



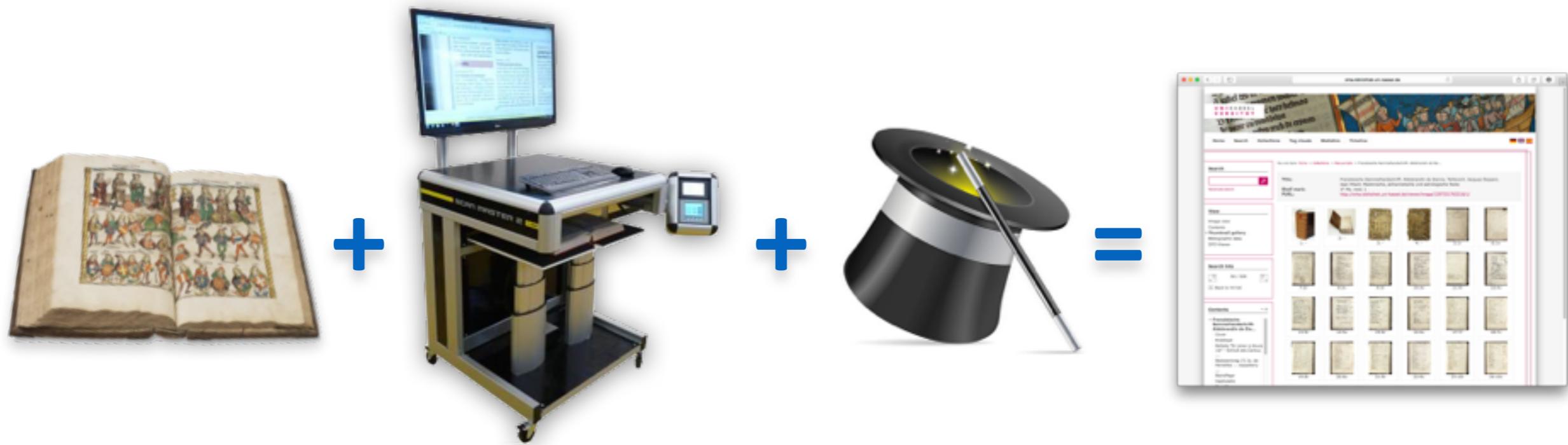
The Goobi workflow system, introduction and demos

Steffen Hankiewicz, intranda GmbH

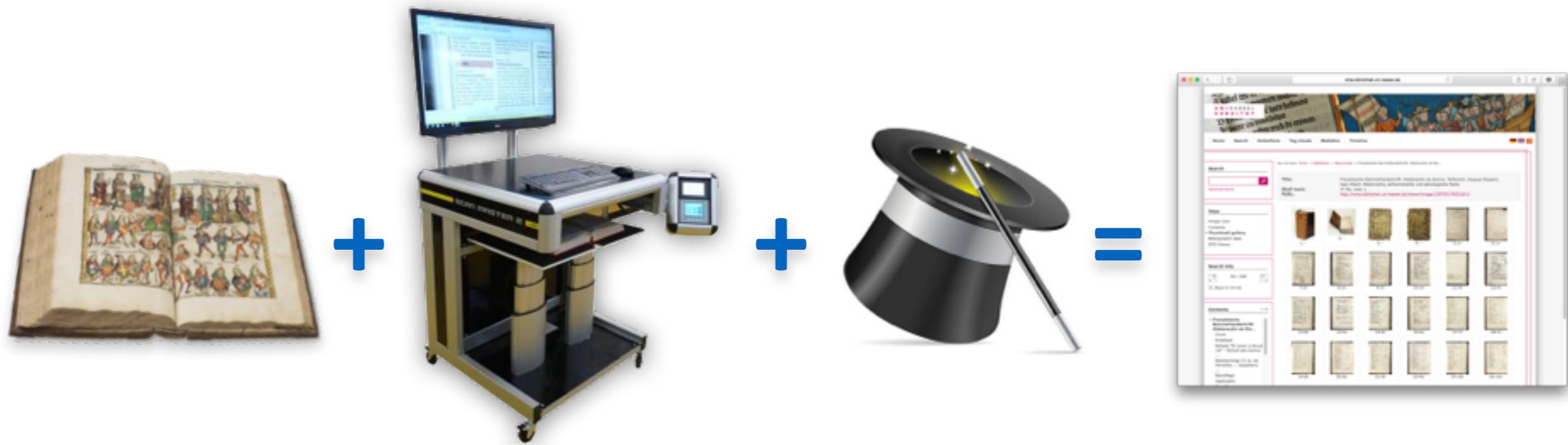
London, 10.11.2014



1. What are we doing?



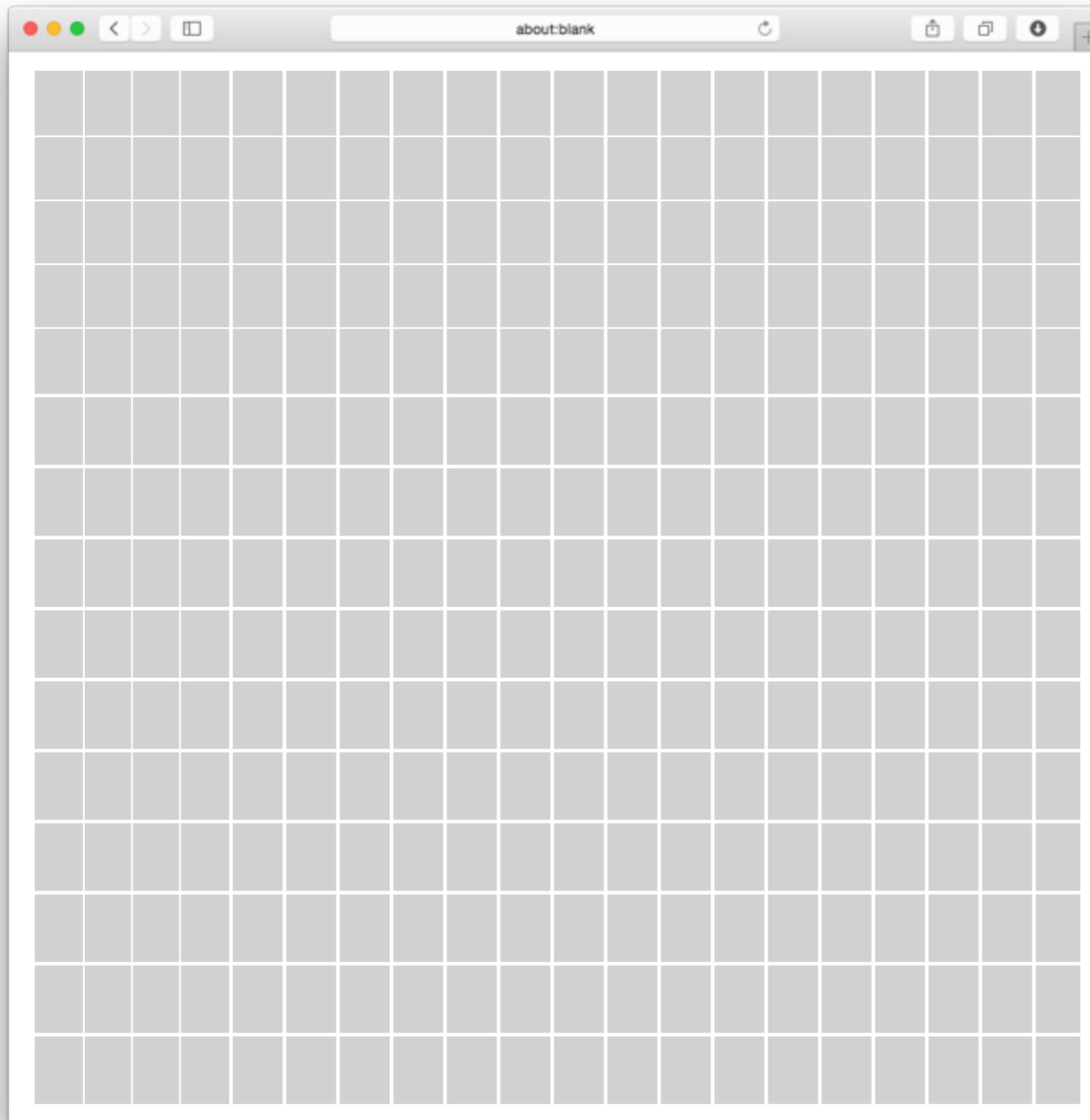
1. What are we doing?



We are content providers 



2. We are content providers ...



Imagine:

- ▶ 10.000 pages

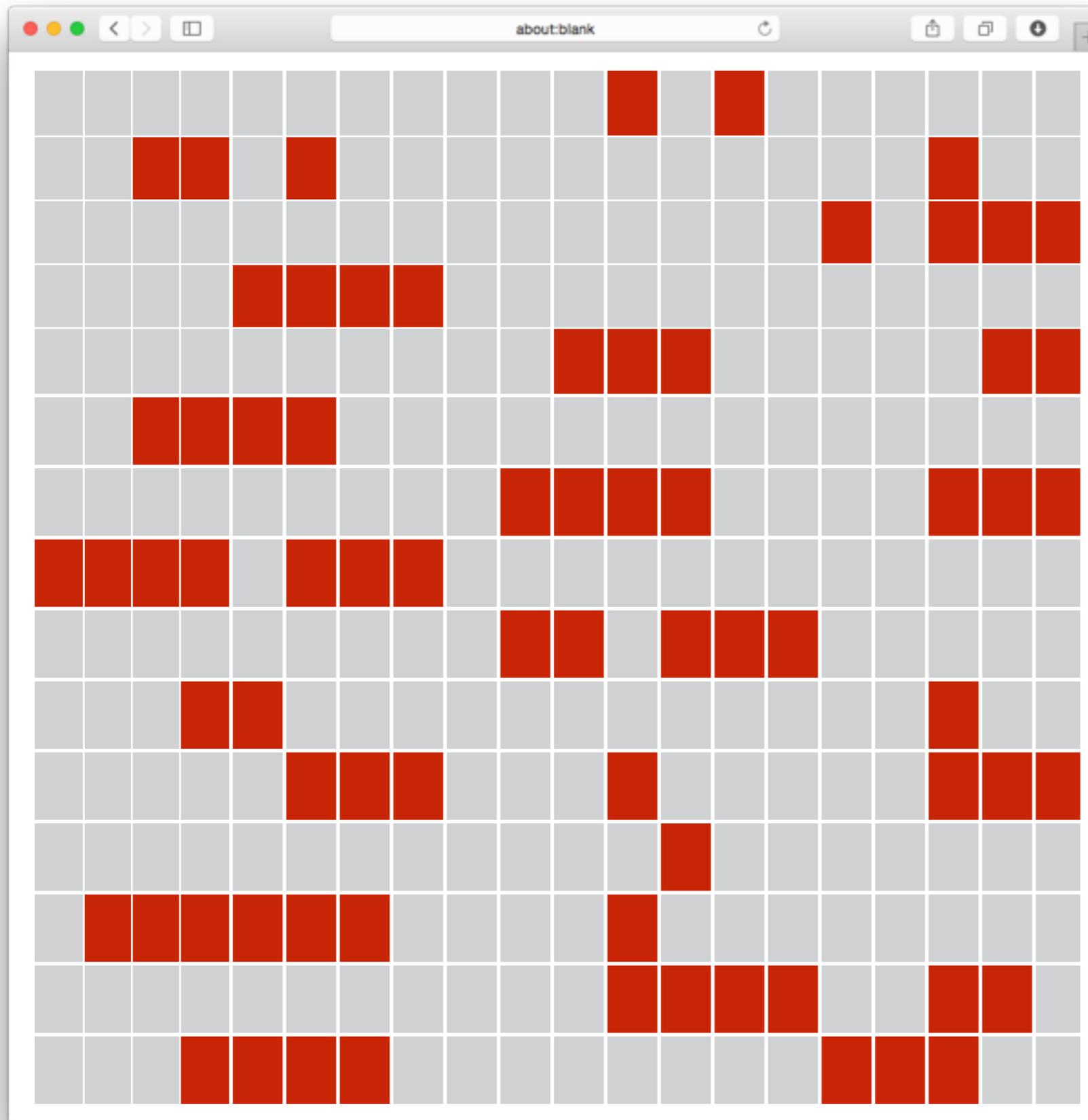
Precondition:

- ▶ digitized images on the web online

Search for: Illustration



2. We are content providers ...



Imagine:

- ▶ 10.000 pages

Precondition:

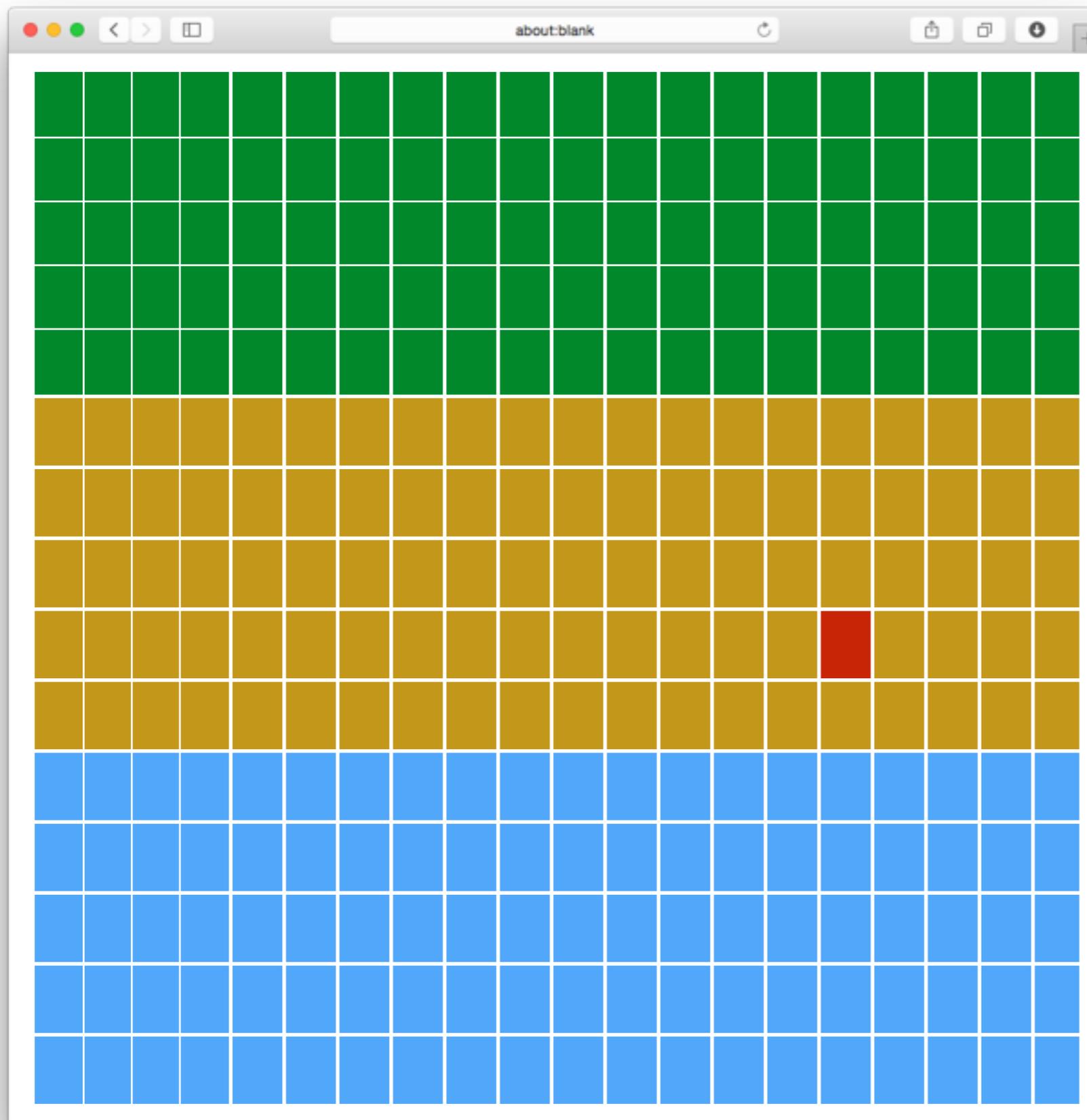
- ▶ digitized images on the web online

Search for: Illustration

- ▶ ... in the full text



2. We are content providers ...



Imagine:

- ▶ 10.000 pages

Precondition:

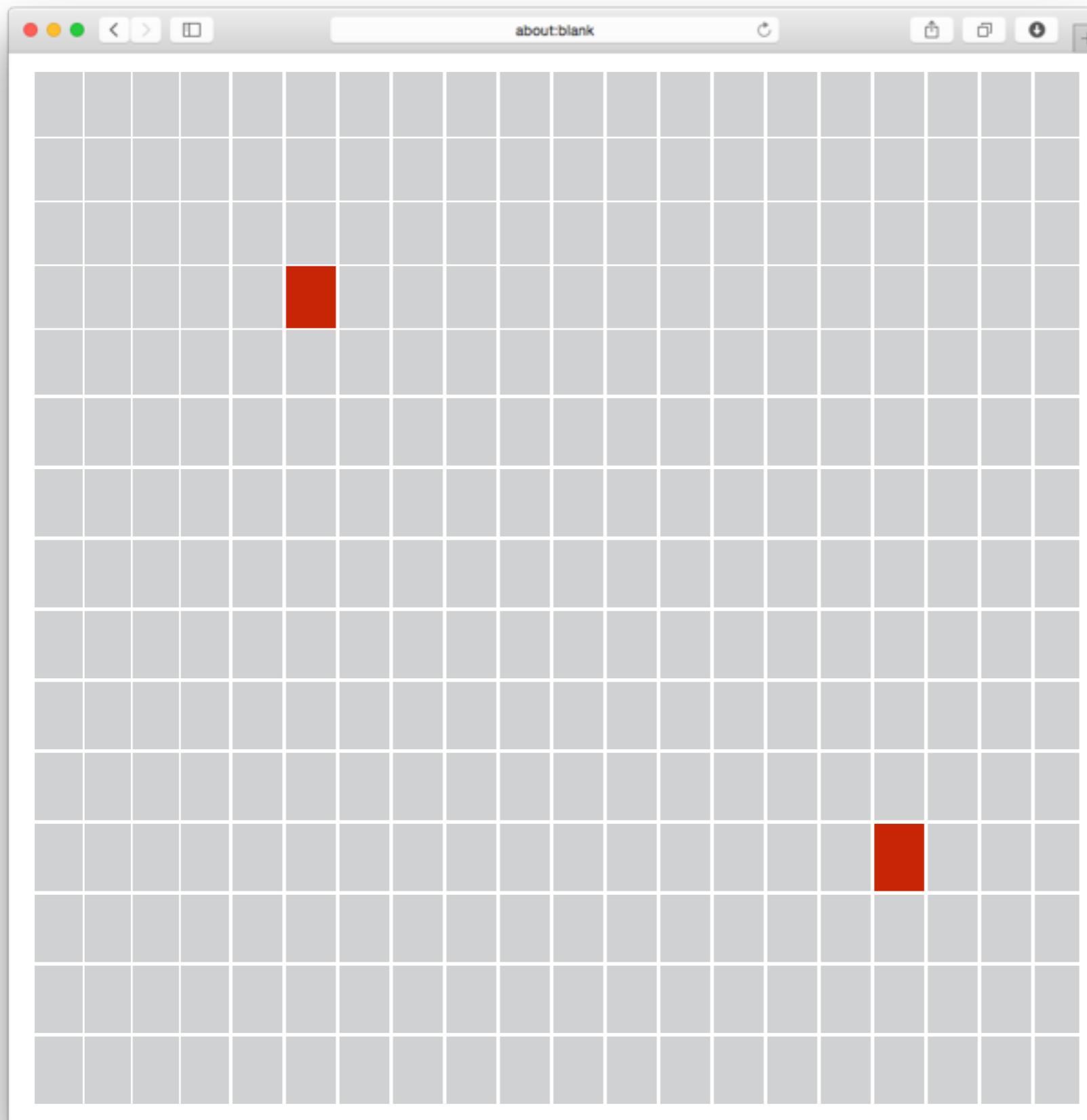
- ▶ digitized images on the web online

Search for: Illustration

- ▶ ... in the full text
- ▶ ... on page **CCXVI**



2. We are content providers ...



Imagine:

- ▶ 10.000 pages

Precondition:

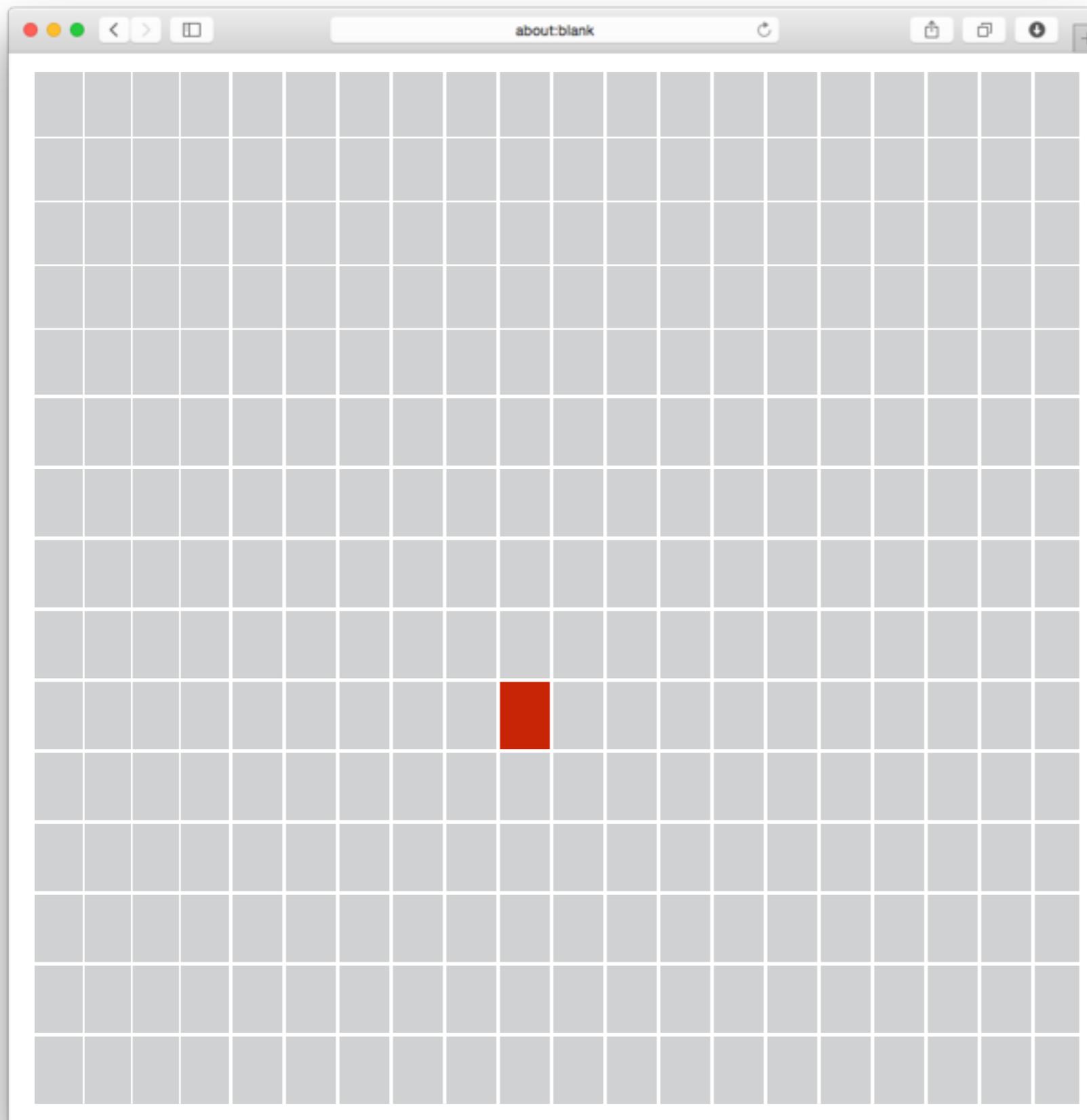
- ▶ digitized images on the web online

Search for: Illustration

- ▶ ... in the full text
- ▶ ... on page CCXVI
- ▶ ... **as structure element**



2. We are content providers ...



Imagine:

- ▶ 10.000 pages

Precondition:

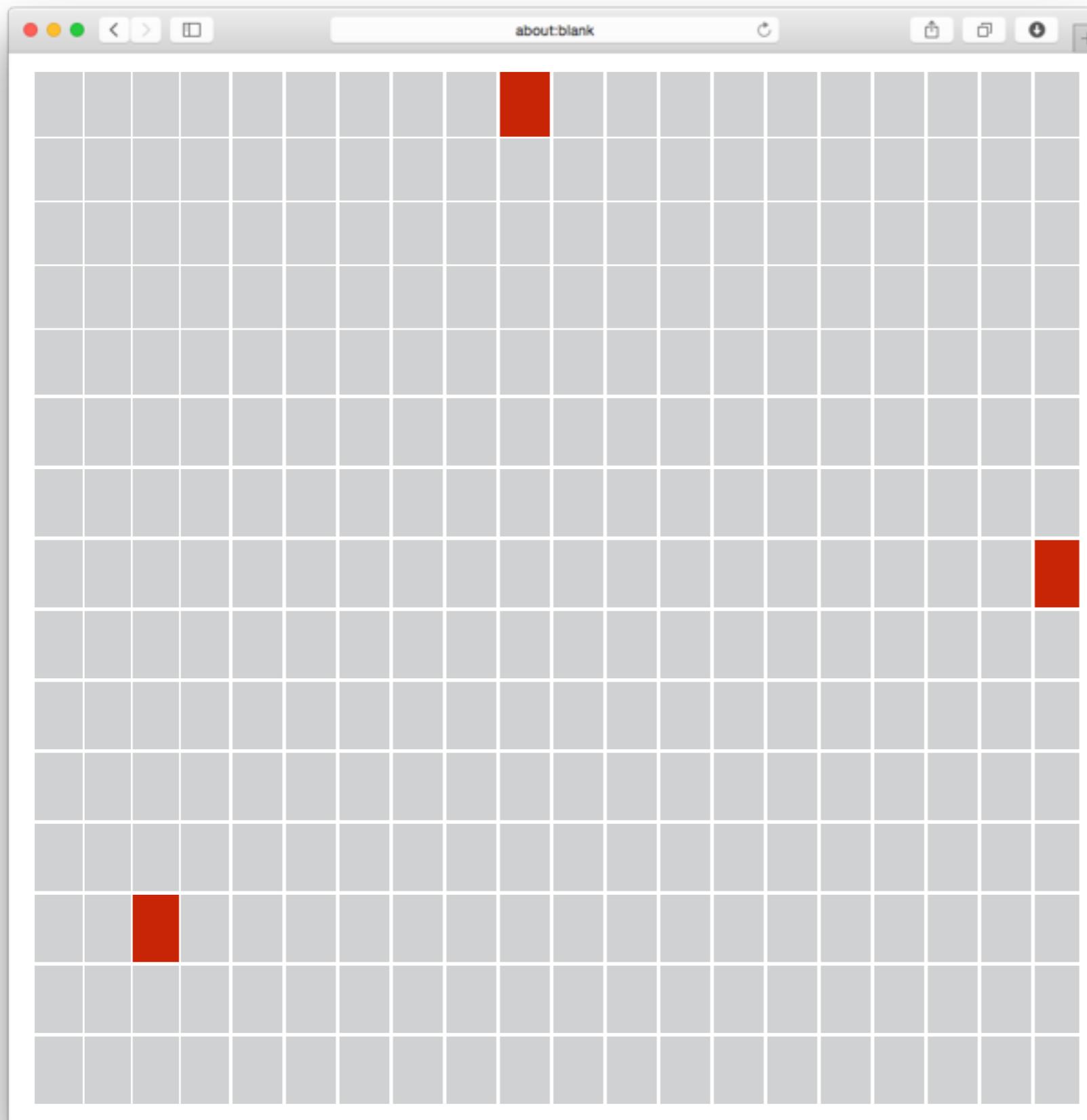
- ▶ digitized images on the web online

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 

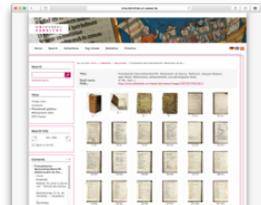


1. Source material



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



4. Goobi - a quick overview

... try to solve common problems

10.11.2014

- ▶ Web application
- ▶ Workflow tool
- ▶ Manage users
- ▶ Organize projects
- ▶ Deadlines
- ▶ Data storage
- ▶ Metadata formats

The screenshot displays the Goobi web application interface. The browser address bar shows the URL `http://goobi.intranda.com`. The page title is "Goobi - intranda edition". The main content area is titled "Process management" and shows a list of processes. The table below is a representation of the data shown in the screenshot:

Process title	Status	Project	Action
alphrudeh_PPN591416573		Sample_Project	XML, METS, PDF, DMS
alsdehou_PPN560091583		Sample_Project	XML, METS, PDF, DMS
AustPrin_000145107		Sample_Project	XML, METS, PDF, DMS
bergsphi_625017145		Sample_Project	XML, METS, PDF, DMS
bianjani_618686231		Sample_Project	XML, METS, PDF, DMS
biannoti_PPN523090374		Sample_Project	XML, METS, PDF, DMS
bojaanat_620650583		Sample_Project	XML, METS, PDF, DMS
daihelv_578780275_0001		Sample_Project	XML, METS, PDF, DMS
daihelv_PPN578780275_0001		Sample_Project	XML, METS, PDF, DMS
dissinsis_626789478		Sample_Project	XML, METS, PDF, DMS

Below the table, there is a "Possible actions" section with the following options:

- Export metadata for DMS
- Set status of process up
- Set status of process down
- Execute GoobiScript
- Generate Excel file
- Create PDF
- Calculate number of metadata and images
- Statistical evaluation

5. Goobi - how it works ...

... a simple approach

10.11.2014

- ▶ 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

The screenshot displays the Goobi intranda edition web interface. The browser address bar shows the URL: `http://goobi.intranda.com`. The page title is "Goobi - intranda edition". The main content area is titled "Process management" and shows a list of processes. A table lists the processes with columns for "Process title", "Status", "Project", and "Action". The "Action" column contains icons for XML, METS, PDF, and DHS. A detailed view of a process is shown, listing tasks such as "Bibliographic import", "Scanning", "Quality check", "Create working images and write tiff header", "Image processing", "Structure- and metadata", "Export to presentation system", and "Archive". The status of each task is indicated by a colored bar (green for completed, yellow for in progress, red for failed). The page also includes a sidebar with navigation links and a footer with performance metrics.

Process title	Status	Project	Action																											
alphrudeh_PPN591416573		Sample_Project																												
alsdehou_PPN560091583		Sample_Project																												
AustPrin_000145107		Sample_Project																												
bergsphi_625017145		Sample_Project																												
bianjani_618685231		Sample_Project																												
biannoti_PPN523090374																														
<table border="1"><thead><tr><th>Number</th><th>Title</th><th>Status</th></tr></thead><tbody><tr><td>1</td><td>Bibliographic import</td><td></td></tr><tr><td>2</td><td>Scanning</td><td></td></tr><tr><td>3</td><td>Quality check</td><td></td></tr><tr><td>4</td><td>Create working images and write tiff header</td><td></td></tr><tr><td>5</td><td>Image processing</td><td></td></tr><tr><td>6</td><td>Structure- and metadata</td><td></td></tr><tr><td>7</td><td>Export to presentation system</td><td></td></tr><tr><td>8</td><td>Archive</td><td></td></tr></tbody></table>				Number	Title	Status	1	Bibliographic import		2	Scanning		3	Quality check		4	Create working images and write tiff header		5	Image processing		6	Structure- and metadata		7	Export to presentation system		8	Archive	
Number	Title	Status																												
1	Bibliographic import																													
2	Scanning																													
3	Quality check																													
4	Create working images and write tiff header																													
5	Image processing																													
6	Structure- and metadata																													
7	Export to presentation system																													
8	Archive																													
bojaanat_620650583		Sample_Project																												
daihelv_578780275_0001		Sample_Project																												
daihelv_PPN578780275_0001		Sample_Project																												
dissinsis_626789478		Sample_Project																												

6. Goobi - the users perspective

... avoid difficulties

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

The screenshot shows the Goobi intranda edition web interface. The browser address bar displays 'http://goobi.intranda.com'. The page title is 'Goobi - intranda edition'. The main content area is titled 'My tasks' and contains a table of tasks. The table has columns for Task, Process, Project, Status, and Action. The tasks listed are all 'Scanning' tasks for 'Sample_Project' with various process IDs. The status of all tasks is 'Yellow'. The page also includes a sidebar with navigation links, a search bar, and a footer with performance metrics.

Task	Process	Project	Status	Action
Scanning	fisclett_609336819	Sample_Project	Yellow	Settings
Scanning	forsindi_600894029	Sample_Project	Yellow	Settings
Scanning	goezerst_578224070	Sample_Project	Yellow	Settings
Scanning	greifswalderWochenblatt	Sample_Project	Yellow	Settings
Scanning	herradfr_PPN612900487	Sample_Project	Yellow	Settings
Scanning	hupsnatu_625019202_0001	Sample_Project	Yellow	Settings
Scanning	kleiuniv_517154005	Sample_Project	Yellow	Settings
Scanning	kreisanzeiger1877	Sample_Project	Yellow	Settings
Scanning	kreisanzeiger1882	Sample_Project	Yellow	Settings
Scanning	laspsesi_605023042	Sample_Project	Yellow	Settings

Page 1 of 3

Possible actions:

- Customize filter
- Customise view

6. Goobi - the users perspective

Steffen



Goobi web interface

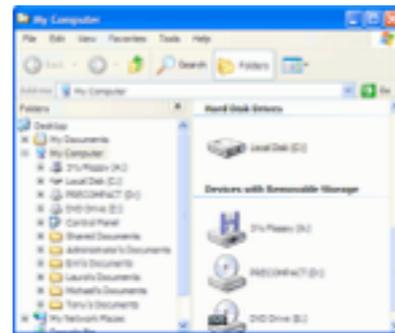


6. Goobi - the users perspective

Steffen



Goobi web interface



working directory
of Steffen

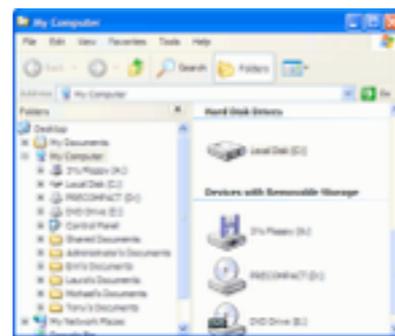


6. Goobi - the users perspective

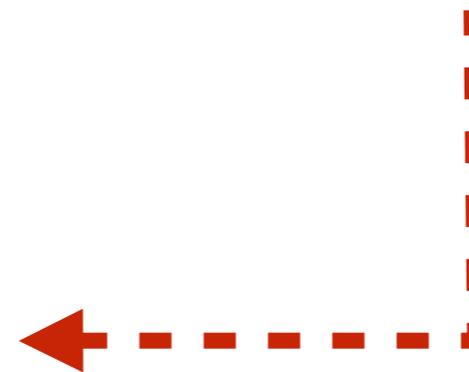
Steffen



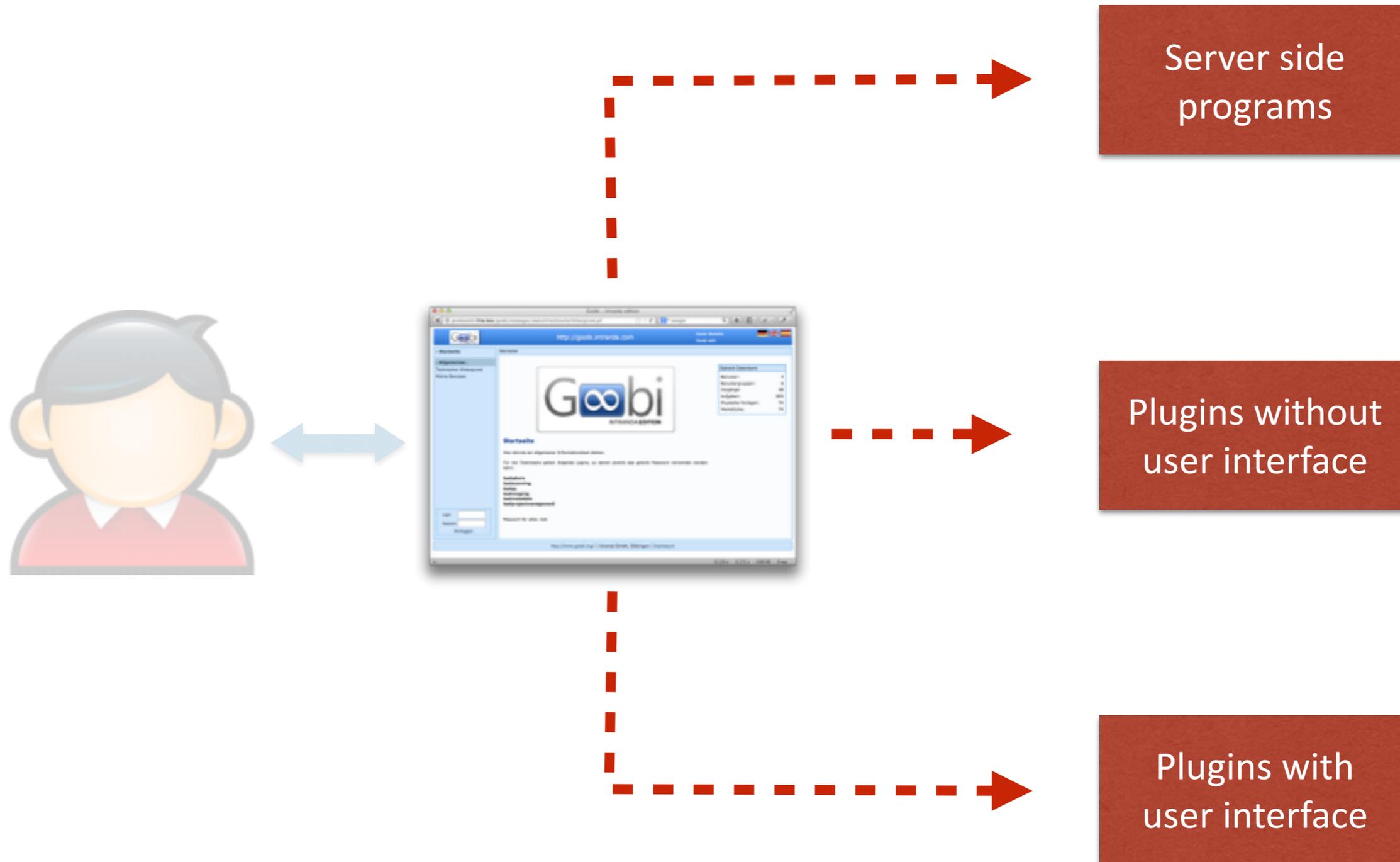
Goobi web interface



working directory
of Steffen



6. Goobi - the users perspective



7. Goobi - Management overview

... manage your workflows

- ▶ Manage all typical configuration in the UI
 - ▶ Workflows
 - ▶ Projects
 - ▶ Users
 - ▶ User groups
 - ▶ Imports
 - ▶ Exports

The screenshot displays the Goobi intranda edition web interface. The browser address bar shows the URL: `http://goobi.intranda.com`. The page title is "Goobi - intranda edition". The main content area is titled "Process details" and shows the following information:

- Process title: `biannoti_PPN523090374`
- Project: `Sample_Project`
- Date created: `25-Aug-2010`
- Rule set: `standard`
- Show in selection list:
- Is template:
- ID: `16`
- Batch:

Below the process details, there is a "List of tasks" table:

Number	Title	Status	Action
1	Bibliographic import	Green	
2	Scanning	Green	
3	Quality check	Green	
4	Create working images and write tiff header	Yellow	
5	Image processing	Red	
6	Structure- and metadata	Red	
7	Export to presentation system	Red	
8	Archive	Red	

Below the task list, there is a "Physical templates" table:

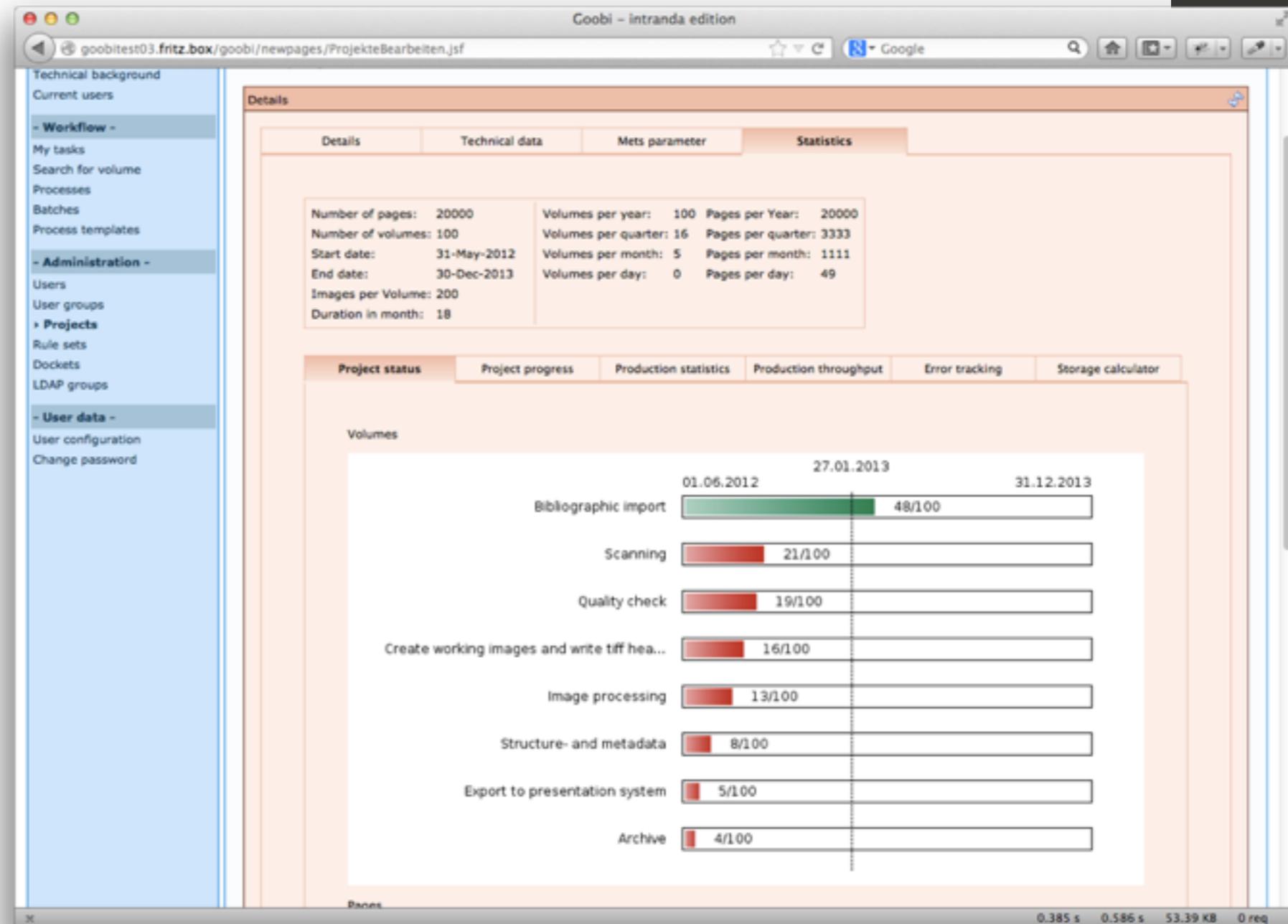
Number	Properties	Action
	Authors: <code>Bianchi, Thomas X.</code>	
	Identifier analog: <code>...</code>	

The interface also includes a sidebar with navigation options: Home page, General (Technical background, Current users), Workflow (My tasks, Search for volume, Processes, Batches, Process templates), Administration (Users, User groups, Projects, Rule sets, Dockets, LDAP groups), and User data (User configuration, Change password). The footer shows performance metrics: 0.180 s, 0.721 s, 186.80 KB, 0 req.

7. Goobi - Management overview

... control your progress

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



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

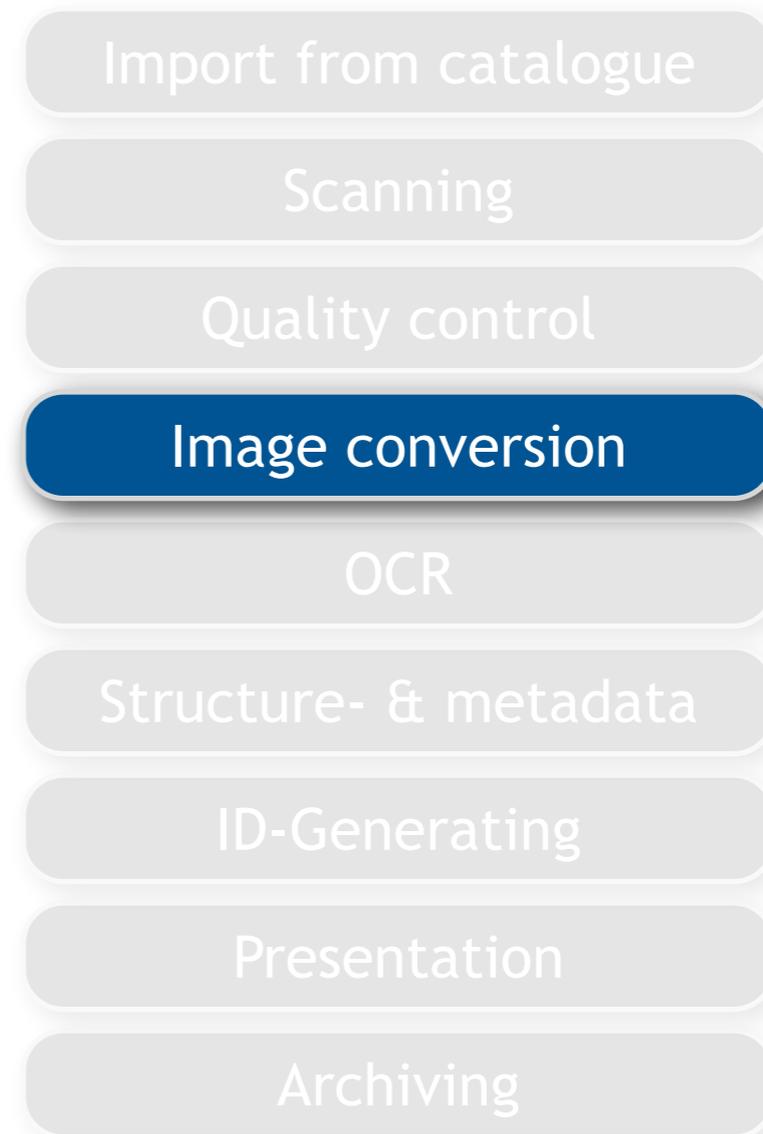
Archiving



8. Goobi - technical background

... what else can be done?

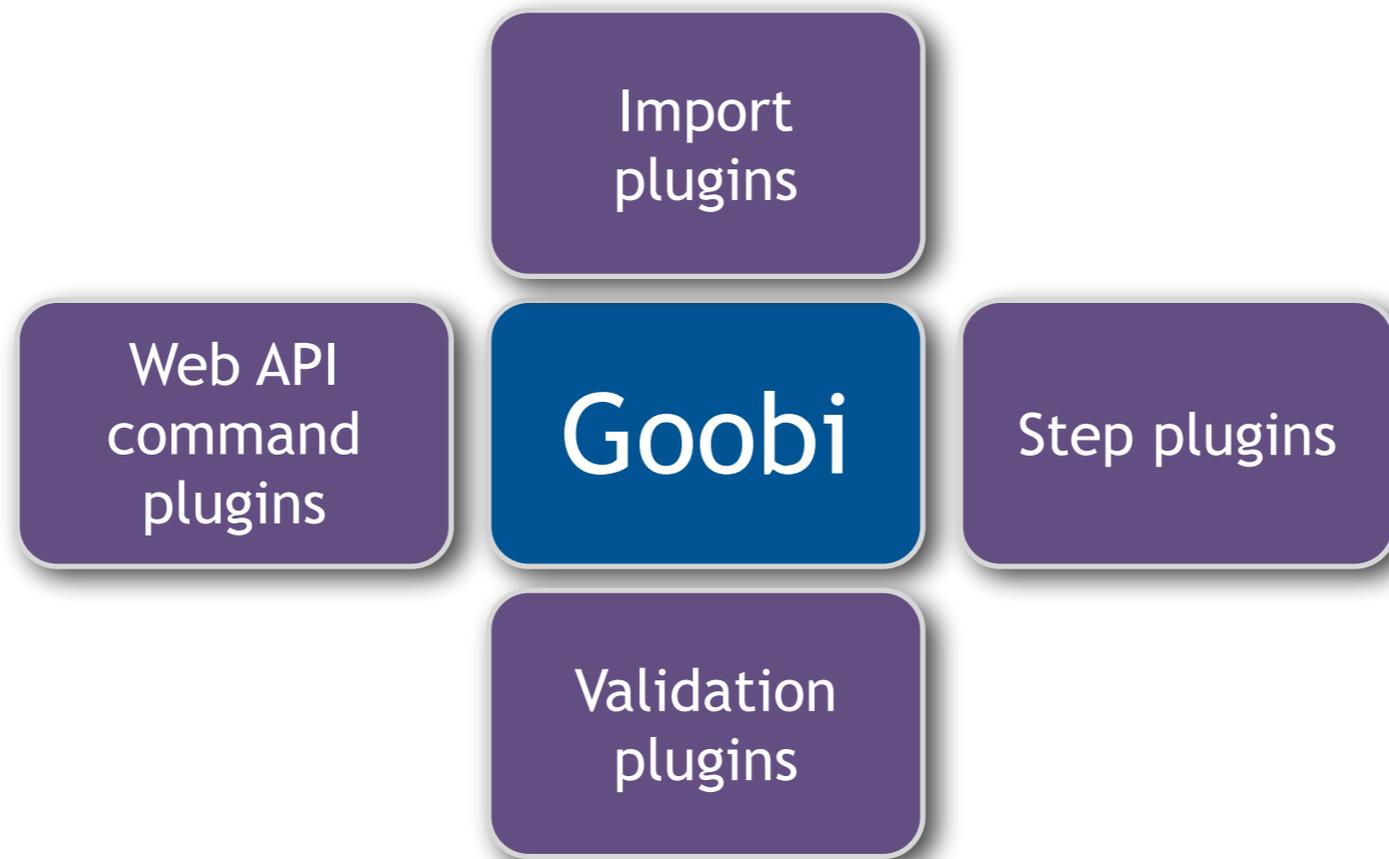
- ▶ Workflow steps can ...
 - ▶ be executed manually by a user
 - ▶ be executed automatically by the server
 - ▶ interrupt the workflow for a given time
 - ▶ contain a validation
 - ▶ allow or forbid access or changes
 - ▶ be triggered by a web-API
 - ▶ call scripts or external programs
 - ▶ have their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



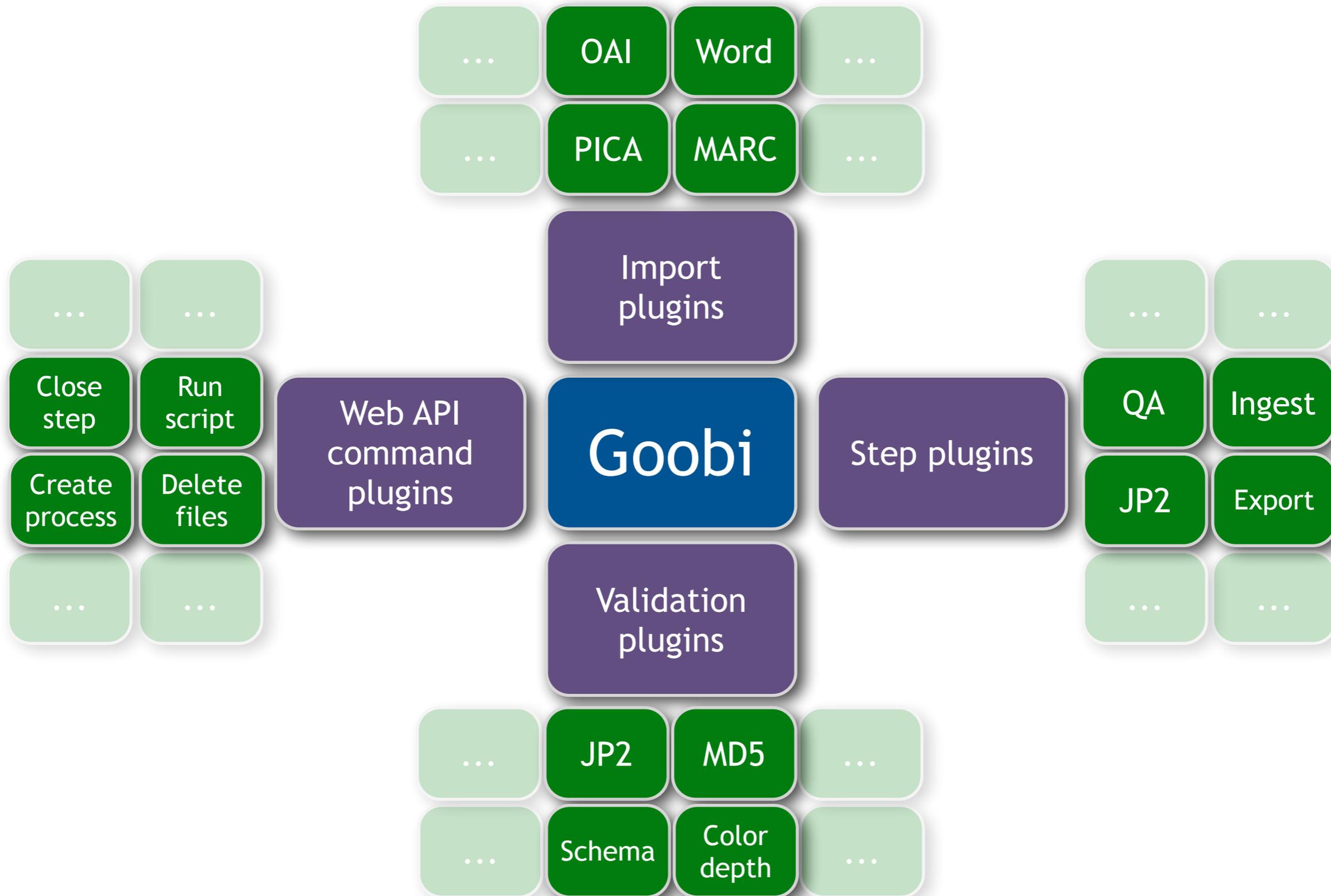
9. Goobi - Extend its functionality



9. Goobi - Extend its functionality



9. Goobi - Extend its functionality



10. Goobi - Scripts & applications

- ▶ OCR
- ▶ JPEG
- ▶ JPEG 2000
- ▶ Jpylyzer
- ▶ Archiving
- ▶ Download-Jobs
- ▶ Exporters
- ▶ Named Entity Recognition

The screenshot displays the intranda TaskManager web interface. The browser address bar shows 'goobitest02.fritz.box/itm/Index.xhtml'. The page title is 'intranda TaskManager' with navigation links for 'Dashboard', 'Search', 'OCR', and 'IA-Download'. The language is set to 'Deutsch English'.

Dashboard

Active jobs: 5 Jobs, 1470 Image files, 5460 MB

ID	Job type	Status	Goobi process title	Priority	Size	Image files	Last change	Action
39	OCR	<input type="checkbox"/>	alprudeh	10	1,736 MB	434	2014-03-24 12:41:32	Details
40	OCR	<input type="checkbox"/>	bbergisphi	10	1,532 MB	383	2014-03-24 11:59:54	Details
41	OCR	<input type="checkbox"/>	manualtofvete	10	932 MB	233	2014-03-24 10:28:21	Details
42	IADOWNLOAD	<input type="checkbox"/>	bianotti	10	588 MB	196	2014-03-24 13:51:48	Details
43	IADOWNLOAD	<input type="checkbox"/>	dissinsis	10	672 MB	224	2014-03-24 14:28:25	Details

Queues

Job	Queue	Plugin status
IA-Download	1	Running
OCR	1	Running

Throughput pages per month

Month	Throughput (pages)
October	0
November	1000
December	400
January	1900
February	1400
March	1100

Licences

Contract	Available
Vertrag 1	84
tesseract	0

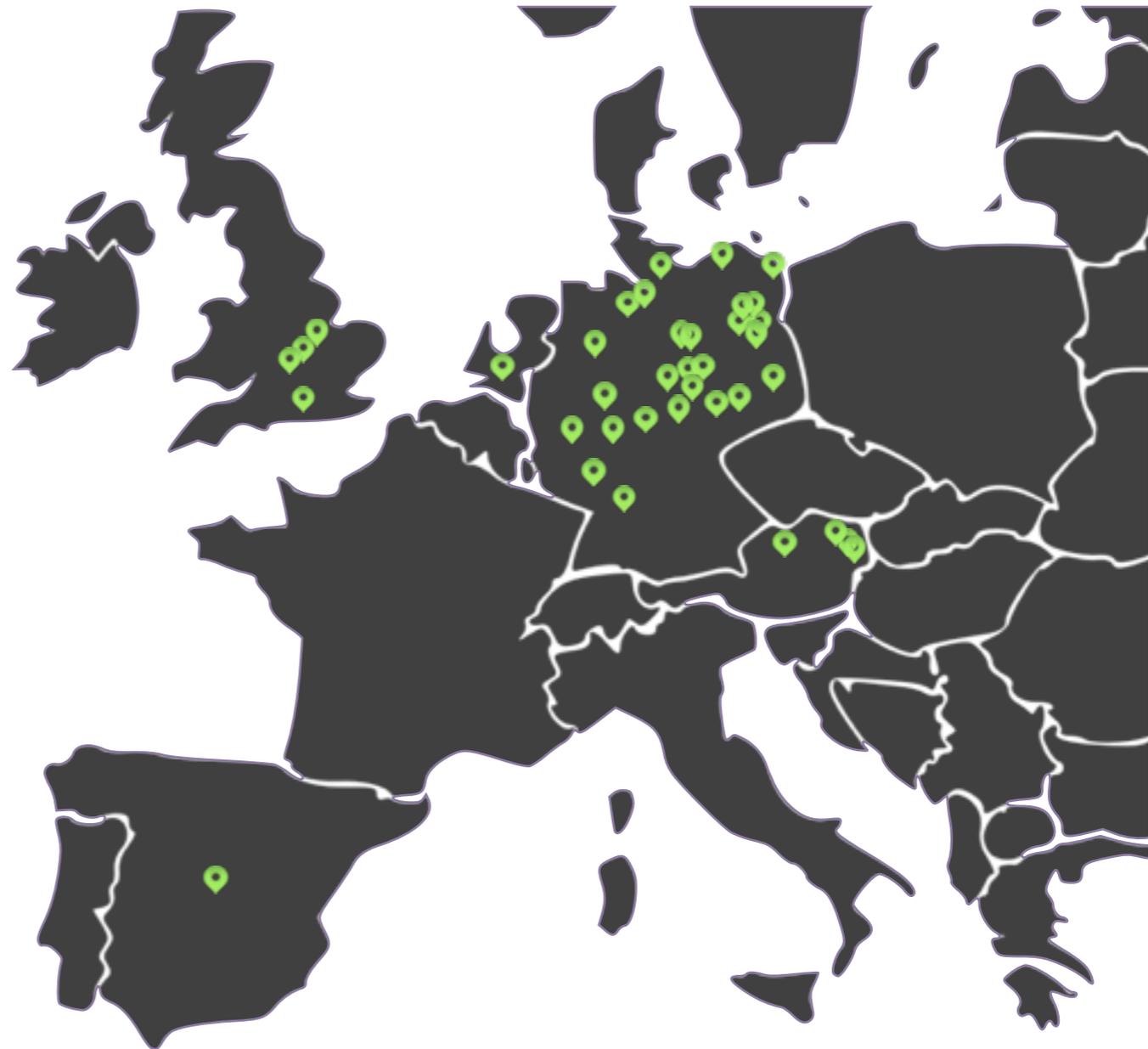
Licences (Usage)

Contract	Used	Total
tesseract	10	84
Vertrag 1	70	84



11. Goobi - production proven

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



A lot of happy content providers ↗



Questions?

intran da GmbH - Steffen Hankiewicz

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

Goobi

 intran da

