The Goobi workflow system, introduction and demos

Steffen Hankiewicz, intranda GmbH
London, 10.11.2014
1. What are we doing?
1. What are we doing?

We are content providers
2. We are content providers ...

Imagine:

- 10,000 pages

Precondition:

- digitized images on the web online

Search for: Illustration
2. We are content providers ...

Imagine:
- 10.000 pages

Precondition:
- digitized images on the web online

Search for: Illustration
- ... in the full text
2. We are content providers ...

Imagine:
- 10,000 pages

Precondition:
- digitized images on the web online

Search for: Illustration
- ... in the full text
- ... on page CCXVI
2. We are content providers ...

Imagine:

- 10,000 pages

Precondition:

- digitized images on the web online

Search for: Illustration

- ... in the full text
- ... on page CCXVI
- ... as structure element
2. We are content providers ...

Imagine:
- 10,000 pages

Precondition:
- Digitized images on the web online

Search for: Illustration
- ... in the full text
- ... on page CCXVI
- ... as structure element
- ... as word in the title of a chapter
2. We are content providers ...

Imagine:
- 10,000 pages

Precondition:
- Digitized images on the web online

Search for: Illustration
- ... in the full text
- ... on page CCXVI
- ... as structure element
- ... as word in the title of a chapter
- ... as synonym for 'drawing'
3. We need some workflow for each item

1. Source material
2. Create digital version
3. Transformation & Enrichment
4. Publish digital version(s)
3. We need some workflow for each item.

1. Source material
2. Create digital version
3. Transformation & Enrichment
4. Publish digital version(s)

- Image conversion
- Image validation
- OCR
- ALTO generation
- Descriptive metadata
- Technical metadata
- Logical structures
- NER
- Authority data
- Persistent Identifiers
- Pagination
- Catalogue enrichment
- Ingest into archive
- Invoicing

Image conversion

Image validation

OCR

ALTO generation

Descriptive metadata

Technical metadata

Logical structures
4. Goobi - a quick overview

... try to solve common problems

- Web application
- Workflow tool
- Manage users
- Organize projects
- Deadlines
- Data storage
- Metadata formats
5. Goobi - how it works ...

... a simple approach

- Workflows cut into small pieces
- Simple sequential order of tasks
- As much validation as early as possible
- Restrict access to the requirements
- Hide everything else from the user
6. Goobi - the users perspective

- Simple UI
- Work with To-Do-List
- Hidden complexities:
  - Storage
  - Projects
  - Infrastructure
- Clean desk

... avoid difficulties
6. Goobi - the users perspective
6. Goobi - the users perspective

Steffen

Goobi web interface

working directory of Steffen
6. Goobi - the users perspective

Steffen

working directory of Steffen

Goobi web interface
6. Goobi - the users perspective

- Server side programs
- Plugins without user interface
- Plugins with user interface
7. Goobi - Management overview

... manage your workflows

- Manage all typical configuration in the UI
  - Workflows
  - Projects
  - Users
  - User groups
  - Imports
  - Exports
7. Goobi - Management overview

- Controlling and statistics
- Manipulate workflows afterwards (e.g. with GoobiScript)
- Collaborate with external partners or agencies

... control your progress
8. Goobi - technical background

... what else can be done?

- Workflows can be ...
  - simple or complex
  - short or long
  - contain tasks
  - have a progress
  - used as template

- Import from catalogue
- Scanning
- Quality control
- Image conversion
- OCR
- Structure- & metadata
- ID-Generating
- Presentation
- Archiving
8. Goobi - technical background

... what else can be done?

- Workflow steps can...
  - be executed manually by a user
  - be executed automatically by the server
  - interrupt the workflow for a given time
  - contain a validation
  - allow or forbid access or changes
  - be triggered by a web-API
  - call scripts or external programs
  - have their own UI as plugins
9. Goobi - Extend its functionality
9. Goobi - Extend its functionality
9. Goobi - Extend its functionality

Goobi

Import plugins

OAI

Word

PICA

MARC

Step plugins

QA

Ingest

JP2

Export

Validation plugins

Web API command plugins

Close step

Run script

Create process

Delete files

JP2

MD5

Schema

Color depth
10. Goobi - Scripts & applications

- OCR
- JPEG
- JPEG 2000
- Jpylyzer
- Archiving
- Download-Jobs
- Exporters
- Named Entity Recognition
11. Goobi - production proven

- Lots of institutions
- Different kinds of material
- Community driven
- Open Source
- Active development
- Inhouse or hosted
- Scalability

A lot of happy content providers
Questions?

intranda GmbH - Steffen Hankiewicz

- http://www.intranda.com
- info@intranda.com
- +49 551 29176100