Integrating Digitization Workflow with the Stanford Digital Repository

Hannah Frost and Peter Mangiafico from Digital Library Systems and Services

Stanford | LIBRARIES

INTRODUCTION

Stanford Libraries has a mature, established digitization program for collection materials of all kinds. After years of running a home-grown automated workflow for creating and managing image files through the production process, we replaced it with an open-source task management application called Goobi and integrated it locally with the Stanford Digital Repository to create an end-to-end workflow to streamline task management, production metrics gathering, and the hand-off of digitized content to the repository for preservation and access.
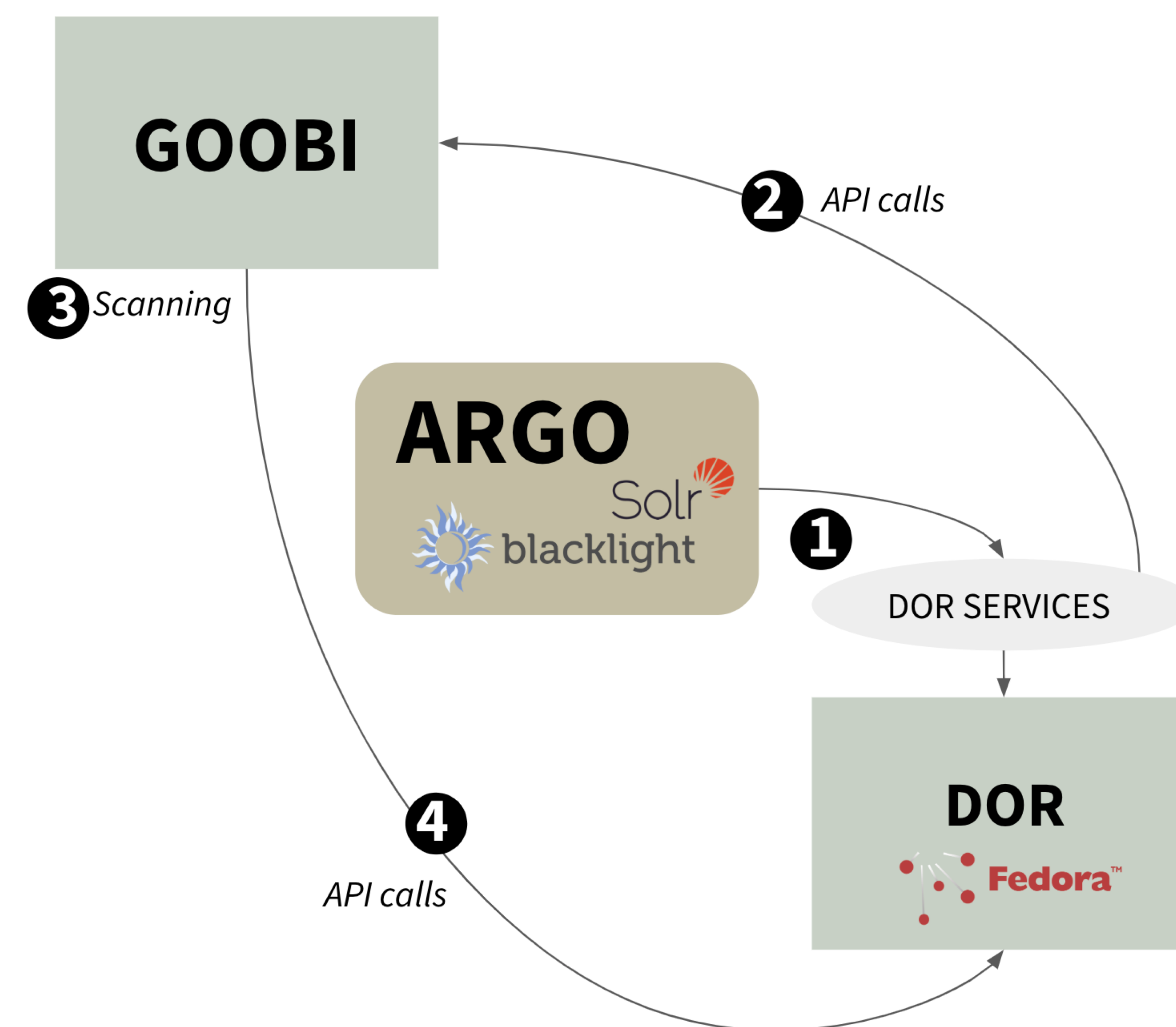
WHY GOOBI?

• Open source; adopted by many libraries around the world
• Using a vendor-supported product reduces the need for local resources to maintain and support one-off code base
• Features:
   • Configurable roles, tasks and workflows at the project level
   • Easy to distribute tasks across teams and locations, including service providers
   • Integratable with other systems, such as OCR processing, ILS
   • Plug-ins for a variety of tasks, such as layout manipulation, image optimization, quality control
   • Project dashboards and statistics

WHAT IS A TYPICAL WORKFLOW?

• Prepare materials to be digitized into digital objects
• Register digital objects to be created during scanning in Argo (SDR admin system), entering title, project tags, and assigning the Goobi workflow and process template (outline of sequenced production tasks for this content)
• Log in to Goobi to find production tasks to be completed in sequence: scanning, cropping, rotation, QA, OCR-processing, final QA, and then "export"

Alternatively, it is possible to complete scanning of registered objects outside of Goobi, and simply drop structured folders of files into a Goobi-aware "home folder" to kick off OCR processing and then export to SDR.

[DIAGRAM]

GOOBI

② API calls

③ Scanning

ARGO
Solr
blacklight

①

DOR SERVICES

④ API calls

DOR
Fedora™

HOW IS IT INTEGRATED?

1 When an object is registered in Argo with the Goobi workflow assigned, a robot in the DOR system developed by the SDR team sends object info in XML to our Goobi system via an API call.

2 Intranda developed an API in our local Goobi instance to receive and process that XML to generate "tasks" in Goobi as specified in the process template assigned to the object at registration.

3 Intranda also customized our Goobi instance to structure files generated through the production process according to our SDR spec.

4 Upon export from Goobi, a package of image files and some relevant metadata is copied to a DOR file system workspace and then queued for assembly and accessioning to the SDR.

WHAT HAVE WE LEARNED?

• Production coordination is made easier by task management features

• Goobi is optimized for book workflows

• Does not handle large images (posters, maps, artwork) well

• Not suited for media workflows

• Production statistics are not sufficiently granular, and stats can be misleading if task completion is delayed

• Our team doesn't have German language skills, unable to understand some terms used in code and database

• Some user documentation is only available in German

• Support by long-distance vendor leads to communication and issue resolution delays

• SDR support for Goobi code is difficult given team's limited Java skills

• No other Goobi adopters in North America

ABOUT GOOBI

Goobi
intranda edition

Goobi is developed by intranda GmbH, a company based in Göttingen, Germany. Intranda has been offering digitization services and producing software for the digitization sector since 2003.

October 2019