UOL Goobi Requirements Document - DRAFT v3.5

18/06/2024 Comments are tagged with the name of the responsible person/group

Purpose of this document

This document is intended as a briefing to intranda from Digitisation.IO about the workflow requirements for UOL to use Goobi as its mechanism to manage digitisation production. This document will detail the various workflows required, the plugins needed and development requirements.

- Based on this version of the document intranda will ask questions in order to clarify the services and developments needed and assign costs to them. This will likely necessitate clarification meetings with Digitisation.IO, UOL and Digirati as an iterative process.
- The final output of this document will be a quotation (offer) from intranda for services which will enable all of the UOL requirements to be met.
- This document will also form the basis for the final Digitisation.IO report to UOL.
- The information which has informed this document was gathered on three separate onsite workshop days delivered by Geoff Laycock.
- Areas where Digitisation.IO believes that Goobi developments required will be tagged as #Development
- There are still areas which require testing and clarification by UOL, Digirati and intranda at the time of writing (see date above)

Headline principles

1. Goobi is to be used to trigger new workflows in one of 2 ways:

- a. #done From an excel import
- b. By polling (requesting) digitisation tickets from the EMu requests module API. #Development
  - #done Geoff What means polling here exactly? Is this something repeated automatically?
  - #done Geoff-reply This means regularly requesting from the system every 15 minutes or so to see if there are new requests. Ideally the plugin could be configured to adjust that time so that, if it needs to be extended to 1 hour then that could be set in the configuration.
  - #done Leeds Goobi to poll Emu every 1 hour for request module
    - #done Timing of polling needs to be configurable

- #done User requests and internal staff requests go into Emu and link to object/collection – these are then polled and pulled into Goobi to create a new digitisation ticket/request. A flag is needed when there are no object/item level records and therefore the selected metadata records in Emu to attach the digitisation output/s to. Is this a new step in the Goobi workflow? This flag would either stop the digitization or the ingestion to Fedora/Preservation API we need confirmation from intranda as to where this sits in the workflow. A report on what tickets/requests are held up at this stage will enable UoL to understand the cataloguing backlog. N.B. Some of the content to be migrated from the Z drive will need this flag in place.
- intranda to confirm they can poll Emu via API for digitisation requests
  - #done Leeds how exactly can we poll new records if we do not know their identifiers, how to find out about new requests? This will be based on the last time the requests have been polled, so goobi needs to record this date/time
- #done Excel imports are for item/object metadata.
- #done Having a separate workflow for digitisation without cataloguing is necessary, but not just for images to be joined with catalogue records. The requirement may actually be two workflows:
  - #done A workflow to capture and store images, ready to be matched with records/IDs when catalogued. Images would be stored in a `holding area` until this is possible.
  - #done A workflow to capture images which will not be joined with records/IDs or added to stores/catalogue (e.g. digitisation of non-library content.) These images would bypass the `matching` step and go straight to the `sending` stage._
- #done Items cannot be transferred into Fedora without a persistent ID, link to an item level record in Alma or Emu, and rights/access/copyright/sensitivity information.

2. Each process created will have a `unique system ID` for either ALMA or EMu this should be stored in the `CatalogIDDigital` field in the METS file for each process.

- #Leeds Ids to use will be confirmed (Alma/Emu or new ID service) – Peter and Anthony to confirm
  - #done Examples of METS https://dcd.intranda.com/viewer/index/

3. Goobi will be required to import selected metadata fields from either ALMA or EMu automatically early in the process template. #Development

- #done Andrey For ALMA the request should work directly (still to be tested by Andrey).
- #intranda For EMu a custom development needs to take place to do the import and must be included in the pricing proposal then.
- #done Geoff-reply What do you need intranda to be able to test this and find out how much development is needed?
  - intranda needs detailed information for:
    - #done Leeds how to poll (know about new records in EMu) [request module - AS to confirm]. Example of request module given to Steffan/Simon

- #done Leeds Anthony to provide API endpoint to query digitization requests, created within a certain time frame, carrying the "digitization pending" status
        - #done Leeds have a concordance between EMu-json and METS (Leeds to confirm metadata to carry across.) Current metadata from EMu has been sent to steffan/simon
        - #done Leeds Create the crosswalk that we talked about in our EMu meeting
    - #done Have intranda been able to test this, can they share the response?

4. Goobi will be used to automatically or manually link master digital media files to processes and their metadata. ~~#Development~~

- #Done This is out of the box functionality. No further features are needed.
- #done Geoff-reply Great news.

5. Goobi will, in certain process templates, be used to automatically crop image files using Layout Wizzard

- #intranda include this in pricing proposal then
- #done Geoff-reply Yes, please include Layout Wizzard

6. Goobi will be used to enhance the metadata of each process in a variety of ways including pagination, structure, OCR to ALTO files, ICR using Transkribus to ALTO files.

- #intranda include this in pricing proposal (using the numbers from the amount named in `Fixed Price Services`)
- #done Geoff-reply Thank you.
- #done Leeds Pagination, structure are crucial for the digitisation workflows, OCR and ICR should be optional.

7. From the launch of the system, the new approach at UOL will be that even if a user requests digitisation of a small percentage of an item (such as a book) then the team will digitise all of the item.

- #done Leeds to review approach at 6 month intervals to understand impact and develop internal decision tree [future action]

8. At the end of processes where a user has requested an item to be digitised Goobi will need to deliver to them their images by offering a download link (similar to WeTransfer) but only of the specific pages/files requested. This may also require the generation of surrogate files for delivery. The system will need to remind the requesting person to download their files and it must also record when the files have been downloaded to enable the closing of the process. #Development

- #done Geoff does this software for the download link and the reminder exist already or shall this be part of the pricing proposal?
- #done Geoff-reply No, the software does not exist. I was hoping that something similar in the open source environment could be adapted for this. In any case the download software will need to be developed.

- #done Leeds Surrogate files will be generated by the IIIF services for all open content in Phase 1 of DLIP and closed content in Phase 2. In the interim WeTransfer will continue to be used. Development by Intranda is not needed for this.

9. #done Leeds Access rights/sensitivity/licence and copyright information must be in place for an item/object and stored in the METS before it can be ingested into the digital preservation repository, at item/object level these will be pulled across from Emu and can be overwritten at a file level.
   - #Leeds Meeting to be arranged (Jodie and Geoff) to discuss workflow needs for recording access rights/licencing/sensitivity and copyright needs during digitization process to ensure this information is in the METS files and content is not made available to users inadvertently.
   - will happen on Fr., June 7th
10. At the end of all processes the master files and the metadata will be delivered to the S3 environment for ingestion into the digital preservation repository. #Development

- #done Geoff this delivery to S3 is a separate action that should take place before the actual ingest? If so then the copying of files and the ingest will be two separate workflow steps.
- #done Geoff-reply From the meeting we had with Digirati my (limited) understanding was that you deliver to S3 first, get an S3 location reference and then trigger the ingest by passing a message to the Digirati API to trigger the ingest to happen. UOL is this assumption correct?
   - #done Leeds Yes, it is.
   - This process is described in https://github.com/uol-dlip/docs/blob/rfc-001-what-store-in-fedora/rfcs/003-preservation-api.md and in the sequence diagram https://github.com/uol-dlip/docs/blob/main/sequence-diagrams/goobi_api_interaction.md.
   - Digirati will produce a `quickstart` guide for Intranda.
   - UoL and Digirati are assuming that the METS Goobi will produce carries Pronom information in the same way METS produced by Goobi at Wellcome does - see https://github.com/uol-dlip/docs/blob/rfc-001-what-store-in-fedora/rfcs/003-preservation-api.md#mets-obligations
   - The ingest is asynchronous. When submitting an ingest job you get an immediate response, an `ImportJobResult` (see https://github.com/uol-dlip/docs/blob/rfc-001-what-store-in-fedora/rfcs/003-preservation-api.md#importjobresult). You can poll this to see how it's getting along (the status field).
      - #done Leeds please make the links accessible for us to be able to read them

11. After successful ingestion has been verified, and after a waiting period, the master files will be deleted from Goobi and only low resolution thumbnails will be kept for reference. #Development

- #intranda take the installation and configuration of multiple additional plugins into account in the pricing proposal.

12. If, after completion, a process needs to be edited then Goobi will pull the original master files from the digital preservation repository so that changes can be made (additions to files, deletion of files etc.) before reexporting the changed files to the digital preservation repository. #Development

- #done Geoff The workflows need to be `extended` if an existing process has to be adapted again, so that new (repeated) workflow steps will be added?
- #done Geoff-reply Yes, my understanding from your comments below is that it is better to include correction steps as deactivated steps in all workflows that can then be reactivated to enable the process to be worked on again. This is preferable to moving to a new process template. On this basis `extending` the process templates would work better.
- #done Geoff we would extend the workflow only if it is needed for corrections. this is easier understandable for users then
- #intranda after having the final API for the ingest this needs to be taken into account for the pricing proposal
- #done Leeds Files will not need to be retrieved if a metadata only update is needed
  - Described in RFC at https://github.com/uol-dlip/docs/blob/rfc-001-what-store-in-fedora/rfcs/003-preservation-api.md#export-creating-a-deposit-from-an-existing-digital-object
  - Earlier sequence diagram: https://github.com/uol-dlip/docs/blob/main/sequence-diagrams/pull-from-repository.md

Fixed Price Goobi Services Requirements

This section details the headline requirements for Goobi as they relate to fixed price services which intranda offers such as installation, maintenance, OCR, ICR, Layout Wizzard, support, configuration, consultancy, and training.

Installation requirements

- Install Goobi workflow to a UOL provided S3 server as a `TEST Environment`
- ~~Install Goobi workflow to a UOL provided S3 server as a `PRE-PRODUCTION Environment`~~
- Install Goobi workflow to a UOL provided S3 server as a `PRODUCTION/LIVE Environment`
- #done Leeds Only 2 environments are required test and production.
- Provide a mechanism for the UOL team to migrate settings and changes from one environment to the next in the order ~~`TEST > PRE-PRODUCTION > PRODUCTION/LIVE`~~ `TEST > PRODUCTION/LIVE` #Development
  - #done Geoff in the past such data migration between different stages was never done automatically. Test systems tend to become outdated usually. How shall such a data migration between these stages look like exactly (metadata, binary data, configuration files, database content, users etc.)
  - #done Geoff-reply My understanding of this is that the migration should relate to the software itself and its settings only. there will be test data (metadata and binaries in processes) in the test and pre production systems that are used to test the various configuration settings only. these would not need to be migrated over. Test would be used for admins to try out different configurations and new plugin developments by intranda. Once testing is complete in test then the settings would be migrated to the pre production server for User acceptance testing. After that, they would migrate to the production server. So, in summary: Goobi

version, plugins and the configuration files would be migrating over. NOT the binaries, the processes, or the users. intranda does that make sense? UOL does this tally with your view on the functionality of the three environments?
  - #done Leeds We would use Puppet to do this.
- Enable single sign on to each environment using Active Directory #Development
  - #done Geoff is it really a regular Active Directory to integrate?
  - #done Geoff-reply I am not exactly sure about this. I know that there is a SSO mechanism and I remember that it was discussed in one of the early meetings with the UOL team. I am also confident that it is Microsoft based. UOL Please can you confirm how this will work in reality?
  - #Leeds UoL use Azure Active Directory, Brian to advise on

Maintenance

- Monitor and Maintain each server environment
- Update the Goobi operating system `at least annually` to the latest version in the order ~~`TEST > PRE-PRODUCTION > PRODUCTION/LIVE`~~ `TEST > PRODUCTION/LIVE`
  - #intranda take fixed update plan into the pricing proposals
  - #done Leeds At times agreed with UoL, with ideally at least four weeks notice, excepting critical security updates which should be scheduled as soon as possible.

OCR - ICR - Layout Wizzard Costs

- Annual OCR licence costs
- Annual ICR Transkribus licence costs
- Annual Layout Wizzard licence costs
- Per 10,000 image OCR costs
- Per 10,000 image ICR costs
- #intranda take these costs into the pricing proposals

Support

- Year 1 - 100 support hours
- Year 2 and subsequent years - 50 support hours
- #intranda take these costs into the pricing proposals
- #done intranda Please include hourly rates and how these hours are accessed.
  - #done Leeds intranda provides access to a support tracker where Leeds can post support requests and used support hours are tracked

Configuration & Consultancy Services

- A pool of hours to configure the workflows as described in this document (to be agreed after consultation between intranda, UOL and Digitisation.IO)
  - #done Geoff the amount of expected work seems a little difficult to calculate as it is a lot. We suggest to start with smaller `phases` or `milestones` instead of `EVERYTHING FROM START ON`

- #done Geoff-reply My understanding is that UOL would like a single fixed price for the project work. It is, of course, OK for the configuration work to be costed in phases (maybe one phase per process template/workflow) but the costs we submit must be for everything needed. intranda is that OK? UOL Is my assumption about this correct?
  - #done intranda Are these hours included in the 100/50 hours above.
    - #Leeds no, configuration and setup are separate tasks from support
  - #done Leeds Yes, we would like one quote with all costs in including optionals being broken down and clear dependencies.
- A fixed number of consultancy days for Digitisation.IO team members to provide hands on support and configuration training to UOL team members.
  - #done Geoff to take this into account for his pricing proposal
  - #done Geoff-reply Of course, I will add that into the proposal

Training

- General Goobi training for a team of between 5 and 12 UOL staff (Online)
- Configuration training for a team of up to 5 UOL staff (Online)
- Administration training for a team of up to 5 UOL staff (Online)
- #intranda take these costs into the pricing proposals

Workflows

There are currently X main workflows (`Process templates`). These workflows will be configured as process templates that will be installed and tested with UOL as part of the Goobi services from intranda and Digitisation.IO. Each process template may have variants with the addition of steps such as OCR, ICR, Cropping, etc. these additional steps will be detailed with the tag #Variant

E-Prints_ingest_workflow

This workflow will be invoked in batches shortly after the commissioning of the system. There are already many digitised records in the E-Prints system at UOL which need to be migrated to the digital preservation repository (the repository). In advance of this workflow being triggered in Goobi the master digitised files and corresponding METS XML files for each item will be stored by UOL in an S3 Bucket. Goobi will need to be triggered to ingest these images and METS files into Goobi as processes. in the process template the system will need to pull in metadata from EMu or ALMA, link the imported media to the metadata in the METS file and then export it all to the repository at the end.

- Leeds Leeds will undertake the migration from EPrints, however they need access to the documentation on how the import plugin/process works and what format is needed for the folder/file structure. Then skip to step 2.
  - #done Leeds which step 2 are you referring to?

Here is the current plan for the process template

Pre-requisite conditions:

Master media files and METS XML files will be in a nominated S3 Bucket ready for import

- #done Geoff can you please provide reference files/folders to see how the data really looks like. Very likely an import plugin or a converter would be needed to make those METS files readable for Goobi.
- #done Geoff-reply UOL Please can you provide example METS files for the E-Prints ingest as requested?
- #done Leeds not needed, please provide Leeds with an example of a Goobi import.
- #Leeds these are existing import plugins for your reference
  - https://docs.goobi.io/goobi-workflow-plugins-en/import/intranda_import_excel
  - https://docs.goobi.io/goobi-workflow-plugins-en/workflow/intranda_workflow_import_json
  - https://docs.goobi.io/goobi-workflow-plugins-en/workflow/intranda_workflow_liechtenstein_volksblatt_importer
  - https://docs.goobi.io/goobi-workflow-plugins-en/workflow/intranda_workflow_fileupload_processcreation
  - https://docs.goobi.io/goobi-workflow-plugins-en/opac/intranda_opac_generic_xml
  - https://docs.goobi.io/goobi-workflow-plugins-en/opac/intranda_opac_json

DRAFT process template:

1. Workflow plugin in Goobi where the S3 Bucket location can be entered into the system #Development

- #done Geoff do you mean an import plugin here? This will need to be developed based on the provided real data and formats
- #done Geoff-reply Yes, an import plugin from the S3 Bucket location given.
- #done Leeds to implement this

2. Goobi imports the METS files and media files, creating new processes #Development

- #done Geoff this is actually the same the first one right? So this should be just one single step
- #done Geoff-reply Yes, I just split them into two because as part of the import the S3 location will need to be given by the user so that the import plugin can locate the files to be imported. For this therefore I am expecting a text box in the user interface with a button to `IMPORT DATA`
- #done Leeds to provide Image folders for a mass creation of processes [RF or PE supply content - Send Bingley or coins if you need safe open content] Download link, password: DLIP2024

1. Goobi takes the `EMu IRN` information (e.g.: 163535) or the `ALMA BIBREF` information (e.g.: 991008723399705181) from the METS files, places them in the CatalogIDDigital field and records which system that it needs to request metadata from as a property #Development

- #done Geoff does this mean that the existing metadata in METS is incomplete and needs to be replaced or enriched?
- #done Geoff-reply I am not sure here. I expect that the METS file will contain limited data including the IRN/BIBREF and that the system will need to pull the remaining metadata from those systems.
- #done Leeds Is that assumption correct or do we expect that the METS file coming in from E-Prints will be complete and ready to push to the repository?
- #intranda plan the catalog request plugin before the export happens

2. Goobi imports selected metadata fields from either the EMu API or the ALMA API into the METS file #Development

- #Geoff can we please have a sample of before/after to see how we know what to request and what to take over into the metadata
- #done Geoff-reply My understanding is that the ball is with you on this as you now have access to EMu API and ALMA to be able to get records out based on the sample IDs provided. If you can send us what data you can get out of the systems then UOL can specify which fields are needed and where they should be mapped to in the METS file. intranda can you confirm that you have been able to extract data from the systems and provide samples of the data you have extracted?
- #done Leeds please advise which fields should be imported to enrich initial EPrint metadata [see initial list above as advised by AS]

3. Goobi links the media files into the METS file

- #done Geoff this is done automatically inside of the previous step and can be removed, right?
- #done Geoff-reply I assume so. UOL Will the linking of the binaries to the METS file be in place when the data is made available for the ingest?
- #done Leeds to implement this

4. Goobi creates the checksums for the digital files to be deposited and records that in the METS files. #Development

- #done Geoff we are waiting for confirmation from your side first
- #done Geoff-reply UOL Is the answer to this pending the finalisation of the ingest API being developed by Digirati?
- #done Leeds Yes, Goobi should create the checksums and store in the METS file. https://github.com/uol-dlip/docs/blob/rfc-001-what-store-in-fedora/rfcs/003-preservation-api.md#mets-obligations
- #done Leeds Add sensitivity/access/copyright/licensing step here. [step in EMu - review flag needed in EMu]
- #intranda include checksum creation and metadata configuration for access conditions in the pricing proposal
- #Digirati specify requirements for checksums: algorithm, SHA256?

1. Goobi creates a digital object in the repository #Development

- #done Geoff is the API of the repository now fix and will not change anymore? Where do we find the most up-to-date description of it?

- #done Geoff-reply You have received documentation I believe. UOL is this the final API version and do intranda have access to it at the moment or is it still being developed?
- #intranda to put the files and METS in S3 and call the preservation API which will then ingest into the Preservation Repository – https://github.com/uol-dlip/docs/blob/rfc-001-what-store-in-fedora/rfcs/003-preservation-api.md
- #intranda to take this into account for the pricing proposal (1. complete ingest; 2. ingest of only metadata OR files), if not done already

2. Goobi deposits the digital object in the repository #Development

- #done Geoff isn't it the same as the step before?
- #done Geoff-reply It might be. Again I remember from our meeting with Digirati and UOL recently that the creation of an object is one step and then the depositing of data for that object is a separate step. UOL have I got that right?

3. Goobi receives back a message/confirmation that the deposit was successful #Development

- #done Geoff isn't it the same as the step before or part of it?
- #done Geoff-reply It could be part of it but I understood that the ingest could take some time and therefore the return message will only be possible after some hours and then Goobi will need to receive and process a message signalling completion of the ingest to trigger the nest step (deletion of the master files).
- #done Leeds Is that assumption correct? [YES]
- #intranda

1. Goobi deletes the master images from the process and retains derivatives for future reference. #Development

- #intranda to include this into the pricing proposal for installation and configuration
- #done Leeds specify desired quality/size of derivatives [RF / PE review size - 72dpi 1280 on the longest side, criteria is that text needs to be legible, this may be too big or small? RF: proposed derivative spec seems fine for purpose. This won't be a one-size-fits-all solution so may need to be changed in future development.]

2. ~~done Goobi closes the process and moves the deposited process to a new project entitled `Deposited Processes` Development~~

- ~~done Geoff are you really sure you want this? I think this is redundant.~~
- ~~done Geoff-reply My logic for this is that if the template is extended to allow an open step to start a corrections process then the processes will have one open step at the point where it should be seen as completed. This will mean that it will still display in the processes page as all the steps will not be green or `completed`. If we move them to a new project and limit user access to that project then the processes can all sit in there with an open step and the users will not be distracted in the other projects. If there is a better/easier way however then we should go for that of course.~~
- ~~done Steffen to include the workflow change plugin and configure it as part of pricing proposal~~

3. ~~done Goobi changes the process template to a new process template called `Editing_Deposited_Processes` with the step: `Select editing action` set to OPEN.~~

   - ~~done Note: the intention here is that the unique identifier passed to Goobi in step 9 above will enable Goobi to pull back the master media from the repository ready for edits to take place.~~
   - #done Geoff Wouldn't it be more straightforward to `extend` this workflow with additional steps just in the moment when (if ever) this workflow needs to be changed again? It will be more natural to keep the original workflow and allows better understanding afterwards as everything is still there. And the best: it could be done over and over again for multiple changes in the future.
   - #done Geoff-reply I Agree that this would make more sense as then the original workflow information can stay with the process at all times. So basically we just add a series of deactivated steps to all process templates that can be activated later.
   - #intranda include workflow process extension in quote

Editing_Deposited_Processes

#intranda rewrite this as additional process steps This process template will be applied to all processes that have been deposited in the repository. All processes that have been deposited will be moved to the project `Deposited Processes` which only administrators will be able to see. After moving to the new project the process template will be changed to the `Editing_Deposited_Processes` template.

Note: it may be that to retain the process log information from the original processes that Goobi either adds the steps from the Editing process template to the end of the previous template or that the editing steps are built into all other process templates as deactivated steps which are automatically activated at the end of the original process using the plugin: Workflow change. this is TBC.

Pre-requisite conditions:

The process must be in the project `Deposited Processes` and the `Editing_Deposited_Processes` Steps must be activated with step 1 (`Select Editing Action`) set to OPEN.

- #done Leeds The `ImportJobResult` described above includes a digitalObject field. It is likely that it was also set by Goobi on the executed `ImportJob` that produced the `ImportJobResult`, but not always - sometimes the Preservation API might decide where to put something in the repository; the location returned on the ImportJobResult is definitive.

DRAFT process template:

1. ~~Select Editing Action: User uses property check boxes to select the actions that are required. These could include:~~

- ~~Delete process~~

- ~~Reimport metadata (from EMu)~~
- ~~Edit metadata~~
- ~~Edit media files~~
- ~~done Geoff we would suggest to use a Workflow Plugin to `Extend the workflow` based on preconfigured options, that will add the appropriate steps to the existing workflows and trigger the first of those added steps then. Would this be a more generic concept with lots of possible future use cases? We could call it the `Workflow Extender Plugin`~~
- ~~done Geoff-reply This sounds good to me and more elegant than moving processes between process templates and projects. I suggest that you base the offer for development on this idea.~~
- ~~intranda include this in the proposal~~

1. ~~Workflow change step to activate subsequent steps in the template~~

- ~~done Geoff with my suggestion this would not be needed anymore~~
- ~~done Geoff-reply Agreed.~~

2. ~~If `Delete process` is selected then Goobi will send a delete command to the repository Development~~

- ~~done Geoff we are waiting for a clarification then~~
- ~~done Geoff-reply Yes, we need to know the process for triggering a delete within the repository using the Digirati API. UOL Please can you ask Digirati if it will be possible to trigger a delete from repository action from the API that they are developing? I am assuming that this could be one of the outcomes from editing a process. Please can you also confirm that assumption UOL ?~~
- #Leeds An ImportJob could delete all the containers and binaries in a digital object. This would produce a new OCFL version and remove from the IIIF services; the previous versions would still be available to anyone.
- #Leeds could you clarify this? [PE is this true? that was my understanding]

1. ~~If `Reimport metadata` is selected then Goobi will reimport the metadata from EMu or ALMA using the `IRN` or `BIBNET` reference as already detailed (this is in situations where the metadata has changed in the EMu or ALMA systems) Development~~

- ~~done Geoff for such a case real data would be helpful again~~
- ~~done Geoff-reply As before, intranda please can you confirm if you have been able to access the EMu and ALMA systems to retrieve data?~~

2. ~~If `Edit metadata` is selected then the users will be enabled to access the METS editor to make changes to the metadata.~~
3. ~~If `Edit media files` is selected then Goobi will pull back the master media files from the Repository.~~

- ~~done Leeds If the user is making changes to metadata that will affect the METS file only, which presumably has its own version of, then Goobi doesn't have to export the entire digital object. If it has the METS file already, the user can perform actions in Goobi that change the METS file, Goobi can put just that METS file in S3 in a Deposit for the existing digital object, and then import it.~~

4. ~~Select media files editing actions: User uses property check boxes to select the actions that are needed. These could include:~~

- ~~Replace files~~
- ~~Delete files~~
- ~~Rotate files~~
- ~~Add files~~
- ~~done Geoff this sound soooo complex and difficult to use for real users then. Couldn't it be easier? Should be done with the `Workflow Extender Plugin` as well anyway.~~
- ~~done Geoff-reply I am happy with anything that simplifies the steps and the process. Please therefore build in the functionality to the workflow extender plugin.~~
- ~~Leeds what are your requirements for editing media files?~~ [#Leeds done internal decision on file mangaement policy]

1. ~~Workflow change to activate subsequent steps in the template.~~

- ~~done Geoff with my suggestion this would not be needed anymore~~
- ~~done Geoff-reply Great news!~~

2. ~~For replacement, deletion, or adding of files the `file upload plugin` will be provided to the user in the task page.~~

- ~~done Geoff still feels too difficult this way. However this functionality is there already.~~
- ~~done Geoff-reply Good news. If there is a file browser or the Image QA plugin allowed deletion of the media and resorting the order etc. then that could also be used? Maybe you at intranda could consider that development as part of the other enhancements suggested?~~
- ~~Leeds what are your requirements for editing media files?~~ [#Leeds done - repeat of line 278]

1. After file upload the QA plugin will be provided for checking the files. (Note: if the selected action is rotate files only then step 8 will remain deactivated and only this step will be provided)

- The QA plugin will need to be developed to allow for the QA of moving image and audio files #Development
  - #done this does exist already and does not need to be extended for audio and video
  - #done Geoff-reply Great news. Another example of parallel development in the Goobi world benefitting everyone! :-)
- The Rename feature of the QA plugin will need to be developed to enable the fully manual renaming and reordering of files #Development
  - #done Geoff what exactly would have to be done then. As long as the reordering is allowed no users should do file naming. Goobi could do that already automatically in the background.
  - #done Geoff-reply That makes sense to me. if drag and drop reordering is allowed and then the number suffix to the file names is re created by Goobi then separate renaming would not be needed.

- #done Leeds do you want to keep sequential file names up to date on reorder? [how will this affect digital preservation and file monitoring? assume it won't as it relates more to OCFL]
- #intranda consider this for quote: extension of QA plugin for renaming and reordering of files
- #intranda after image reordering in QA plugin, file names should be updated sequentially

1. Metadata edition: after all media steps the metadata edition step will be enabled to allow for pagination or structure changes.

- #done this does exist already

2. #Variant if OCR was part of the original process template then it will be re run here if selected

- #done this does exist already

3. #Variant if ICR using Transkribus was part of the original process template then it will be re run here if selected

- #done this does exist already

4. ~~Re-export to the repository: Goobi will re-export to the repository using the standard procedure (See steps 6 to 10 of the E-Prints ingest workflow above)~~

- ~~Steffen to take this into account for the pricing proposal, if not done already~~

5. ~~Reset process template: the Editing_Deposited_Processes steps will be reset to a state where future edits can be done if needed.~~

- ~~done Geoff I can only highly suggest not to work this way. It is so difficult to understand later what happend. Let's use the `Workflow Extender Plugin` better for this use case.~~
- ~~done Geoff-reply I agree Steffen. Let's proceed on that basis.~~

Digitisation_Requests_Workflow

- This workflow is for item level requests for digitisation that are delivered by the studio team using a variety of equipment depending on the materials requested for digitisation. This is the most complex workflow as it involves pulling the digitisation request information from EMu and also a newly developed mechanism to deliver a selected amount of digitised content to the requesting user.
- Note: A user may request (and pay for) selected content from an item, for example 10 pages from a 100 page book. In that case the studio team will digitise all of the book and the entirety of the book will be delivered to the digital repository. However only the pages paid for by the client will be delivered to them directly.

Pre-requisite conditions:

- A request tracker ticket with the status `Pending digitisation` will be exist in EMu with a request identifier.

- In the request there will be a list of EMu IRNs (Minimum 1 but there could be multiple) of items to be digitised and other information such as page numbers requested etc.

- The physical item will have been retrieved and will be on a shelf for the studio team to collect or, if it is large then the location of the item will have been communicated to the studio team.

- Goobi functionality will be developed to enable sub projects to be created. In this way there will be a master project called `Digitisation Requests` with sub projects for each EMu digitisation request ticket. Under each ticket there will be 1 or multiple processes running independently. #Development

  - #done Geoff why not use the Batch functionality of Goobi to group items together?
  - #done Geoff-reply Happy with that as long as it makes it easier for a user to see all the aspects that they are working on. I like the idea of each sub-project being all the processes for a request as it makes it easier to track completion.
  - #done Geoff a `sub project` means process here, right?
  - #done Geoff-reply No. The project is `Digitisation Requests`. The `sub project` is a request (ID: 12345 from the EMu requests module which you have ingested into Goobi) Under the `sub project` there are `processes` from 1 to unlimited that are part of the request. the request (`sub project`) is not complete until all of the `processes` are complete.
  - #done intranda show how this can be done with batches

- Goobi functionality will be developed to enable the assignment of a responsible user for each sub project. This will take the form of a Workflow plugin `Assign Digitisation Request Responsibility` where sub projects will be held until a manager has assigned responsibility for the sub project to one person. (possible development of the projects list page) #Development

  - #done Geoff if we use the user assignment plugin for steps and work with Batches then we would have that functionality already
  - #done Geoff-reply Fine with me. I wasn't even aware that there was a `user assignment plugin`.
  - #intranda include in pricing proposal: extend user assignment plugin to cover whole batches

- Goobi reporting to be extended to enable status tracking of the sub projects so that a report of current status is viewable and exportable. #Development

  - #done Geoff the Batch functionality gets visualised a charts on the dashboard. Not exportable, but easy visible. The alternative would be a specific statistics plugin

- #done Geoff-reply OK, that is fine then if we go for the batches approach. Please read my earlier answer to see if batches would still work to achieve what I believe UOL needs here.

- Goobi Dashboard to be developed to allow for pie chart or graphic progress views on the dashboard for all live master projects #Development

    - #done Geoff what is a master project? Doesn't this fit to what I suggested here with Batches?
    - #done Geoff-reply As before, I am not sure that the batches would work here (see above) I ==DO== think though that there needs to be pie charts of project progress developed to go onto the dashboard for easy visibility.
    - #intranda ask for clarification about this

- Goobi QA plugin to be set to only allow QA to be carried out by a different user than carried out the digitisation step. #Development

    - #intranda take this into account for the pricing proposal if not done already

- Goobi Development to set at a project level the percentage of processes that will be required to go through Quality Assurance. So, for example, at the beginning of a large project the QA percentage could be set at 100% as default causing all processes to have the Quality assurance step activated. Later this could be changed, say, to 50% and this will trigger Goobi to randomly activate QA on 50% of processes. The percentage should be able to change at any point and then Goobi will refresh all processes where QA has not happened yet to the percentage level. #Development

    - #done Geoff what is the percentage expressing? Number of images? Number of processes? Number of errors in Scans? How is this calculated or proven?
    - #done Geoff-reply The percentage would be number of processes. the system would not need to calculate the number of errors and therefore when the percentage set for the project should be lowered. that would be a decision from the team based on what they have found. The key thing is that when their percentage is changed to, say 10%, only 1 in every 10 processes would go to the QA step. That percentage needs to be set at a project level in the configuration and then, after setting the correct number of processes are sent for QA only.
    - #intranda take this into account (QA on a percentage of processes in a project, defined on a project level)

- Ingest into Digirati

- Send mail out to the customer and to Request tracker system with URL for access

- ~~Goobi Development to enable users to select which images from a complete item are needed by a customer. Image selection tool?~~

    - ~~#done Geoff Image Selection Plugin does exist already. Can we use this?~~
    - ~~#done Geoff-reply Yes, if it works better and is designed so that it is easy to use.~~
    - ~~This is not needed anymore. All images of an object will get ingested to Digirati~~

- ~~Goobi Development to deliver to clients their ordered digitised content by offering a download link (similar to WeTransfer) The system will need to remind the requesting person to download their files and it must also record when the files have been downloaded to enable the closing of the process.~~

    - ~~#done Geoff What kind of system is it? Does it need to be developed? Does it exist already. Does it provide an API?~~
    - ~~#done Geoff-reply This needs to be developed. I thought that maybe something similar already may exist in the open source world that you could base it on maybe. Otherwise developed from scratch I think.~~
    - ~~for the delivery to the customer we have the information about the customer as properties (originally from EMu)~~
    - ~~based on the customers order specific derivatives (just for the requested pages) might be created~~

DRAFT Process template:

1. Goobi will poll the EMu request tracker API at fixed intervals (TBC every 30 minutes during working hours and days) to pull back NEW requests with the status `Pending digitisation`. Each New request will be imported into Goobi with the following steps:

- The request ticket will be imported as a sub project under the master project `Digitisation requests` #Development
    - #done Geoff please see my comments above
    - #done Geoff-reply replied above
    - #done Leeds every 60 mins
    - #Leeds make sure that the requests to EMu allow us to get a list all orders younger than a specific date. Email of EMu requests sent to Simon/Steffan
    - #intranda test if the communication works as written in the mail from Anthony
- The metadata from the requests will be imported as project properties (e.g.: specifications, information, email address of requesting person, Request tracker ticket number etc.) #Development
    - #done Geoff please see my comments above, as we need real data for testing
    - #done Leeds please can you provide access to the requests module data via the API to intranda
    - #intranda check the samples that Leeds provided: Any request with the markers `Digitisation` / `Download Request` which are also marked `Completed` are standard examples. See RRREQ-2024-171 and RRREQ-2024-116 as two recent examples.
- The project properties will be viewable in the task page of each sub process in a properties box (possibly through the copying of the sub project properties to the process properties) possible need for a new properties area in the task page #Development
    - #done Geoff what do you expect here as new functionality needed?
    - #done Geoff-reply Basically the properties that have been imported from the EMu module will have the `Brief` for the project in there, things like amounts paid, quantities, deadlines, email address of the requesting person etc. It would be helpful to have these showing as information in all the processes for a single

request so that the users have it as reference. My suggestion would be to allow multiple properties boxes on the task page so you had a read only area with all this information and then an editable properties area if we need it to gather information from the users like check boxes etc.

- #done intranda present a similar development tomorrow
- #intranda plan an extension of the batch functionality to allow properties per batch that get shown inside of the tasks as well
- Each item in the request will be created as a process under the sub project with the EMu IRN as the Process title. #Development
  - #intranda the import mechanism needs to take this batch-assignement into account
- Goobi will automatically assign a deadline (set centrally at, for example, 1 month) to the sub project #Development
  - #done Geoff this would work with Batches already. However: What to do if a deadline is exceeded?
  - #done Geoff-reply good news. I guess the batch should be flagged in some way (maybe on the dashboard) if it has exceeded the deadline?
  - #intranda include in quote (for batch)
- A manager will assign a user to the sub project in the new workflow plugin `Assign Digitisation Request Responsibility` as a property. #Development
  - #done Geoff this means: per task, right?
  - #done Geoff-reply no to all the processes in the sub project.
  - #intranda plan the extension of the user-assignment-plugin to be usable for entire batches and to allow assignment of groups and users (not users only)
- Goobi will assign the user assigned to the digitisation tasks (i.e. Item collection, assessment, scanning steps etc.) in the process template of each of the processes in the sub project (so that they only appear in the task list of that specific person). #Development
  - #done Geoff This does need more clarification
  - #done Geoff-reply What I mean here is that when the user is assigned to the processes in a sub project (a digitisation request with multiple item level processes in it) then the user needs to be assigned only to the manual steps. In other words not to the automatic ones. This means that the tasks for the manual steps (scanning etc.) will only appear in the task list of the assigned user and not everyone else in the user group.
- All the processes will be open at the step assessment in Goobi.
  - #done Geoff we need concrete samples to understand how the users request would look like and which data we get exactly there
  - #done Geoff-reply agreed, I have requested that data from UOL above.
- #done Leeds decide how to proceed: use batches or sub projects or single processes
- #done Leeds decide how user assignment should be done [Leeds to document]

1. Item Collection: the assigned person will collect the item from the shelf or location and will record their workstation ID as a property.
2. Assessment: the assigned person will look at the original and will assess it in comparison to the sub project properties (request ticket information). The user will select from 2 property options: Clarification needed or Ready to digitise.

3. Workflow change: If `Clarification needed` is `TRUE` then: the clarification step will be activated and opened, the `waiting for clarification` will be set to `LOCKED`. If the `ready to digitise` process is selected then the `Digitisation Triage` step will be set to `OPEN`.
4. Clarification needed: User enters comments into a text box and closes the task. Goobi then emails a specific email address with the Request tracker ticket number property (imported from EMu as a sub project and then process property) plus the phrase `Digitisation clarification needed: ITEM ID:` and the EMu IRN of the process in the subject line, the comments text in the email body and the email address of the assigned user in the CC box. #Development

- #intranda take all this configuration and customization into account for the pricing proposal

5. Waiting for clarification: Process property dropdown (Required) with the options: `Please select/Proceed with digitisation/Cancel process` and a box for the user to copy and paste the response text to the enquiry into a process property.

- #done Geoff can you please explain further what happens here?
- #done Geoff-reply Essentially here we are waiting for clarification which happens outside of Goobi in a different system (that we are not linking to!). So it is a task step where the option is selected and then any text from the clarification is pasted into a property.

6. Workflow change: if `Proceed with digitisation` is recorded then the `Digitisation Triage` step will be set to `OPEN`. If `Cancel process` is recorded then Goobi will cancel the process and move it into a project of cancelled processes. #Development

- #intranda plan this for proposal

7. Digitisation triage: Here the user will select the digitisation methodology for the item based on its size and format (Dropdown). These options to include:

- Moving image
- PhaseOne
- Canon 35mm
- ATIZ book scanner
- Flatbed scanner
- Versascan
- Microfilm/fiche
- Sheet fed scanner
- Roboscan
- Audio

8. Workflow change: If the method chosen will result in images as an output (Phase One, scanners etc.) Then the `Image output settings` step will be set to `OPEN` and the `Image upload` step will be set to `LOCKED`. If the method is `Moving image` then the `Moving image supplier liaison` step will be set to open and all other Moving image steps will be set to `LOCKED`. If the method is `Audio` then The step

`Upload Audio File` will be set to `OPEN` and the `Audio` transcoding step will be set to `LOCKED`.

9. Image output settings: Here the user will record:

- whether cropping will be required,
- whether OCR is required,
- whether ICR by Transkribus is required,
- if a specific Transkribus model identifier is to be used, and if so,
- the Transkribus model identifier that has been trained for the material.
- is Pagination required
- is structural metadata required

10. Image upload: the user will upload the images using the file upload plugin.
11. Moving image supplier liaison: check box for `Supplier engaged`
12. Moving Image supplier commissioned: check box for `PO raised`
13. Moving Image logistics: check box for `Logistics arranged`, date picker for the collection date and notes for the collection details.

- #intranda Date picker is not available in properties yet, put into pricing proposal

14. Moving Image media prepared: check box for `Materials packed and ready`
15. Moving Image media dispatched: check box for `Moving images media dispatched`
16. Moving Image media files received: check box for `Digitised files received`
17. Moving image file upload: user uploads the file to Goobi using the file upload plugin.
18. Upload Audio file: user uploads the file to Goobi using the file upload plugin.
19. #done intranda Where will access rights/licencing/sensitivity checks be done in the process and added to METS, is this ahead of QA?
20. #done Leeds after QA, based on images
21. Quality Assurance: after upload for all file types they will be reviewed in the QA plugin. this plugin must be developed to ensure that a different user than the one that did the digitisation reviews the output. #Development The plugin must be developed to allow for the display of moving image and audio files. #Development The Reorder feature of the QA plugin will need to be developed to enable the fully manual renaming and reordering of files #Development.

- #intranda plan this development and take it into the pricing proposal
- #done Geoff please explain what it needed to be adapted in the renaming
- #done Geoff-reply based on your earlier comment I don't think the renaming is needed here as long as reordering is possible and Goobi changes the file suffix automatically after that reordering has happened.

20. Automatic file renaming: File rename plugin based on reordered files
21. #Variant If cropping is required - Automatic image analysis.
22. #Variant If cropping is required - Layout Wizzard (Note: Layout Wizzard workflow plugin should also be installed)
23. #Variant If cropping is required - Automatic cropping (to Cropped master folder retaining the master file)
24. #Variant If OCR is required - OCR step

25. #Variant If ICR is required - Transkribus step (using specific model ID if one was provided in step 9 above)
26. #Variant If Moving image - Transcoding plugin using FFMPEG to create Mp4 with H264 codec #Development

- #done Geoff can we please have reference files for testing here?
- #Leeds please can you arrange for some reference files (source and target format) to be transferred to intranda as requested? [no open content to transfer, can find low risk content if needed]

27. #Variant If Audio file - Transcoding from FLAC to WAV/Mp3 #Development

- #done Geoff can we please have reference files for testing here?
- #done Leeds please can you arrange for some reference files (source and target format) to be transferred to intranda as requested? [RF transfer some Liddle files] Download link, password: DLIP2024

28. #Variant if pagination is required then the user carries out pagination in the METS editor.
29. #Variant if structural data is required then the user assigns structural data in the METS editor.
30. Goobi re-imports selected metadata fields from either the EMu API or the ALMA API into the METS file #Development

- #intranda take this into account for the pricing proposal if not done already

31. Final QA: carried out by a different user than the main person assigned to the sub project and using the METS editor tool.

- #intranda take this into account for the pricing proposal if not done already

32. #Variant For moving images - Item received back from client: check box to select when item has been received. Date picker for received date. Quarantine time period (drop down list)

- #intranda take this into account for the pricing proposal if not done already

33. #Variant For moving images - Quarantine: wait for x time period (set at the previous step)

- #done Geoff How long is the waiting? Will it be flexible?
- #done Geoff-reply No, it is not flexible. it will be set by the UOL team
- #done Leeds How long is the quarantine period usually please? 2 weeks

34. #Variant For moving image - Pest Checking: Check box to record pest checking carried out. Sub process here for what to do if pests are found.

- #done Geoff what is meant here? How does it affect Goobi?
- #done Geoff-reply we are using Goobi to record physical steps here only. Similar to when we say `item has been returned to the store`
- #done Leeds 2 week period for quarantine

37. Item returned to storage: Check box for user to record that the item has been returned and a text box to enable the recording of the return location (Shelf)
38. Goobi creates the checksums for the digital files to be deposited and records that in the METS files. #Development

- #done Geoff we are waiting for a confirmation here first.
- #done Geoff-reply Yes, confirmation requested from UOL
- #intranda double check the documentation for the Digirati system which checksums are expected and how to ingest

1. Goobi creates a digital object in the repository #Development

- #intranda take this into account for the pricing proposal if not done already

2. Goobi deposits the digital object in the repository #Development

- #intranda take this into account for the pricing proposal if not done already

3. Goobi receives back a message/confirmation that the deposit was successful *TBC* #Development

- #done Geoff we are waiting for a confirmation then
- #Leeds confirmation requested

4. Goobi deletes the master images from the process and retains thumbnails for future reference. #Development

- #intranda take this into account for the pricing proposal if not done already

5. ~~Goobi closes the process and moves the deposited process to a new project entitled Deposited Processes Development~~

- ~~done Geoff Why not leave the items in the projects?~~
- ~~done However this functionality does exist already~~
- ~~done Geoff-reply With the workflow extender plugin you suggested maybe this is no longer needed?~~

6. ~~Goobi changes the process template to a new process template called Editing_Deposited_Processes with the step: Select editing action set to OPEN.~~

- ~~Note: the intention here is that the unique identifier passed to Goobi in step 43 above will enable Goobi to pull back the master media from the repository ready for edits to take place.~~
- #done Geoff I would use the `Workflow Extender Plugin` to be developed first for such use cases.
- #done Geoff-reply Agreed.
- Send out E-Mail to user and the Tracking system with link to published documents (one mail per object)
  - #intranda plan this

Large_Scale_Digitisation_Workflow_Internal

This workflow is for large scale digitisation projects where the work will be carried out by UOL staff. examples of this could be the digitisation of Theses. It is envisaged that this will likely be for image based digitisation in the medium term but a #Variant could be created for moving image or audio projects in the future, based on this process template.

Pre-requisite conditions:

- All items to be digitised are catalogued to item level on EMu or ALMA.
- A list of records (one per line) to be digitised with EMu IRN or the ALMA BIBNET references are pasted into the mass import interface of Goobi.
  - #done Geoff can we have such a file please?
  - #done Leeds Please can you create a sample of such identifiers as a list (for alma and for EMu) for intranda [see Internet Archive record - https://archive.org/details/b21524956 image to download from there, and record is here https://explore.library.leeds.ac.uk/special-collections-explore/241077]
  - #intranda take this into account for the pricing proposal (no excel anymore)
- A project has been set up in Goobi with a total number of processes, estimated number of pages, a start date, and a deadline for completion.
- The initial percentage for quality assurance has been set. See #Development above.
  - #intranda to take this into account for the pricing proposal, if not done already
- Goobi mass Upload plugin has been developed to allow barcode uploads for specific projects while allowing non barcode (filename based) uploads in other projects. This is because, at the moment, it is only possible to turn the barcode function on or off. #Development
  - #done Geoff how exactly will the files look like then and how are the processes named then?
  - #done Geoff what needs to be developed here exactly? You are talking about the mass upload plugin here, right? Not the mass import plugin.
  - #done Geoff-reply Apologies it should have read Mass upload The issue is that currently in the settings barcodes are either allowed or not. If they are set to allowed then filename mass uploads do not work because the system is looking for barcodes. If they are not set to allowed then mass uploads with barcodes would not be linked to processes because the filenames would not match. The plugin needs to be developed to allow both depending on the project in the configuration. So: Project A (external digitisation by a company where they are required to do the file naming properly before sending the data back to UOL) is happening at the same time as project B where the internal team are doing mass digitisation using barcodes.
  - #intranda add a selection into the user interface to let the user decide between barcode detection or file name analysis
- Goobi project settings will be extended to allow for the setting of requirements for cropping, OCR, ICR, surrogate requirements, the need for pagination, the need for structural metadata etc. #Development
  - #done Geoff can you please explain further what is needed here exactly?
  - #done Leeds Geoff-reply Thinking about it I don't think we need this any more because we can use process templates to set these things much more easily.

- #done gets ignored for now

DRAFT Process template:

1. User will paste a list of identifiers into the mass import interface with identifiers from EMu or ALMA.

- #done Geoff can you please share these Excel sheets as sample with us?
- #done Geoff-reply Files requested above.
- #intranda check if development is needed for this

2. ~~Goobi will Automatically create the processes.~~

- ~~done Geoff this should be removed as it is part of the previous step~~

3. The user will download and print the batch docket and insert the barcode into the front cover of all the items to be digitised.
4. Item Collection: an assigned person will collect the item from the shelf or location and will record their workstation ID as a property.

- #done Geoff Who is an assigned person? Will the person select himself or is someone else assigning?
- #done Geoff-reply The studio manager (Rob) will assign the person.
- #intranda plan the user and user-group assignment here for multiple steps

5. Assessment: the assigned person(s) will look at the original and will assess its condition and if it needs conservation treatment before digitisation with the options: Conservation Assessment or Ready to digitise.
6. Workflow change: If `Conservation Assessment` is TRUE then: the `Conservation Assessment` step will be activated and opened, the `waiting for conservation` will be set to LOCKED. If the `ready to digitise` process is selected then the `Image upload` step will be set to OPEN.
7. Conservation Assessment: Conservation user enters comments into a text box and closes the task.
8. Goobi then emails a specific conservation email address with the phrase `Conservation assessment needed: ITEM ID:` and the EMu IRN of the process in the subject line, the comments text in the email body. #Development

- #done Geoff what is the `comments text` here? Where is it coming from?
- #done Geoff-reply A property.
- #intranda include in quote: e-mail plugin

9. Waiting for conservation: Process property dropdown (Required) with the options: `Please select/Proceed with digitisation/Delete process` and a box for the user to copy and paste the response text to the enquiry into a process property.

- #done Geoff Can you please explain this a bit deeper? I don't get it.
- #done Geoff-reply OK, so if after initial assessment the studio team are concerned about the condition of the item then they will want the conservation team to assess it and make a decision about whether it should be digitised or not. After they have decided they will email the studio team. The team will then record the decision as a

property in the task and then workflow change will kick in (next step) to turn things on or off based on the property information.

- move process to a `deactivated item`-project in case it gets cancelled

10. #done Workflow change: if `Proceed with digitisation` is recorded then the `Upload images` step will be set to `OPEN`. If `Delete process` is recorded then Goobi will delete the process entirely. #Development

- #done most of this does exist already
- #done check if complete deletion is available, include in quote otherwise
- #done this is not needed anymore as we more items into a inactive project if cancelled

11. Image upload: During digitisation the user will capture the barcoded docket page first and then the item. The user will upload the images using the mass upload workflow plugin. Goobi will then read the barcode and automatically link all the images including the barcode image to the correct process in Goobi.

- #done this does exist already
- #intranda take this into account for the pricing proposal if not done already (see above: change of UI to select between barcode and file naming)

12. Quality Assurance: after upload for all file types they will be reviewed in the QA plugin. this plugin must be developed to ensure that a different user than the one that did the digitisation reviews the output. #Development

- #intranda plan this development and take it into the pricing proposal

13. ~~Automatic file renaming: File rename plugin (the barcode docket image will be renamed with the number `_00`~~

- ~~done Geoff are you sure? Why not move it to the end instead of to the start? Or move it out of the images folder would even be better.~~
- ~~done Geoff-reply That is fine but I am just repeating what we did here when you first developed the functionality for me in the italy project.~~

14. Automatic removal of the barcode page: Goobi will delete the barcode image leaving only the master images of the item in the process.

- #done Geoff Why first rename the file automatically and then delete it afterwards?
- #done Geoff-reply As above, I thought it had to happen that way as at QA the user should see the barcode.

15. #Variant If cropping is required - Automatic image analysis.
16. #Variant If cropping is required - Layout Wizzard (Note: Layout Wizzard workflow plugin should also be installed*)
17. #Variant If cropping is required - Automatic cropping (to Cropped master folder retaining the master file)
18. #Variant If OCR is required - OCR step
19. #Variant If ICR is required - Transkribus step (using specific model ID if one was provided in step 9 above)

20. #Variant if pagination is required then the user carries out pagination in the METS editor.
21. #Variant if structural data is required then the user assigns structural data in the METS editor.
22. ~~Variant If surrogate JPEGs or PDF files are required then Goobi will create the surrogates in a folders~~

- ~~Geoff how do the expectations for the JPEGs and the PDFs look like exactly?~~
- ~~Geoff-reply I have asked that question of UOL~~
- ~~Leeds IIIF Services will create the surrogate files, Goobi does not need to create.~~

13. Goobi re-imports selected metadata fields from either the EMu API or the ALMA API into the METS file #Development

- #intranda take this into account for the pricing proposal if not done already

14. Final QA: carried out by a different user than the initial QA person and using the METS editor tool.

- #intranda Same as in Step 11, needs to be developed and added to pricing proposal

15. Goobi creates the checksums for the digital files to be deposited and records that in the METS files Note, this is to be confirmed. #Development

- #done Geoff we are waiting for a confirmation then
- #done Geoff-reply Yes, as above
- #intranda plan this (and the ingest) development as mentioned above as pool of hours

16. Goobi creates a digital object in the repository #Development

- #intranda take this into account for the pricing proposal if not done already

17. Goobi deposits the digital object in the repository #Development

- #intranda take this into account for the pricing proposal if not done already

18. Goobi receives back a message/confirmation that the deposit was successful TBC #Development

- #done Geoff we are waiting for a confirmation then
- #done Geoff-reply Yes, as above
- #intranda add a delay plugin here and then do a request to the digirati system if the ingest was successful
- #intranda add a delay a second time before the actual deletion is started

1. Goobi deletes the master images from the process and retains thumbnails for future reference. #Development

- #intranda take this into account for the pricing proposal if not done already

2. ~~Goobi closes the process and moves the deposited process to a new project entitled Deposited Processes Development~~

- ~~done Geoff Why not leave the items in the projects?~~

- ~~done However this functionality does exist already~~
- ~~done Geoff-reply Already replied to this above.~~

3. ~~Goobi changes the process template to a new process template called `Editing_Deposited_Processes` with the step: `Select editing action` set to OPEN~~

- ~~Note: the intention here is that the unique identifier passed to Goobi in step 27 above will enable Goobi to pull back the master media from the repository ready for edits to take place.~~
- ~~Geoff I would use the `Workflow Extender Plugin` to be developed first for such use cases.~~
- ~~Geoff-reply Yes, as above, agreed.~~

Large_Scale_Digitisation_Workflow_Outsourced

This workflow is for large scale digitisation projects where the work will be carried out by third party providers. It is envisaged that this will likely be for image based digitisation in the medium term but a #Variant could be created for moving image or audio projects in the future, based on this process template.

Pre-requisite conditions:

- All items to be digitised are catalogued to item level on EMu or ALMA.
- Each item is entered in the mass import interface to be important via EMu IRN or the ALMA BIBNET reference.
    - Dockets are printed out then and put into the books
- A project has been set up in Goobi with a total number of processes, estimated number of pages, a start date, and a deadline for completion.
- The supplier has received the materials with the barcodes and a detailed specification.
- The supplier scans always the barcode-docket page before each book. Naming of the files is not important anymore then
- The images are uploaded into a given S3-bucket into subfolders
- The mass upload plugin reads from the S3-bucket (ignores the naming of files and folders), checks the barcodes and puzzles the images to the correct Goobi processes
    - #intranda extent the mass upload to read from a S3-bucket

DRAFT Process template:

1. User will inserts identifiers from EMu or ALMA.

- #intranda take this into account

2. Image upload through mass upload and barcode scan

- #intranda plan this

3. #done Leeds Checking of access rights/licensing/copyright and sensitivity warnings and adding/updating METS.
4. Quality Assurance: after upload for all file types they will be reviewed in the QA plugin.

- #done this does exist already

5. #Variant If cropping is required - Automatic image analysis.
6. #Variant If cropping is required - Layout Wizzard (Note: Layout Wizzard workflow plugin should also be installed)
7. #Variant If cropping is required - Automatic cropping (to Cropped master folder retaining the master file)
8. #Variant If OCR is required - OCR step
9. #Variant If ICR is required - Transkribus step (using specific model ID if one was provided in step 9 above)
10. #Variant if pagination is required then the user carries out pagination in the METS editor.
11. #Variant if structural data is required then the user assigns structural data in the METS editor.
12. Goobi re-imports selected metadata fields from either the EMu API or the ALMA API into the METS file #Development

- #intranda take this into account for the pricing proposal if not done already

14. Final QA: carried out by a different user than the image QA person.

- #intranda Same as in other workflow, needs to be developed and added to pricing proposal

15. Goobi creates the checksums for the digital files to be deposited and records that in the METS files Note, this is to be confirmed. #Development

- #done Geoff we are waiting for a confirmation then
- #done Geoff-reply As above

16. Goobi creates a digital object in the repository #Development

- #intranda take this into account for the pricing proposal if not done already

17. Goobi deposits the digital object in the repository #Development

- #intranda take this into account for the pricing proposal if not done already

18. Goobi receives back a message/confirmation that the deposit was successful TBC #Development

- #done Geoff we are waiting for a confirmation then
- #done Geoff-reply As above

19. Goobi deletes the master images from the process and retains thumbnails for future reference. #Development

- #intranda take this into account for the pricing proposal if not done already

Next steps

- This document will be circulated to #intranda and #Leeds for comments, questions and clarifications.
- Based on that feedback #intranda will ask further questions in order to clarify the services and developments needed and assign costs to them. This may necessitate clarification meetings with Digitisation.IO, UOL and Digirati as an iterative process.

End of Document