java 8](#) on your system. If [Java 8](#) is not already installed, you can download it for different operating systems at:

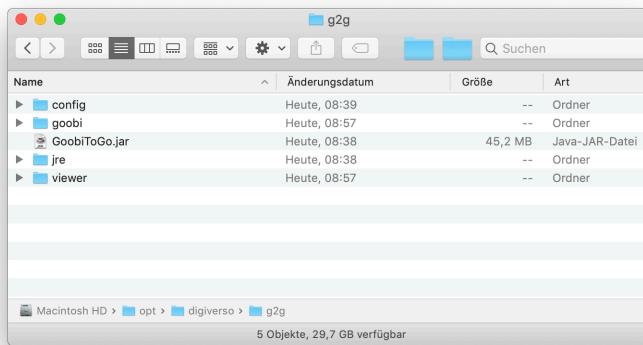
<https://www.java.com/en/download/manual.jsp>

If [Java 8](#) is already installed, you can download the [Goobi to go](#) package as a zip file ([g2g.zip](#)) from:

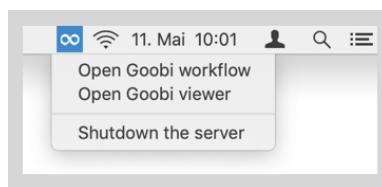
<https://files.intranda.com/archiving2019>

2.1 Installing Goobi to go for macOS

After downloading the file [g2g.zip](#), you will first need to unzip it by double-clicking. The file system will now show a folder entitled [g2g](#) with the following contents:

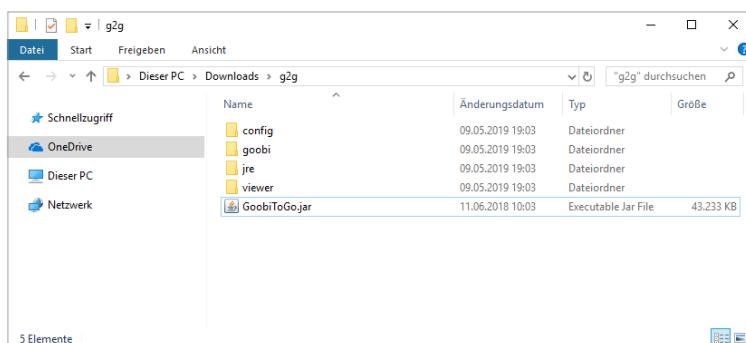


To launch **Goobi to go**, double-click the file **GoobiToGo.jar**. Once it has opened, the launch screen is hidden so that **Goobi workflow** and the **Goobi viewer** can be opened using the icon in the menu.

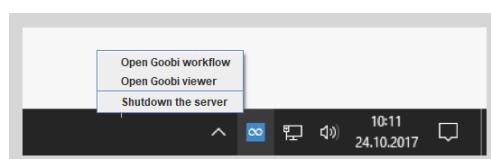


2.2 Installing Goobi to go for Microsoft Windows

After downloading the file **g2g.zip**, you will first need to unzip it. The file system will now show a folder entitled **g2g** with the following contents:

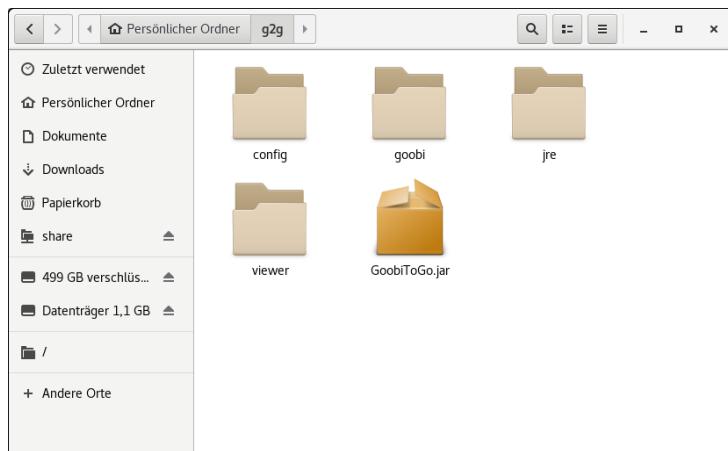


To launch **Goobi to go**, double-click the file **GoobiToGo.jar**. Once it has opened, the launch screen is hidden so that **Goobi workflow** and the **Goobi viewer** can be opened using the icon in the Systray menu.

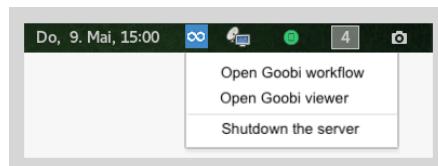


2.3 Installing Goobi to go for Linux

After downloading the file **g2g.zip**, you will first need to unzip it. The file system will now show a folder entitled **g2g** with the following contents:



In the case of Linux, [Goobi to go](#) is usually launched via a command-line call. Once it has opened, the launch screen is hidden. Both [Goobi workflow](#) and the [Goobi viewer](#) can then be opened. In some Linux distributions you will now see an icon in the main menu. This can be used to open the applications.



If the menu is not displayed, you can open [Goobi workflow](#) and [Goobi viewer](#) directly in a web browser at:

```
http://localhost:8888/goobi/  
http://localhost:8888/viewer/
```

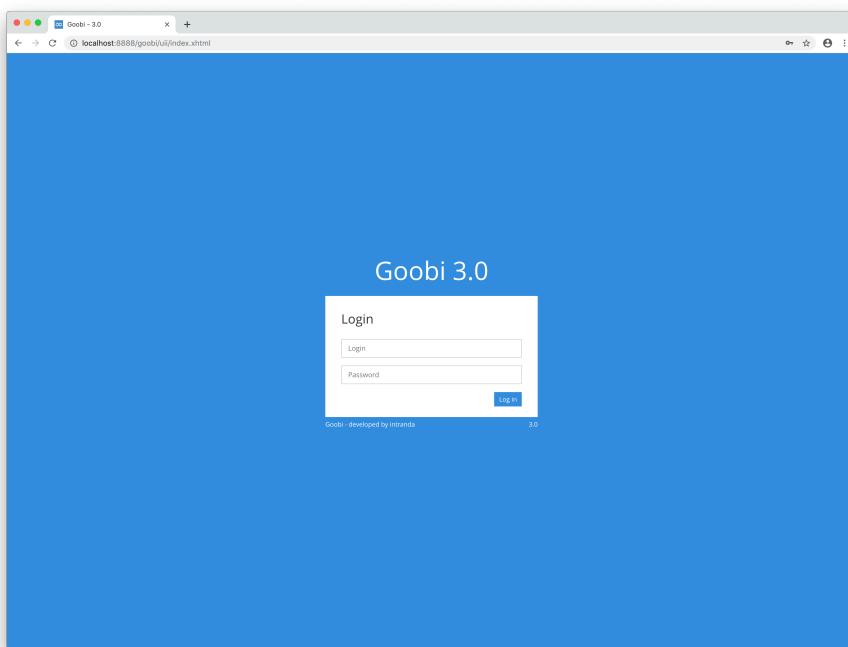
All the required command-line instructions are shown below:

```
# Download Goobi to go  
wget -O g2g.zip https://files.intranda.com/archiving2019  
  
# Unzip zip file  
unzip -q g2g.zip  
  
# Open unzipped folder  
cd g2g  
  
# Launch Goobi to go  
java -jar GoobiToGo.jar
```

3 Hands-on: Goobi workflow

The [Goobi workflow](#) GUI is called from the following URL:

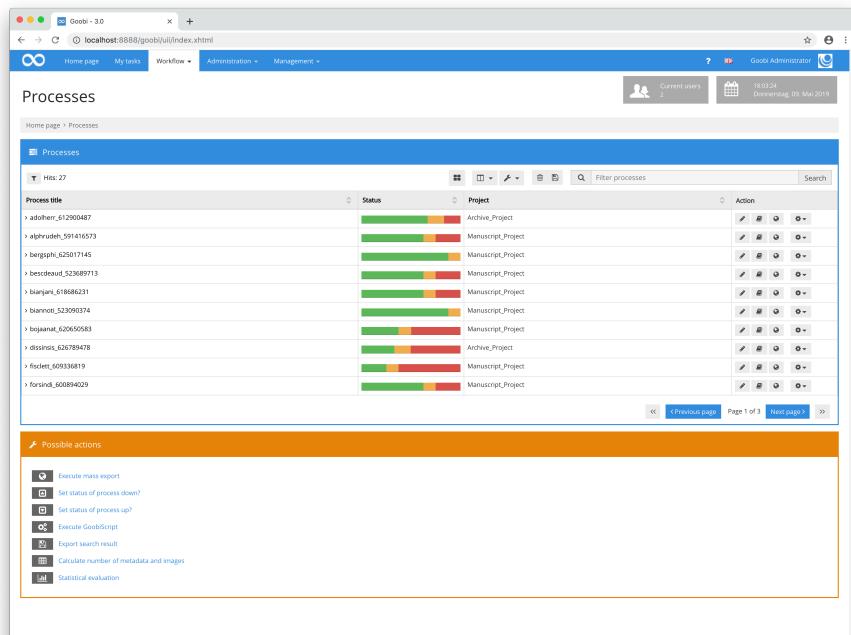
```
http://localhost:8888/goobi/
```



You can now enter one of the following user names in the login area:

Login	Password	Description
goobi	goobi	Administrator
testscanning	test	Scanning officer
testqc	test	Quality control officer
testmetadata	test	Metadata officer
testprojectmanagement	test	Project manager
testimaging	test	Image optimisation officer

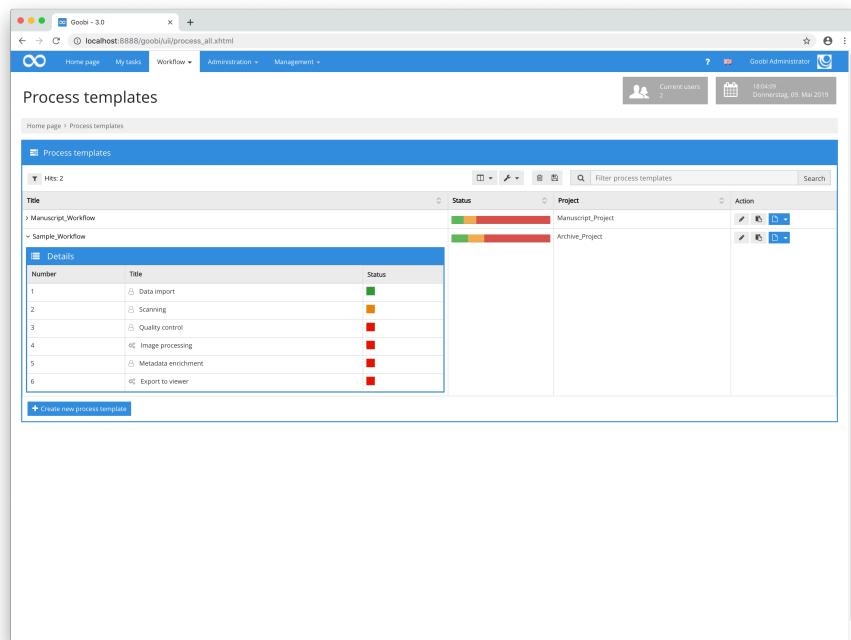
Depending on your authorisation level, you will now be able to access certain areas of the [Goobi workflow](#), e.g. a list of all the current processes.



The screenshot shows the 'Processes' page in the Goobi workflow interface. The main area displays a table of processes with columns for 'Process title', 'Status', 'Project', and 'Action'. The 'Status' column uses color-coded bars to represent different states. The 'Project' column shows categories like 'Archive_Project' and 'Manuscript_Project'. The 'Action' column contains icons for various operations. Below the table is a 'Possible actions' section with options such as 'Execute mass export', 'Set status of process down?', 'Set status of process up?', 'Execute GoobiScript', 'Export search result', 'Calculate number of metadata and images', and 'Statistical evaluation'.

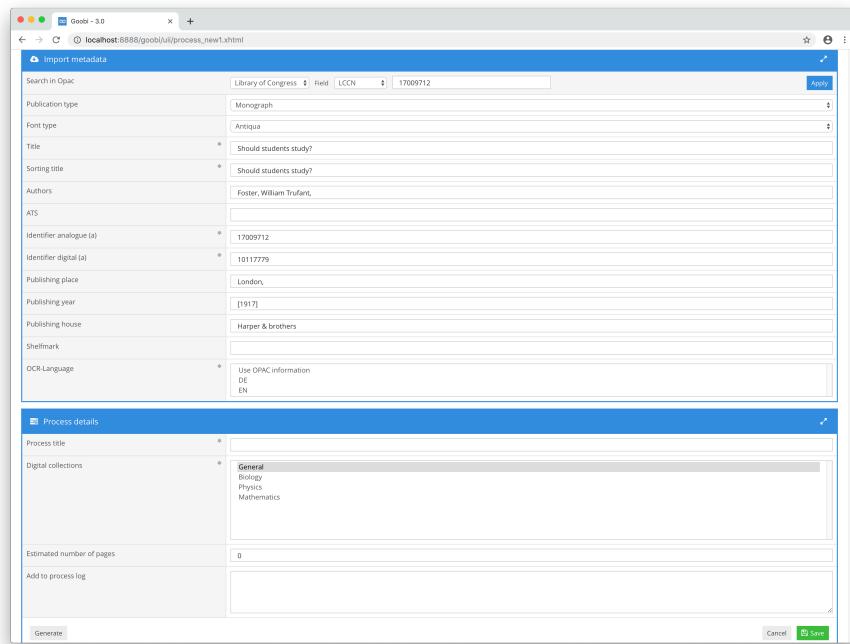
3.1 Create new objects

As an **Administrator** or **Project Manager**, you can create new objects in **Goobi workflow**. To do this, first open the list of production templates.



The screenshot shows the 'Process templates' page in the Goobi workflow interface. The main area displays a table of process templates with columns for 'Title', 'Status', and 'Action'. The 'Title' column lists 'Manuscript_Workflow' and 'Simple_Workflow'. The 'Status' column shows 'Manuscript_Project' and 'Archive_Project'. The 'Action' column contains icons for various operations. A modal window titled 'Details' is open, showing a table of steps with columns for 'Number', 'Title', and 'Status'. The steps listed are: 1. Data import (Status: green), 2. Scanning (Status: orange), 3. Quality control (Status: red), 4. Image processing (Status: red), 5. Metadata enrichment (Status: red), and 6. Export to viewer (Status: red). At the bottom of the modal is a blue button labeled '+ Create new process template'.

Once you have clicked the button to select a **Production template**, **Goobi workflow** will display an input form in which you can enter the relevant metadata. This can be done manually, or you can import the data from an external source, e.g. a library catalogue.

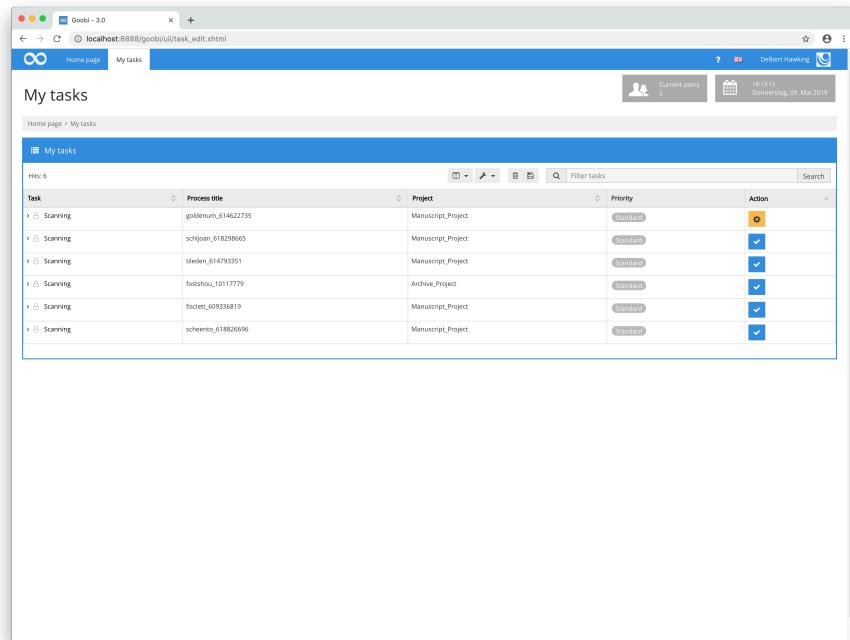


The screenshot shows the Goobi 3.0 interface for creating a new process. The top section, "Import metadata", contains fields for publication type (Monograph), font type (Antiqua), title ("Should students study?"), sorting title ("Should students study?"), authors (Foster, William Trufan), and various identifiers (LCCN, ISBN). The bottom section, "Process details", includes fields for process title, digital collections (General, Biology, Physics, Mathematics), estimated number of pages (0), and a log entry. A "Save" button is visible at the bottom right.

You can now click **Save** to create a new Goobi process using the data you have provided and open the workflow.

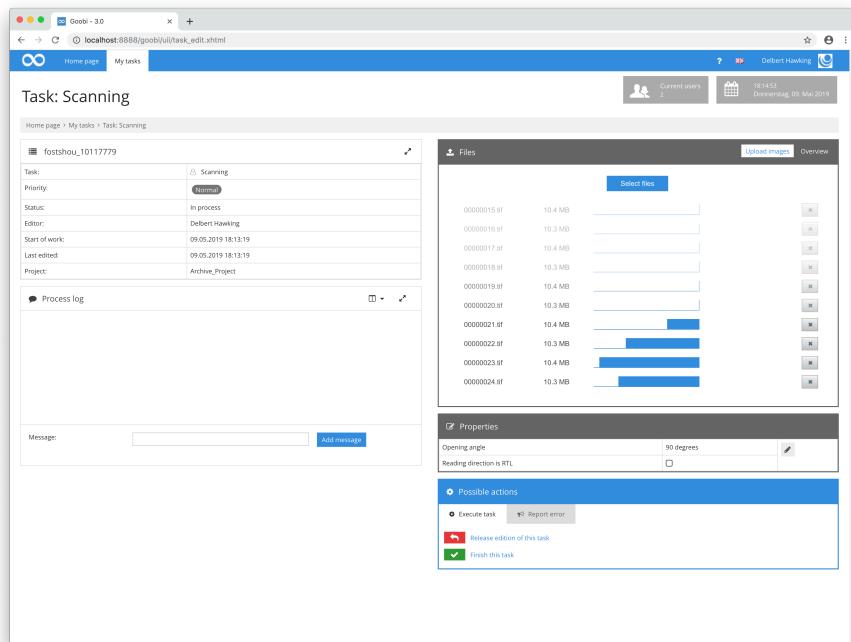
3.2 Reading in media/images

As a **Scanning officer**, you can read images into Goobi. First select the right object from your own tasks and accept the processing job.



The screenshot shows the "My tasks" interface in Goobi 3.0. It displays a list of six scanning tasks, each with a title like "Scanning goldenen_61462273S" and a status column showing "Completed". The tasks are categorized by project (Manuscript_Project, Archive_Project) and priority (Standard). A "Search" bar and a "Filter tasks" button are at the top of the list table.

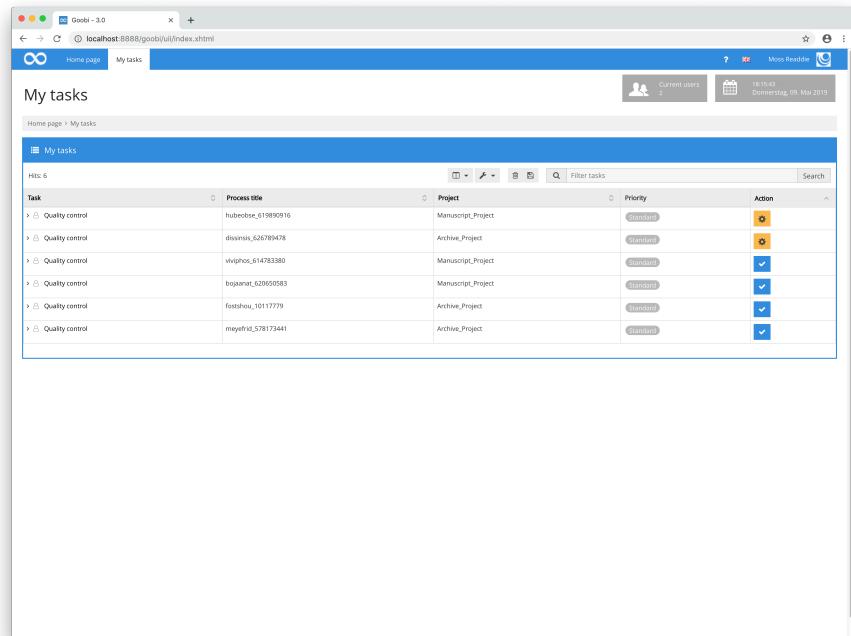
The digitised images can now be uploaded simply by dragging them into the upload area of the form.



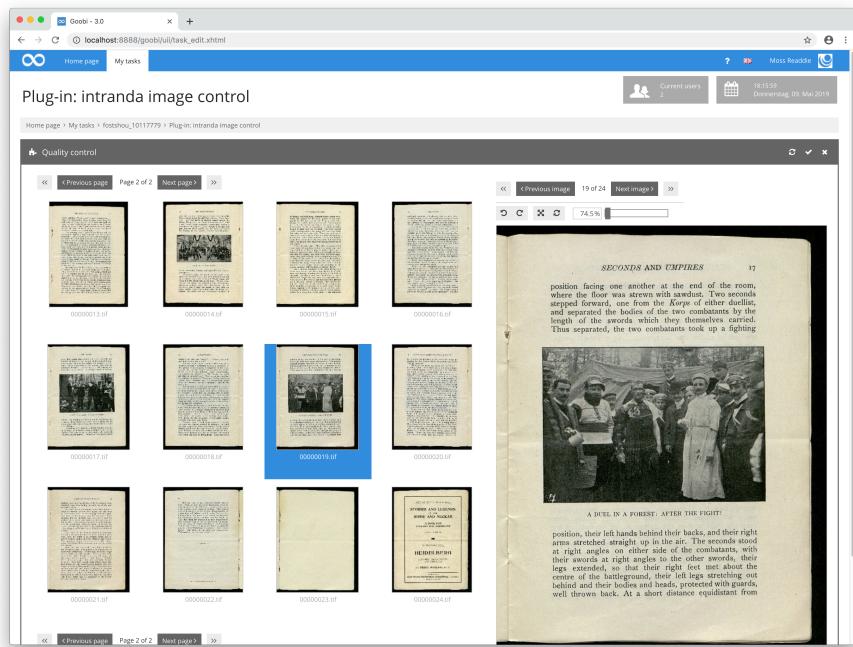
Once the upload is complete, click the [Finish this task](#) button and activate the next stage of the workflow.

3.3 Quality control

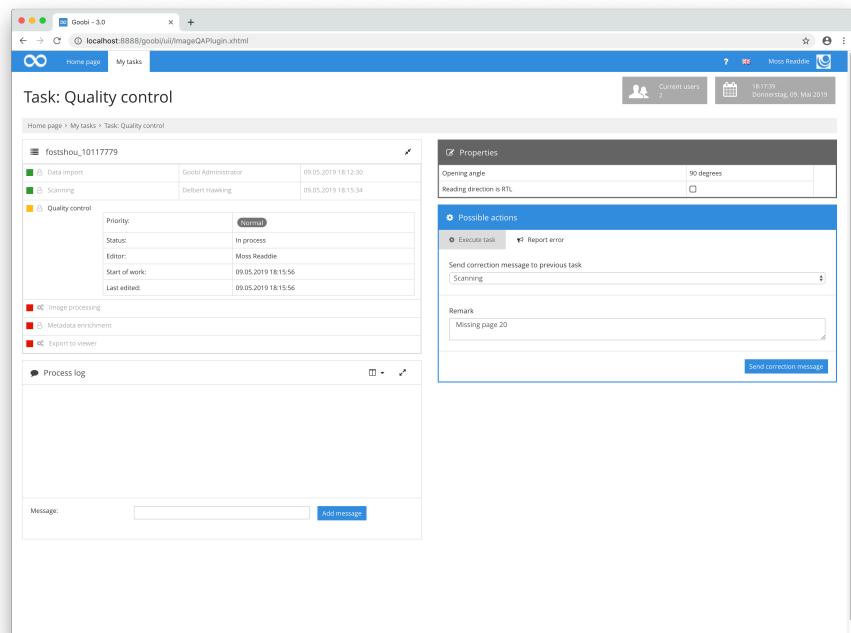
As a [Quality control officer](#), your role is to check the quality of the images. You will need to accept this task before you can start work.



Within this task, you can enter the [Image QA plugin](#). This allows you to carefully review all the previously uploaded images of the selected object.

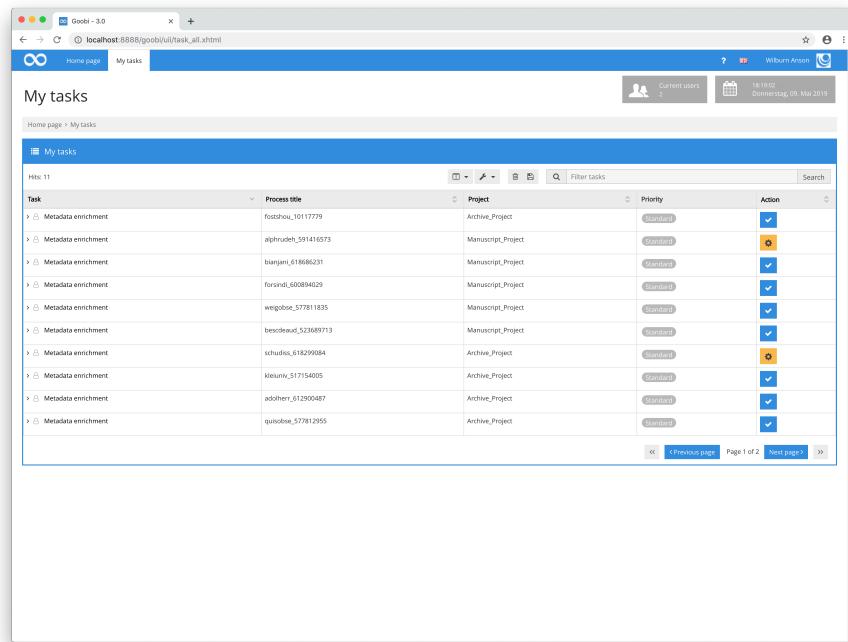


After performing quality checks, the user can exit the plugin and conclude the task or (if necessary) submit a correction message.

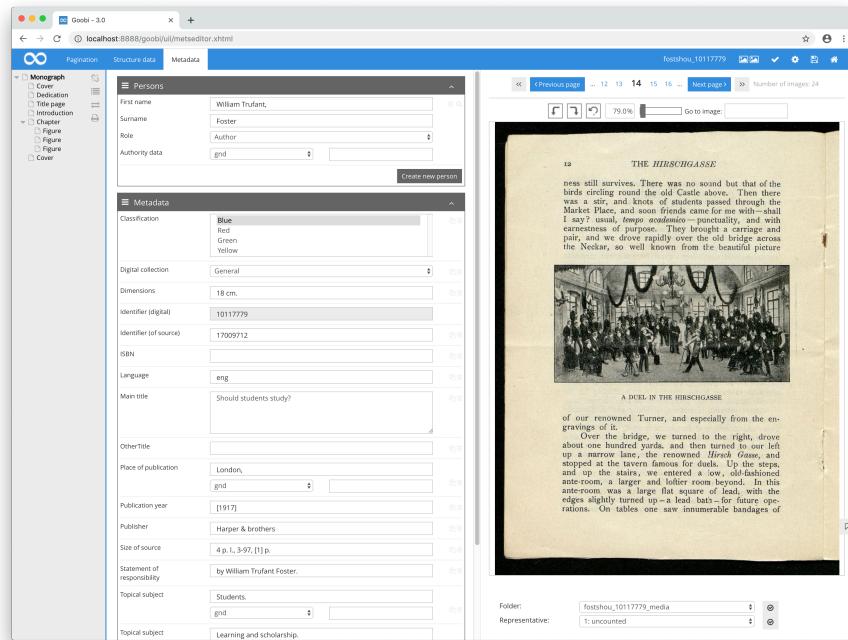


3.4 Editing metadata

As a **Metadata officer**, you should first select the object you want to work on from the task list and then accept that task.



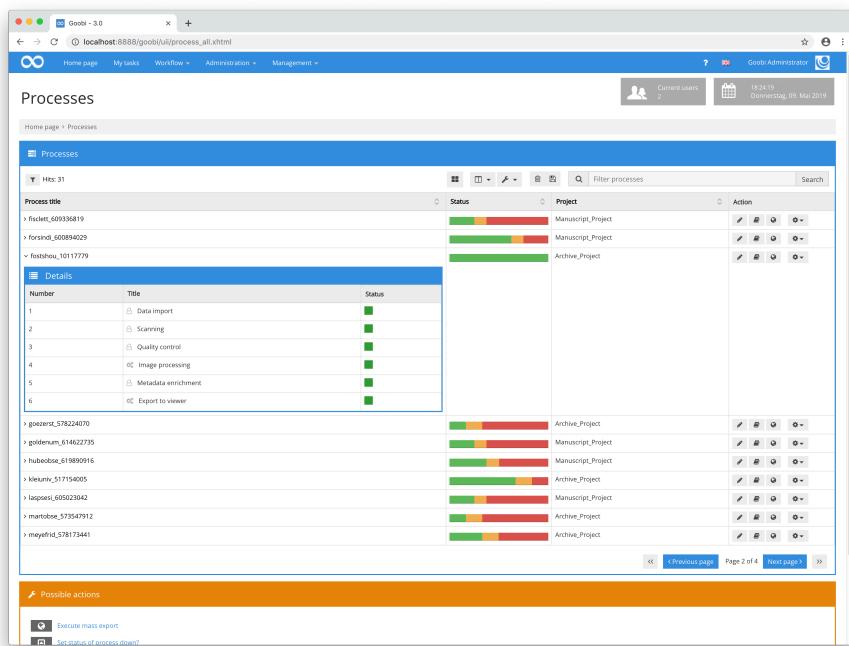
Within the task, the user can enter the [Goobi Metadata Editor](#) and edit the data under the headings [Pagination](#), [Structure data](#) and [Metadata](#).



After completing the task, the user can save the data and exit the [Metadata Editor](#). This will end the task and trigger the next step in the workflow.

3.5 Export to Goobi viewer

Once the metadata have been entered, the application will perform an automatic export to the [Goobi viewer](#). The digitised object is now published and can be viewed by visitors.



The screenshot shows the Goobi workflow interface. At the top, there's a navigation bar with links for Home page, My tasks, Workflow, Administration, and Management. On the right side of the header, there are icons for Current users (2), a calendar showing 18:24:19 Donnerstag, 09. Mai 2019, and a user profile for Goobi Administrator.

The main content area is titled "Processes". It displays a list of processes for an object with ID fischer_609336819. The list includes:

- Process title: fischer_609336819
- Process details:

Number	Title	Status
1	Data import	Green
2	Scanning	Green
3	Quality control	Green
4	Image processing	Green
5	Metadata enrichment	Green
6	Export to viewer	Green
- Project status: Manuscript_Project (represented by a green and yellow bar)
- Other processes listed below include: forstd_600894029, foreshow_10117779, goerz_578224070, goleidem_614622735, hubedose_519890916, kleiner_517154005, lappas_605023942, martodse_573547912, meyrfid_578173441.
- Possible actions at the bottom: Execute mass export and Set status of process down!

As an [Administrator](#), you can examine the process list to see exactly which workflow steps have been completed for a given object.

3.6 Other workflow steps

Apart from the typical workflow steps outlined above, there are many other ways in which the workflow can be customised. There are currently more than 180 plugins covering different areas of Goobi. As well as individual steps such as image straightening and cropping, [Goobi workflow](#) offers many other features, e.g. data validation, individual export formats, ingests to different external systems such as permanent archives and repositories, imports from various data sources, APIs and statistics. It is also possible to configure tasks that cannot be performed in Goobi as part of the workflow. This means that they can be worked through systematically for each object and their progress monitored.

4 Hands-on: Goobi viewer

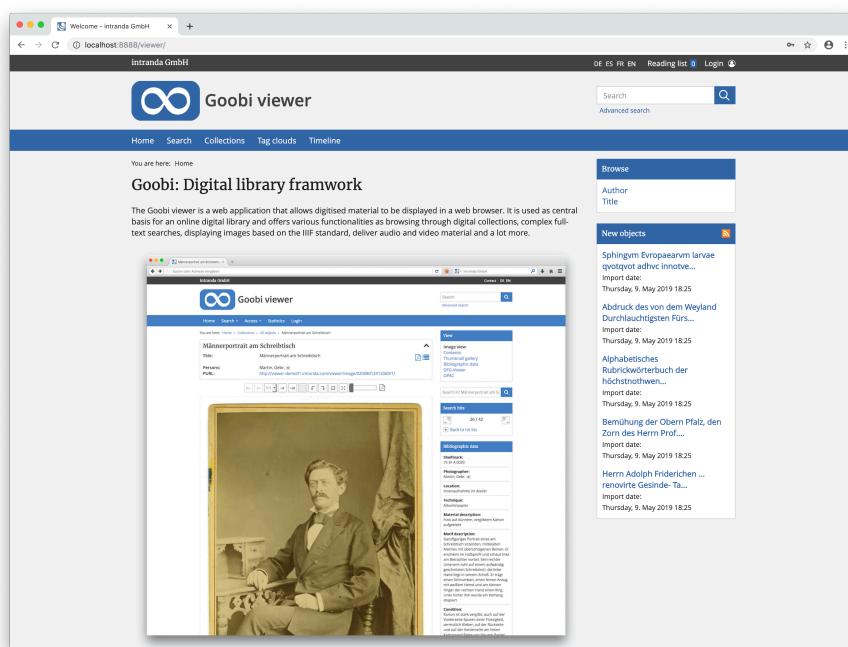
4.1 Working with the Goobi viewer as a user

The [Goobi viewer](#) GUI is called from the following URL:

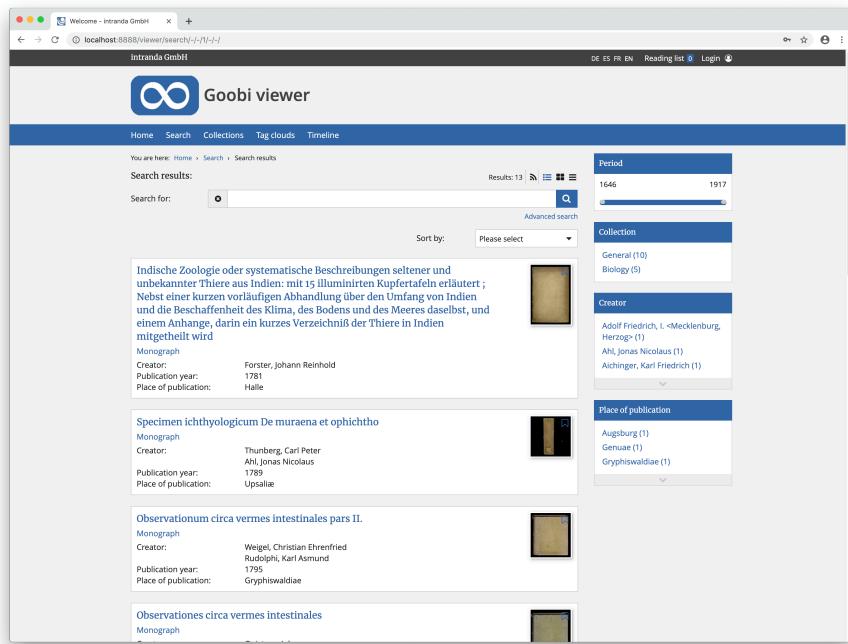
```
http://localhost:8888/viewer/
```

4.1.1 Working with the GUI

If you open an instance of the [Goobi viewer](#) from the URL, you will see the following welcome screen:



From here the user can access the two main components of the [Goobi viewer - Search](#) and [Work view](#). The search function may also be available from the menu or in the page header area:

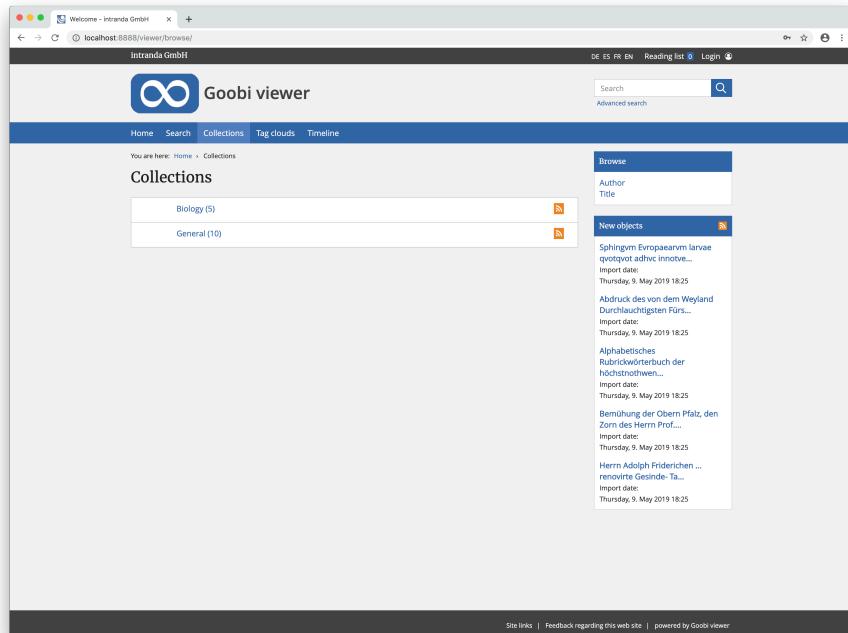


The screenshot shows the Goobi viewer interface with a search results page. The search term is "Indische Zoologie oder systematische Beschreibungen seltener und unbekannter Thiere aus Indien". The results list includes:

- Indische Zoologie oder systematische Beschreibungen seltener und unbekannter Thiere aus Indien:** mit 15 illuminirten Kupferstichen erläutert; Nebst einer kurzen vorläufigen Abhandlung über den Umfang von Indien und die Beschaffenheit des Klima, des Bodens und des Meeres dasselbst, und einem Anhange, darin ein kurzes Verzeichniß der Thiere in Indien mitgetheilt wird.
- Specimen ichthyologicum De muraena et ophichtho**
- Observationum circa vermes intestinales pars II.**
- Observationes circa vermes intestinales**

On the right side, there are filters for Period (1646-1917), Collection (General 10, Biology 5), Creator (Adolf Friedrich, I. <Mecklenburg, Herzoge 1, A.H. Jonas Nicolaus 1, Aichinger, Karl Friedrich 1), and Place of publication (Augsburg 1, Genuae 1, Gryphiswaldiae 1).

There are several ways of asking the viewer to display an object. One of these is through the collection list:



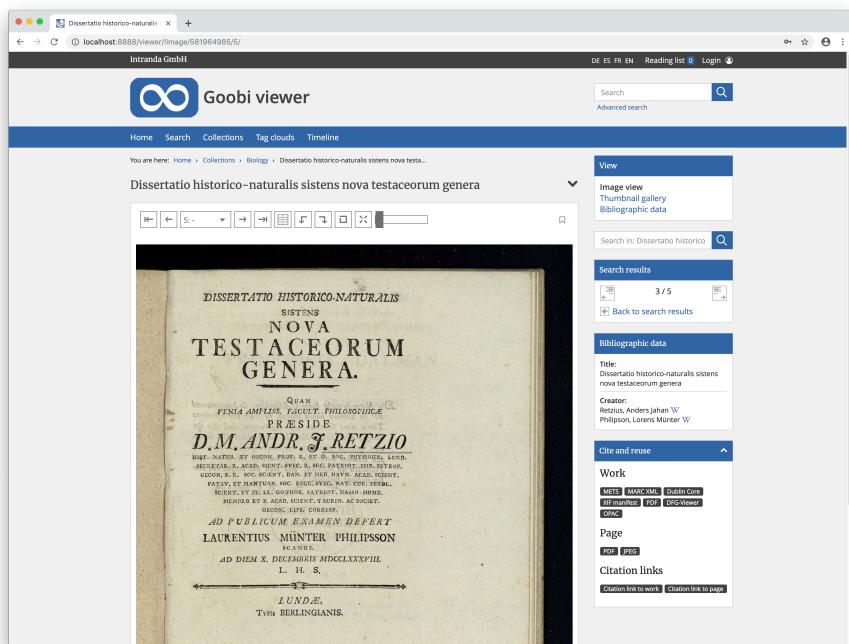
The screenshot shows the Goobi viewer interface with a collections page. The collections listed are:

- Biology (5)
- General (10)

On the right side, there are filters for Browse (Author, Title) and New objects, which lists recently imported items:

- Sphingim Europaeum larvae quoque adhuc innovata...
- Import date: Thursday, 9. May 2019 18:25
- Auszug des von dem Weyland Durchsuchtigsten Fürs...
- Import date: Thursday, 9. May 2019 18:25
- Alphabetisches Rubrickwörterbuch der Höchstrathen...
- Import date: Thursday, 9. May 2019 18:25
- Bemühung der Obern Pfalz, den Zorn des Herrn Prof...
- Import date: Thursday, 9. May 2019 18:25
- Herrn Adolph Friderichen ... renovirte Gesinde-Ta...
- Import date: Thursday, 9. May 2019 18:25

Once an object has been opened, the viewer will offer the usual navigation and display functions for working with digitised images. The [Cite and reuse](#) widget in the side panel shows relevant licence information and links that may apply to further use of the data.



4.1.2 Interfaces and formats

All the data available in the Goobi viewer can be requested and reused via standardised interfaces provided that they have been set up and suitably configured.

Goobi viewer can process metadata in [METS/MODS](#), [LIDO](#) und [TEI](#) formats. It supports full texts in [Plaintext](#), [ALTO](#) and [TEI](#) formats and images in [TIFF](#), [JPEG](#), [PNG](#) and [JPEG2000](#) formats. Image sizes and image sections are automatically calculated, generated and supplied in real time. As well as image formats, the Goobi viewer supports 3D objects and several audio and video formats.

Images are compatible with the [IIIF Image API](#). [IIIF Presentation Manifests](#) can be called whatever the source format of the metadata. Information about changes is provided by another integrated [IIIF Change Discovery API](#). Image annotations are issued as [Open Annotations](#) or [Web Annotations](#) within the [IIIF Manifest](#).

The OAI interface supports the formats [METS/MODS](#), [MARCXML](#), [DublinCore](#), [LIDO](#), [TEI](#), [CMDI](#), [ESE](#) and [xepicur](#).

Bibliography management programs are supported by [COinS](#) and through the [SRU](#) interface.

4.2 Working with the Goobi viewer as an administrator

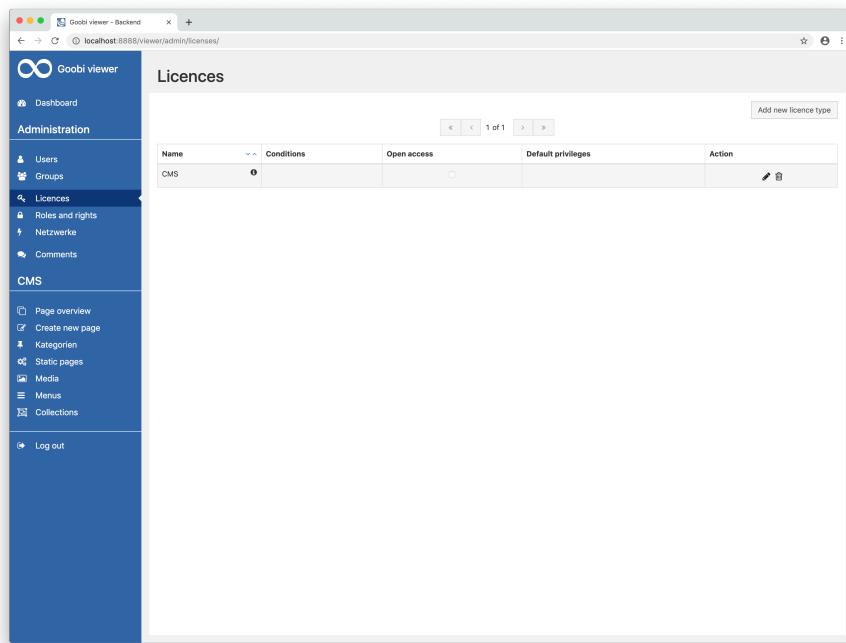
After logging in, users with the required authorisation level can use the back-end. The following data can be used to log in within [Goobi to go](#):

Login	Password
goobi@intranda.com	goobi

The back-end is divided into two areas, **Administration** and **CMS**. The **Administration** area can be used to manage user groups or configure access licences. The **CMS** area allows users to perform a range of actions, e.g. create new pages or modify the menu structure.

4.2.1 Configuring access restrictions

After logging in to the back-end of the **Goobi viewer**, the user can configure access restrictions in the **Administration** area. This involves issuing licences, adding the required authorisations and linking them to user accounts, groups or IP networks.

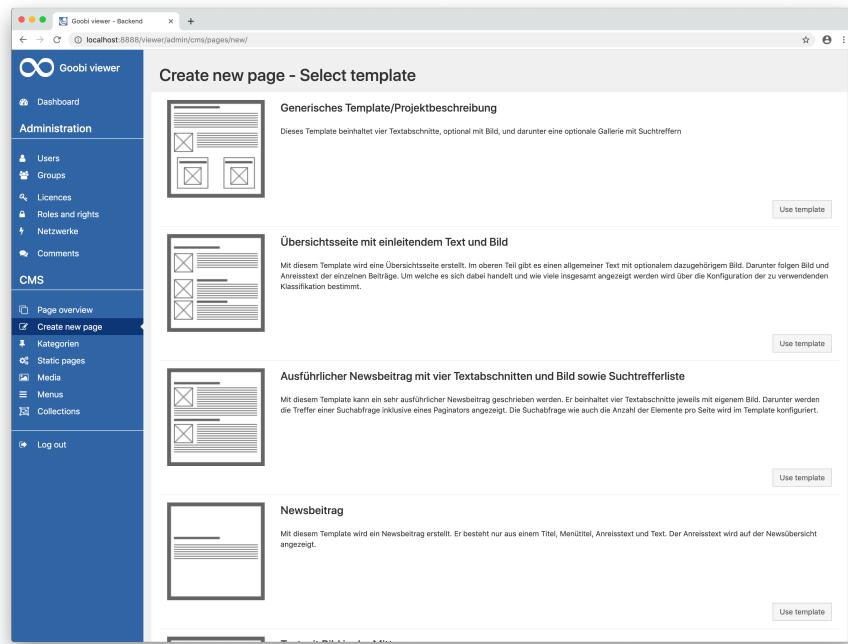


A detailed configuration of access restrictions can be found in section 6.8 of the **Goobi viewer** documentation at:

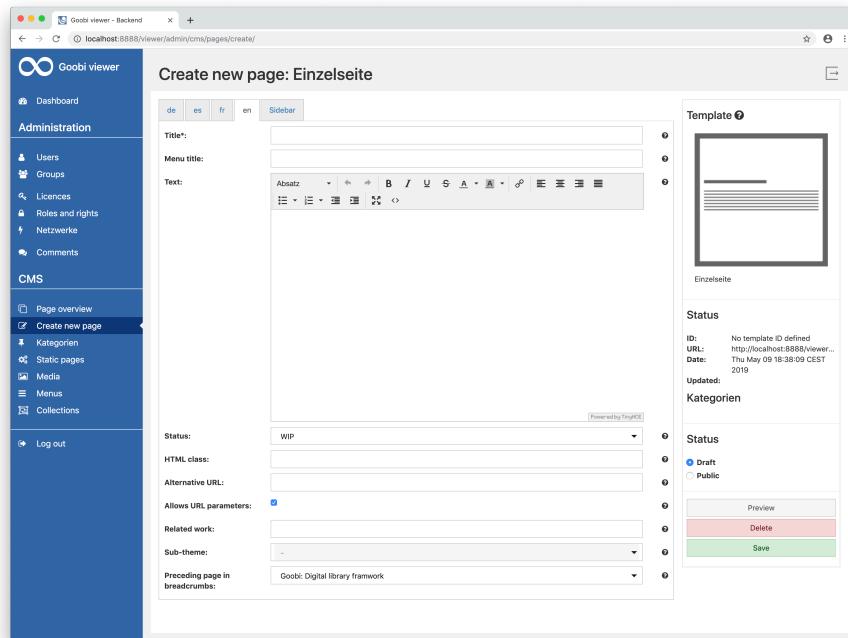
<https://docs.intranda.com/goobi-viewer-de/6/6.8>

4.2.2 Creating new pages via CMS

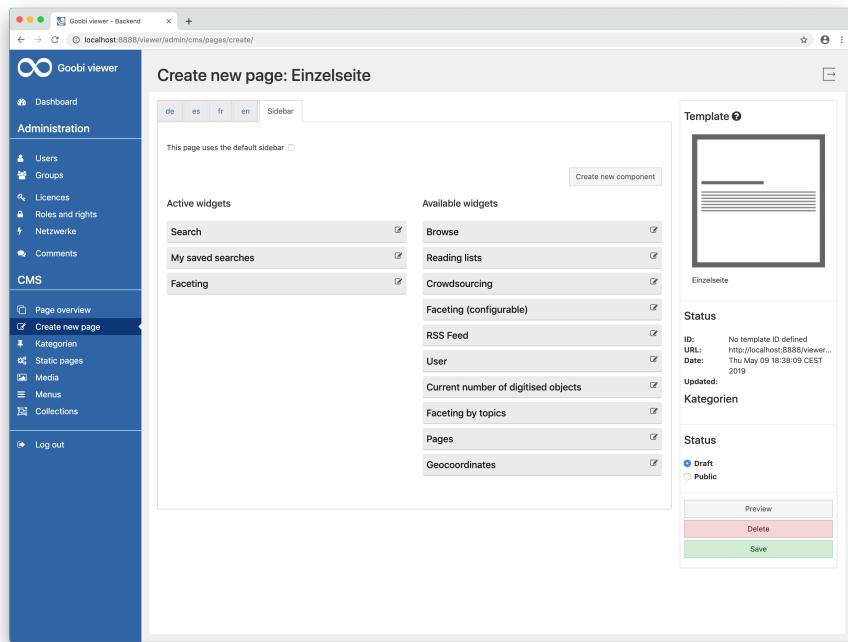
After logging in to the **Goobi viewer** back-end, the administrator can view a list of all the available templates by clicking the link **Create new page** in the **CMS** area. By clicking on the **Use template** button, you can create a new page based on the chosen template.



Within the page, depending on the chosen template, there are various configuration options that allow the administrator to define the **Page title**, **Text**, **Image** and **Solr query** for available search hits.

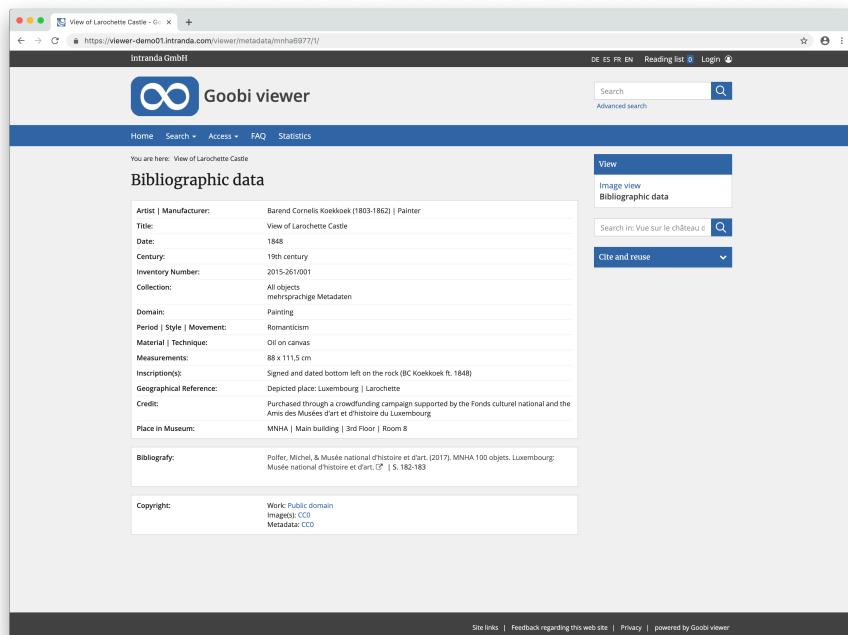


On the different tabs, the administrator can also set individual language versions for the pages. The **Sidebar** tab can be used to configure the **sidebar widgets** used on the page. If the page has been saved and its status set to **public**, it can be viewed in the front-end and a link created, for example, in the menu.



4.2.3 Configuring how metadata are displayed

The way metadata are displayed can be configured very precisely in the [Goobi viewer](#) front-end.



A typical configuration is shown below:

```
<metadata label="MD_BIBLIOGRAPHY_GROUP" value="MASTERVALUE_BIBLIOGRAPHY" group="true" type="2">
<param type="field" key="MD_ALEPH_ID" value="MASTERVALUE_BIBLIOGRAPHY_ALEPH_ID_BEFORE" />
<param type="field" key="MD_PREFERRED_CITATION"/>
<param type="field" key="MD_ALEPH_ID" value="MASTERVALUE_BIBLIOGRAPHY_ALEPH_ID_AFTER" />
<param type="field" key="MD_PAGE" value="MASTERVALUE_BIBLIOGRAPHY_PAGE"/>
<param type="field" key="MD_CATALOGUE" value="MASTERVALUE_BIBLIOGRAPHY_CATALOGUE"/>
</metadata>
```

```
MD_BIBLIOGRAPHY_GROUP=Bibliography
MASTERVALUE_BIBLIOGRAPHY_ALEPH_ID_BEFORE=<a href="http://a-z.lu/primo-explore/fulldisplay?docid={0}&vid=BIBNET&lang=en_US" target="_blank" style="color:#4c4c4c">
MASTERVALUE_BIBLIOGRAPHY_ALEPH_ID_AFTER=<i class="fa fa-external-link" aria-hidden="true"></i></a>
MASTERVALUE_BIBLIOGRAPHY_PAGE= | p. {0}
MASTERVALUE_BIBLIOGRAPHY_CATALOGUE= | ill. {0}
MASTERVALUE_BIBLIOGRAPHY={1}{3}{5}{7}{9} <br />
```

More information on how to configure the way metadata are displayed can be found in section 2.19.1 of the [Goobi viewer](#) documentation:

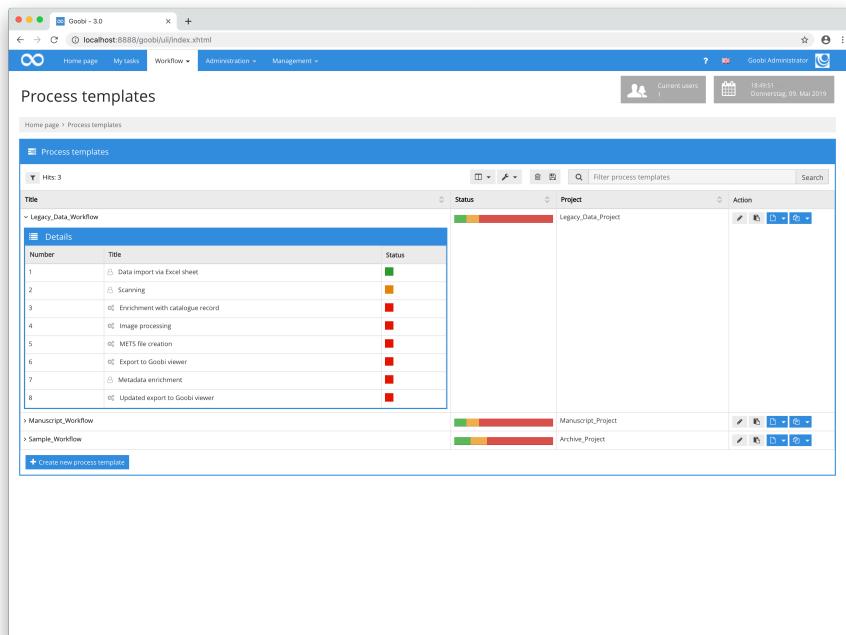
<https://docs.intranda.com/goobi-viewer-de/2/2.19/2.19.1>

5 Processing legacy data

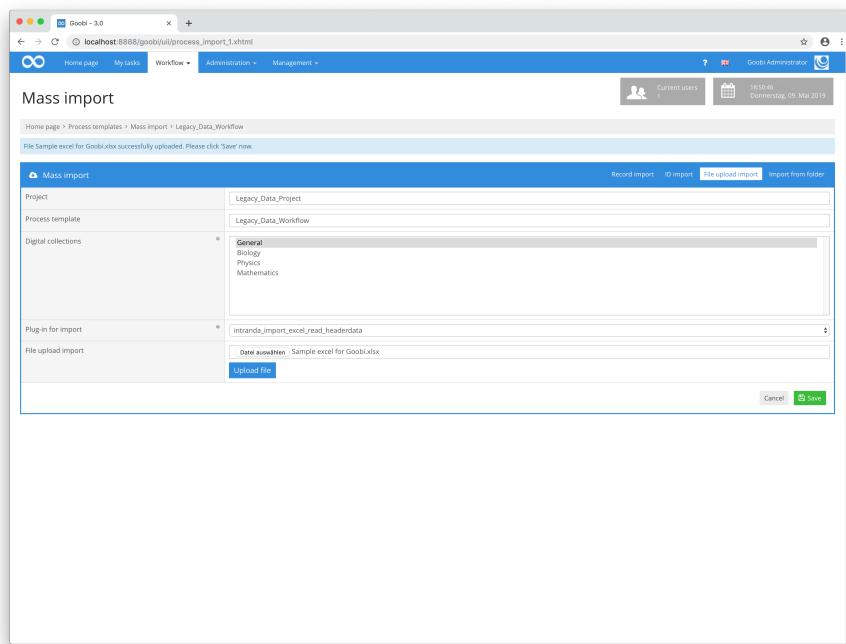
Users frequently need to import legacy data. In the following example, an Excel file with two metadata (identifier and title) per dataset is uploaded so that processes can be created automatically using the data. The next step is to upload the corresponding images. Various automated tasks are then performed in the rest of the workflow. By way of example, these might involve exporting the processes to the [Goobi viewer](#). Once these automated steps have been completed, the [Goobi workflow](#) remains in the [Metadata enrichment](#) step to enable final quality checks and re-export.

5.1 Importing data from Excel files

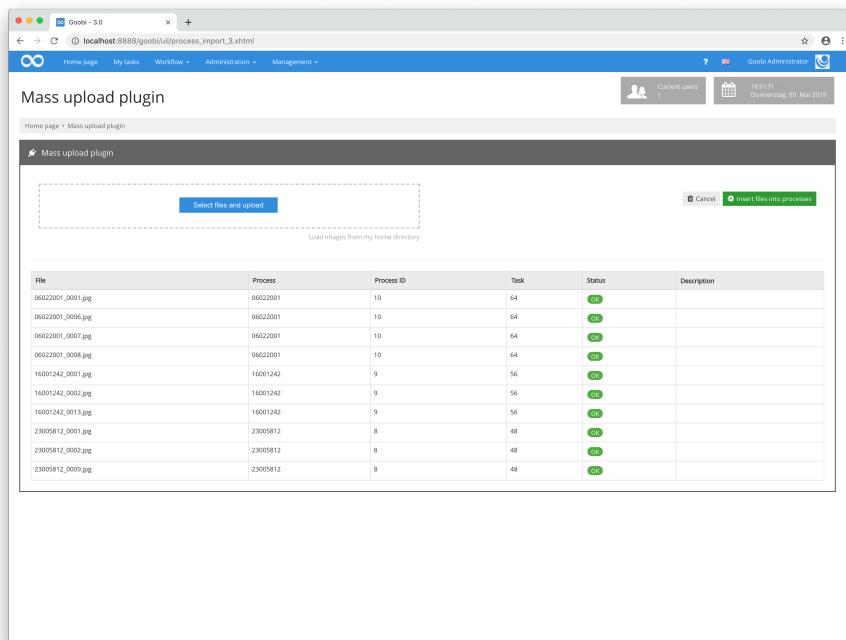
The first step here is to open the [Mass import](#) dialogue box for the chosen [Process template](#).



Here you can select the required import variant and the right import plugin and enter the necessary information.



Images can then be uploaded using the [Mass upload plugin](#). Images are allocated automatically to the right process on the basis of the prefix. Once the images have been uploaded, the workflow is updated and the next step is activated for each object that has been read in.



5.2 Automated catalogue requests

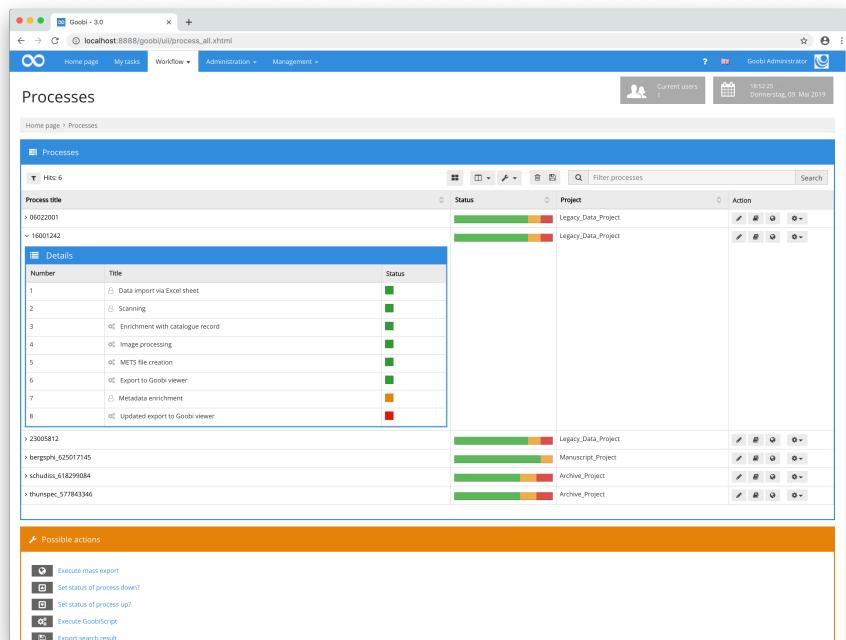
Since the metadata from the Excel table were very basic when imported, they are now updated automatically. This involves submitting requests to the Library of Congress catalogue.

5.3 Other automated workflow steps and export

Several workflow steps can be performed to prepare the data for export. These include generating a derivative of the uploaded images and updating the pagination within the METS files. The objects are then exported to the [Goobi viewer](#) in a fully automated process and can then be viewed.

5.4 Metadata enrichment and subsequent re-export

The next step is to open the workflow step [Metadata enrichment](#), either to perform a final quality check or to permit further enrichment of the metadata. Once this step has been completed, the entire object is re-exported definitively to the [Goobi viewer](#).



Number	Title	Status
1	Data import via Excel sheet	Green
2	Scanning	Green
3	Enrichment with catalogue record	Green
4	Image processing	Green
5	METS file creation	Green
6	Export to Goobi viewer	Green
7	Metadata enrichment	Yellow
8	Updated export to Goobi viewer	Red

6 Next steps

6.1 Documentation

Comprehensive user documentation is available for every area of Goobi, including the main programs, the plugins, scripts and the different use cases. The documentation is regularly updated and compiled on our central platform:

<https://docs.intranda.com>

We welcome suggestions, additions and criticism. You can contact us by email (support@intranda.com) or post a contribution to the [Community Forum](#).

6.2 Community and Forum

The [Goobi Community](#) is made up of hundreds of people - from over 80 institutions in 17 countries - who work with Goobi every day. The [Goobi Forum](#) is the main platform used by the community to discuss questions, suggestions and developments. The most common languages used on the forum are German and English:

<https://community.goobi.io>

6.3 User meetings

User meetings are held at least once a year. Members of the community report on the latest developments, projects and challenges in the world of digitisation. The meeting is not limited to users, however. It is open to any institution that wishes to learn more about the software suite and the active community.

The next user meeting will be held in Göttingen (Germany) on 25 and 26 September 2019.

6.4 Contact

For more information about Goobi, see:

<https://goobi.io>

If you require further details, please do not hesitate to contact our Goobi and plugin development teams.



intranda GmbH
Bertha-von-Suttner Str. 9
D – 37085 Göttingen

<https://www.intranda.com>
info@intranda.com