RCOS

国立情報学研究所
National Institute of Informatics

WEKO3: an Invenio 3 based Institutional Repository Flavor
- current development status and future plan -

National Institute of Informatics(NII)
Masaharu Hayashi
Kazutsuna Yamaji

IUGW2019@the University of Hamburg

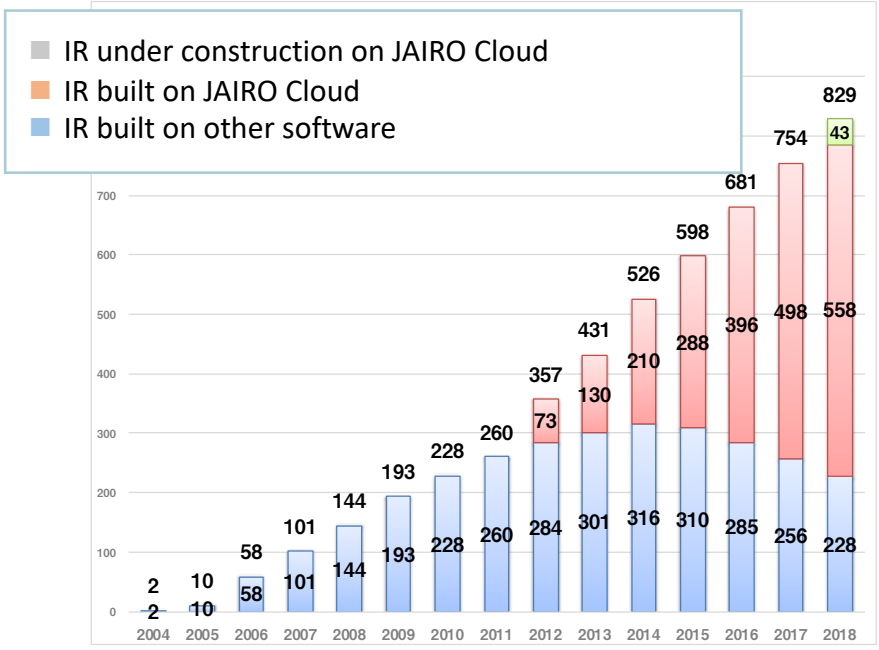
Table of Contents

- Background
- Introduction of WEKO3 project
- Characteristic functions of WEKO3
- Current issues in WEKO3 project
- Summary and future plan

Background

Japanese Institutional Repositories Online Cloud

- **JAIRO Cloud is a shared repository service** which is provided by NII since April 2012.
  - It aimed to support Japanese universities that have difficulty in building their IR with own efforts.



In March 2019, 558 IRs are running on JAIRO Cloud

Background

Repository Module on NC2
WEKO

- **JAIRO Cloud is based on** homegrown repository software, "**WEKO**".
- WEKO is a sort of **"out-of-the-box" repository software** which is developed by NII with reference to DSpace and EPrints functionality since 2008.
- **WEKO has** the following remarkable features.
  - **Metadata flexibility**: Metadata can be defined by the user, and can be OAI-PMH output by mapping to Dublin core schema.
  - **Easy operation**: All operations related to the repository can be done in the web browser.
  - **CMS integration**: Powerful site design capabilities are provided by Content Management System(CMS).
- WEKO has grown during capturing the user needs.
  However, due to the end of life of CMS,
  we were required to create new version of WEKO.

Why choose Invenio version 3

- Because…
  - **Invenio v3** has a **flexible data model**.
    - Therefore, We thought that it was possible to support the flexibility of metadata like the WEKO level.
  - Invenio v3 has **excellent extensibility and scalability**.
    - Scalability requirements were essential when assuming JAIRO Cloud operation. Invenio v3 is loosely coupled with the middleware was also attractive.
  - Invenio v3 has a **simple architecture**.
    - We thought that this would help the developer's understanding and donate to future maintenance.

We started "**WEKO3 project**"

that is **a development project**

of Invenio v3 based repository flavor

Brief history of WEKO3 Project

- FY2017: Initial development phase
  - Comparison study of open source repository software
  - Sent a engineer to CERN for learning Invenio v3
  - Started to develop invenio v3 based software
- FY2018: System development phase
  - Starting to implement the functions of previous version of WEKO
  - Sent a engineer to CERN for learning Invenio v3 more deeply
  - NII and CERN signed a memorandum of understanding on joint development of next generation repository system
- FY2019: pilot operation phase
  - Now

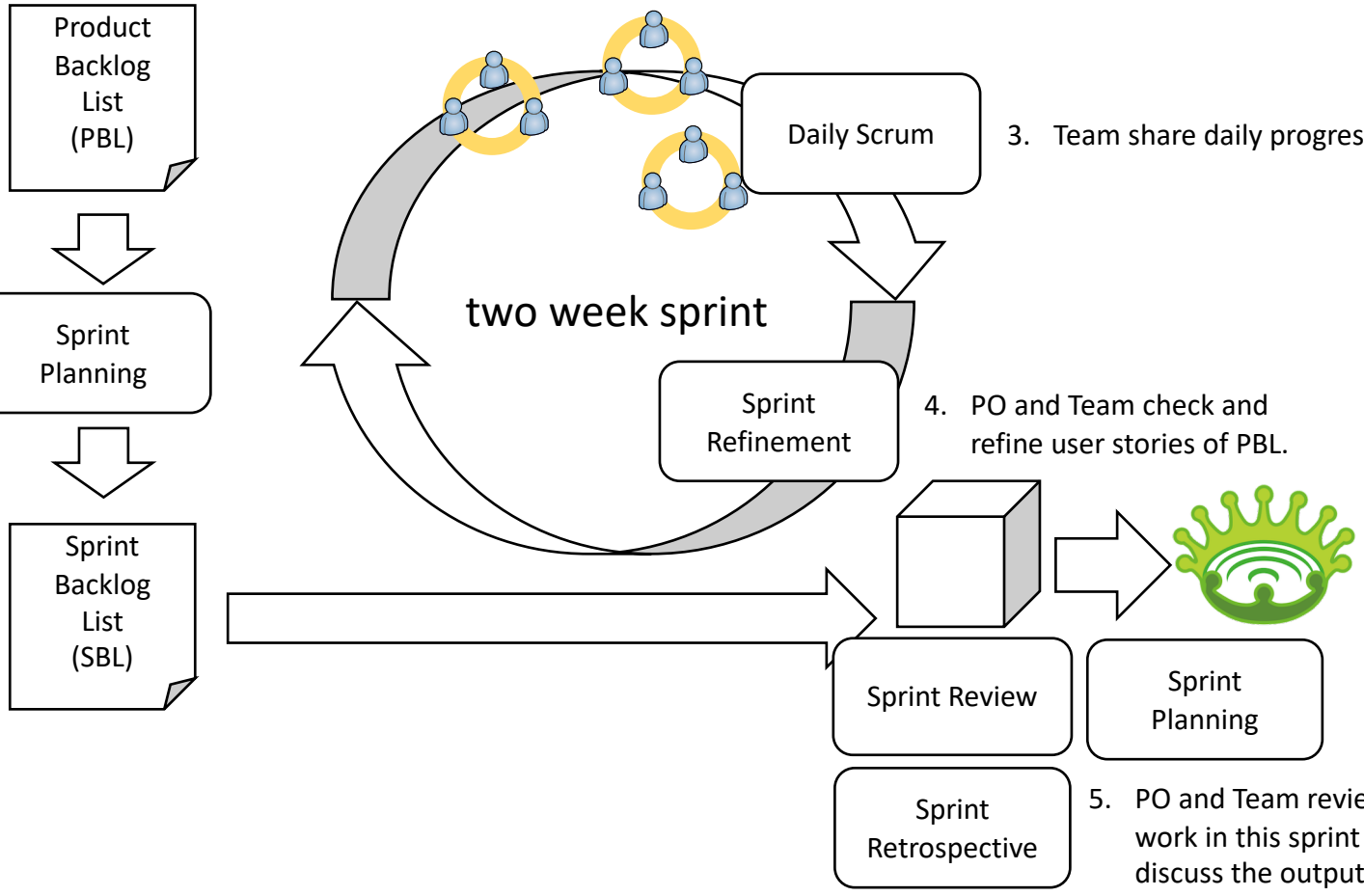
WEKO3 Development Style

Adopting Scrum, an agile software development style

NII: Product owner(PO) , 5 people
Software Development Company: Team,24 people
(SE x 6, PG x 13 including offshore members)

1. PO determine the user story priority

2. PO and Team put user stories of PBL into SBL in Sprint Planning Meeting.



3. Team share daily progress

4. PO and Team check and refine user stories of PBL.

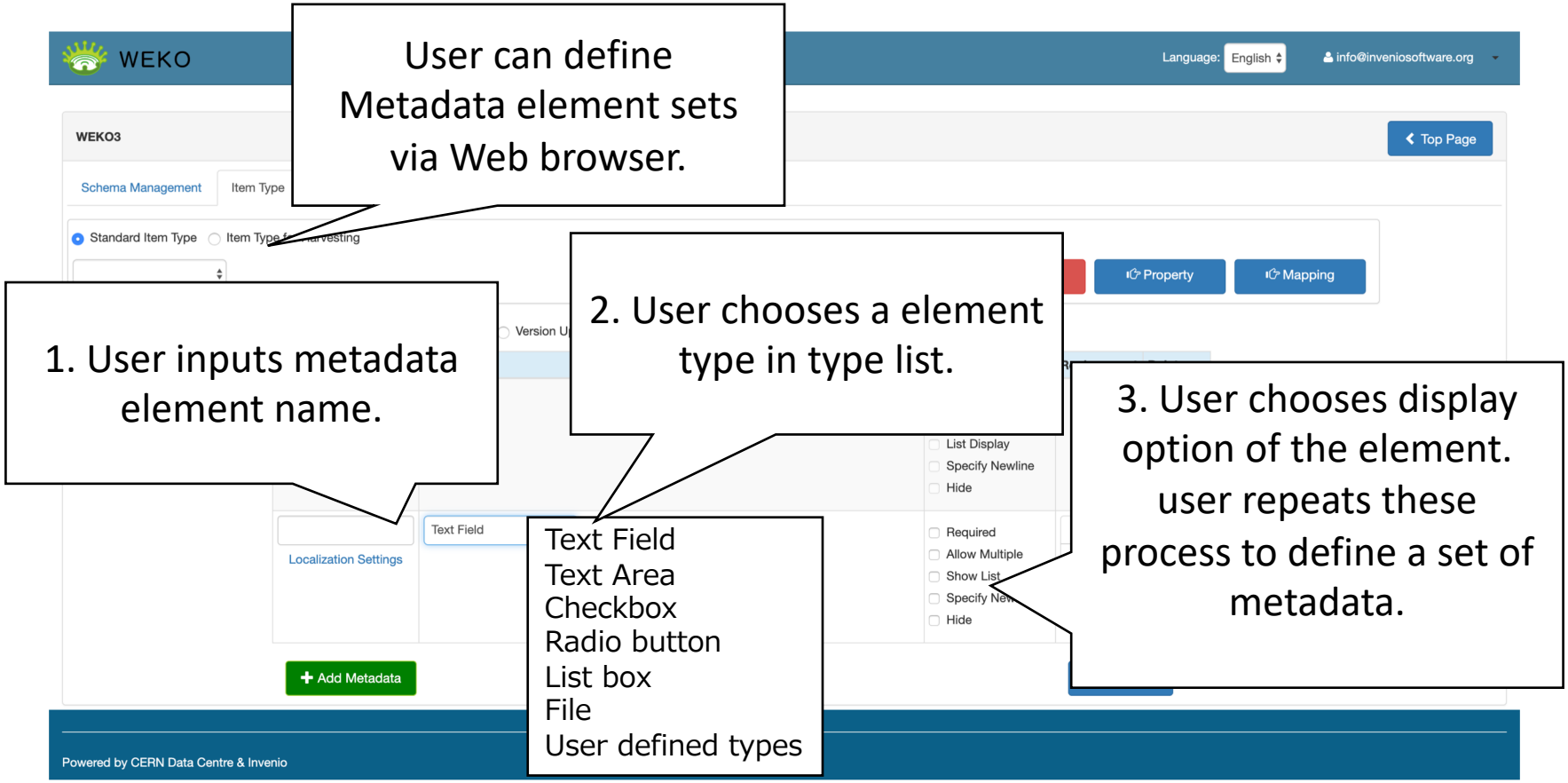
5. PO and Team review the work in this sprint and discuss the output and reflect next sprint.

Characteristic function of WEKO3

- WEKO3 has 3 characteristic functions:
  1. Metadata management function
  2. Repository site designing function
  3. Customizable workflow function

Metadata management function(1/3)

WEKO3 provides flexible metadata management function

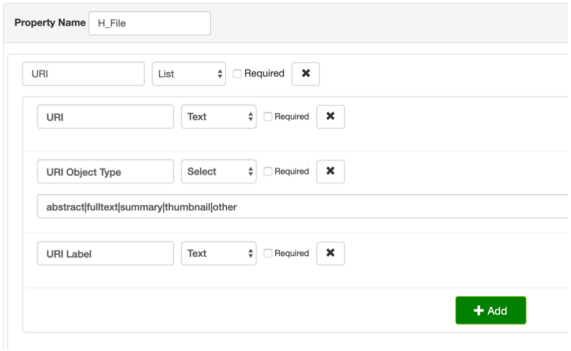


All operation related metadata management can be done on the web browser

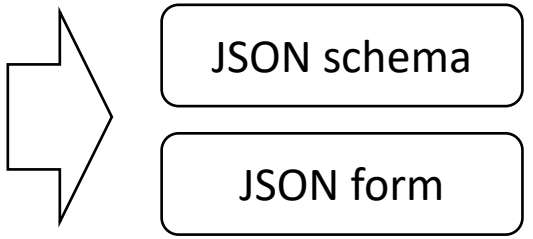
Metadata management function(2/3)

WEKO3 has JSON schema + JSON form editor as "metadata element sets editor”

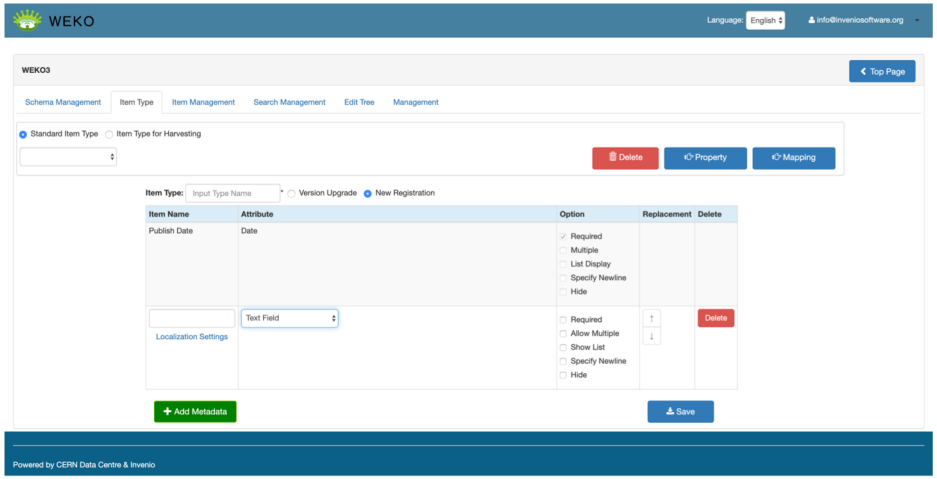
1. User defines an original element type in “element type editor”.



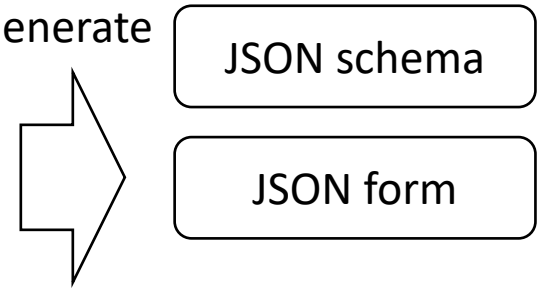
Generate



2. User combines built-in and user-defined element types to create a metadata set.



Generate

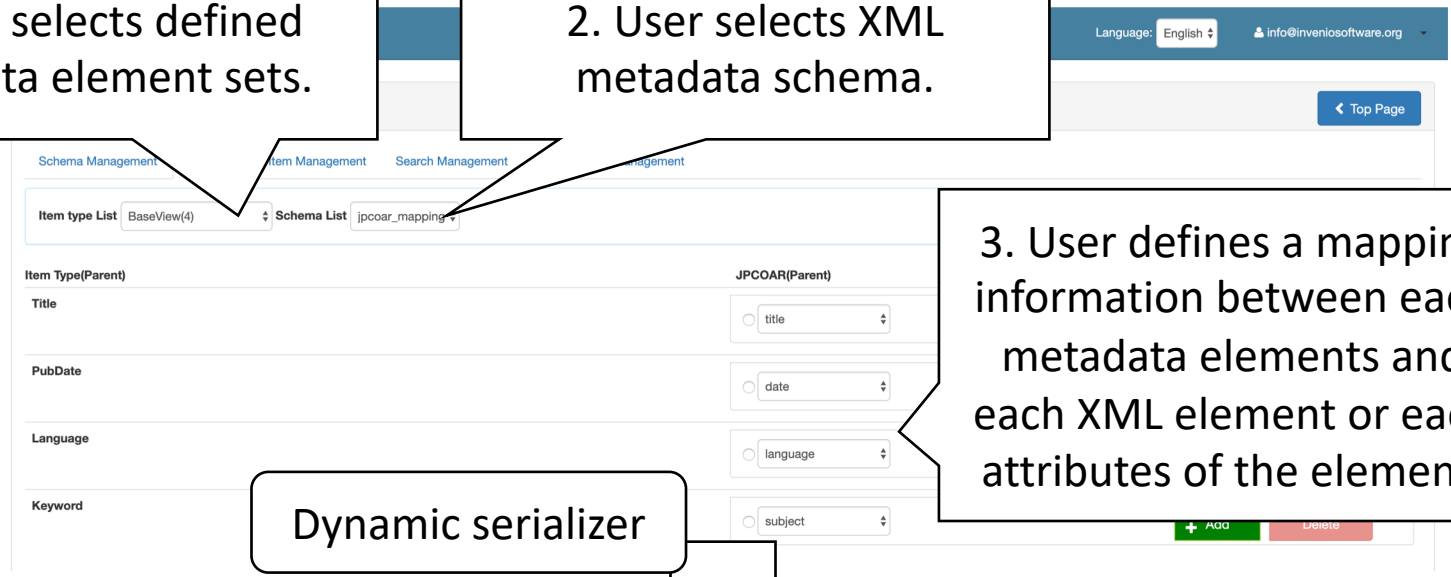


Metadata management function(3/3)

WEKO3 has a sort of dynamic serializer with XML Schema mapping

1. User selects defined metadata element sets.

2. User selects XML metadata schema.



3. User defines a mapping information between each metadata elements and each XML element or each attributes of the element.

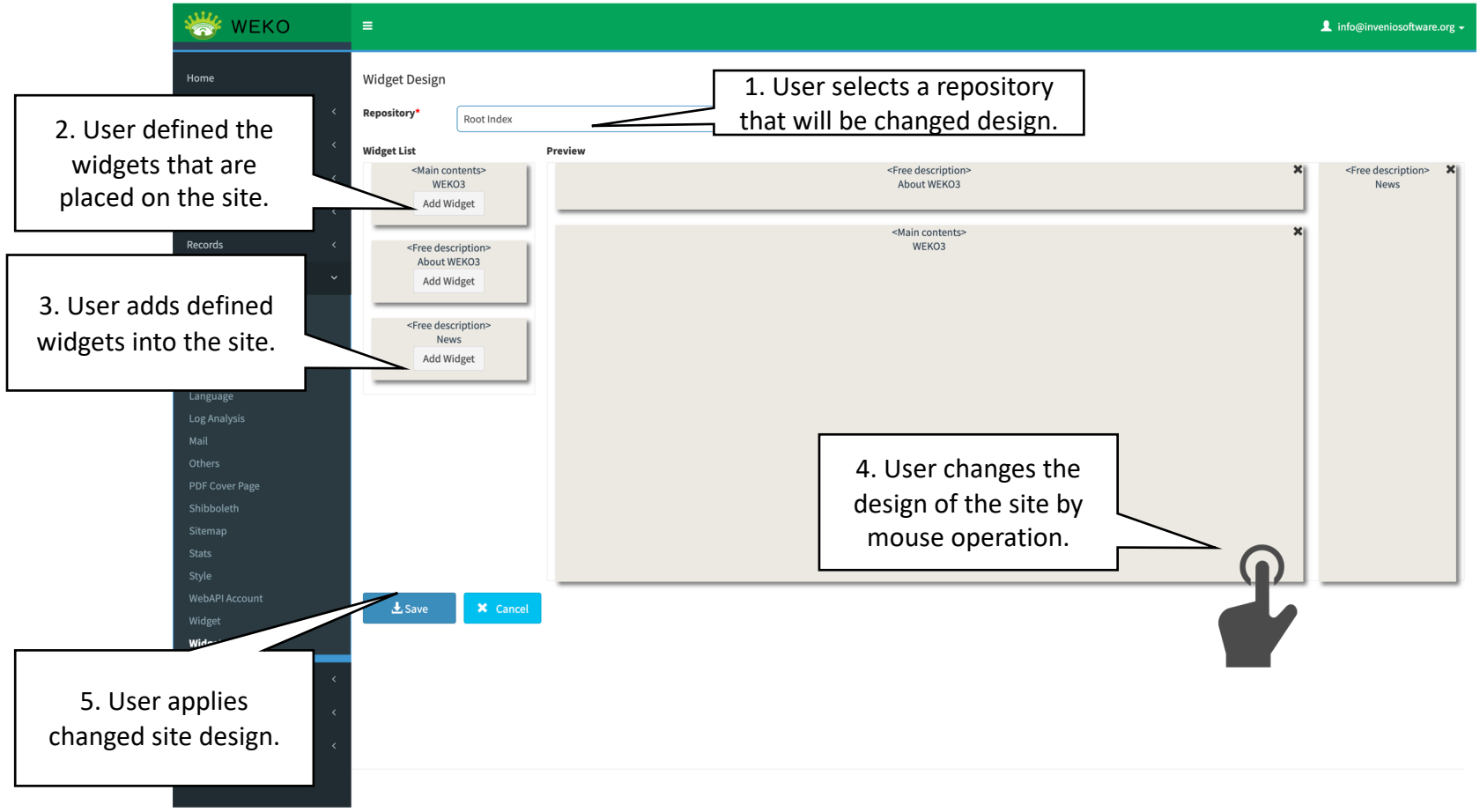
Dynamic serializer

Output formatted XML data (like a oai_dc)

WEKO3 possible to support a wide variety of metadata formats

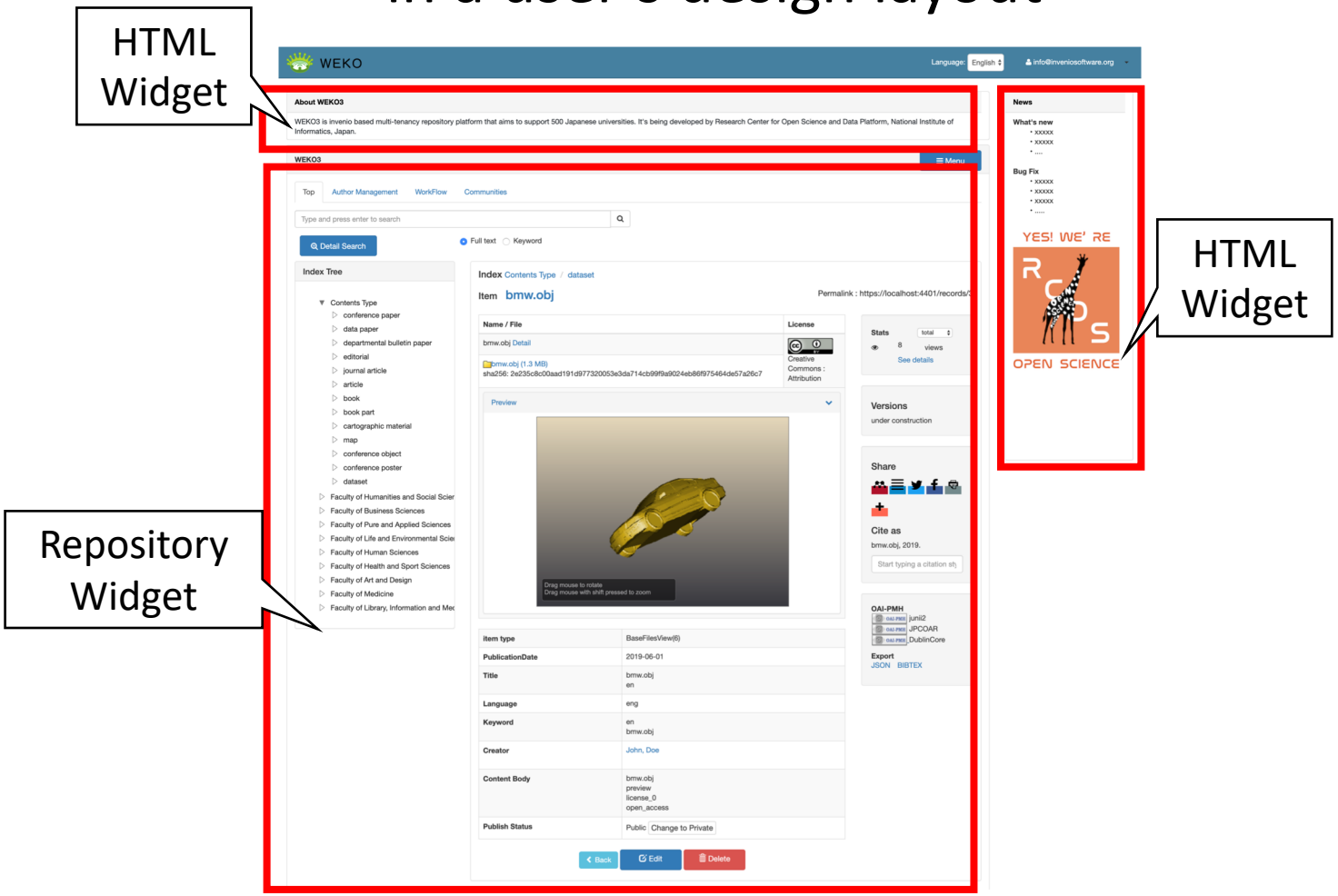
Repository Site Designing function(1/2)

WEKO3 can customize site design like a previous version of WEKO



Repository Site Designing function(2/2)

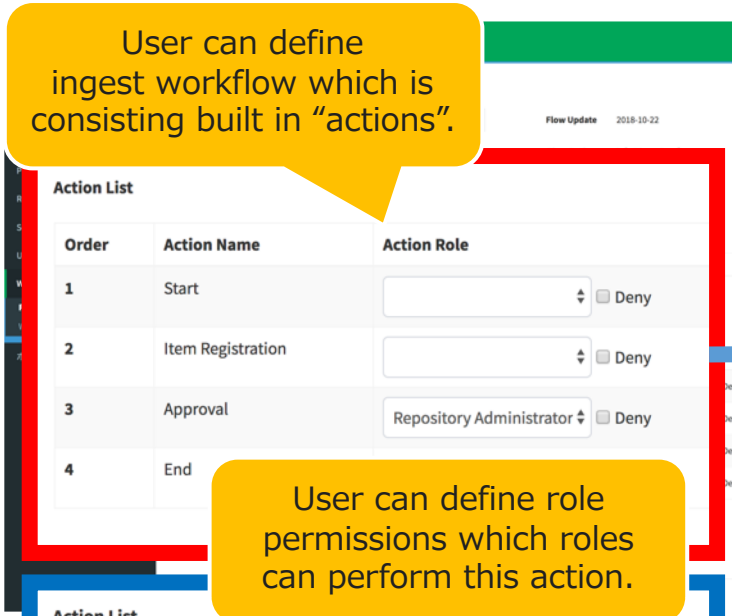
WEKO3 displays widgets, repositories in a user's design layout



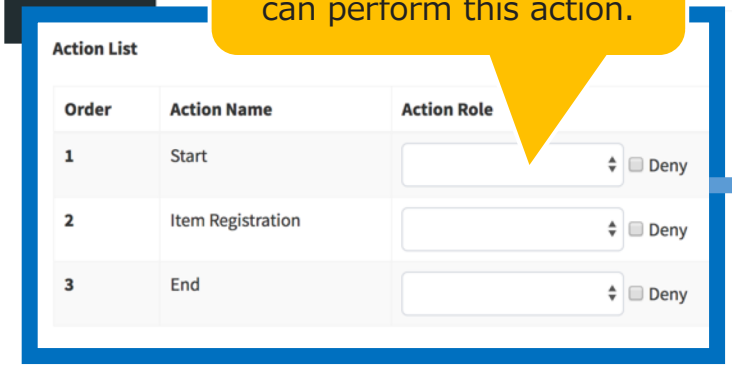
Customizable workflow function

In WEKO3, user can customize workflow according to the repository policy

User can define ingest workflow which is consisting built in “actions”.



User can define role permissions which roles can perform this action.



ingest workflow with approval

ingest workflow without approval

The workflow function is important for the repository to handle research data

Our development and forked/extended modules list

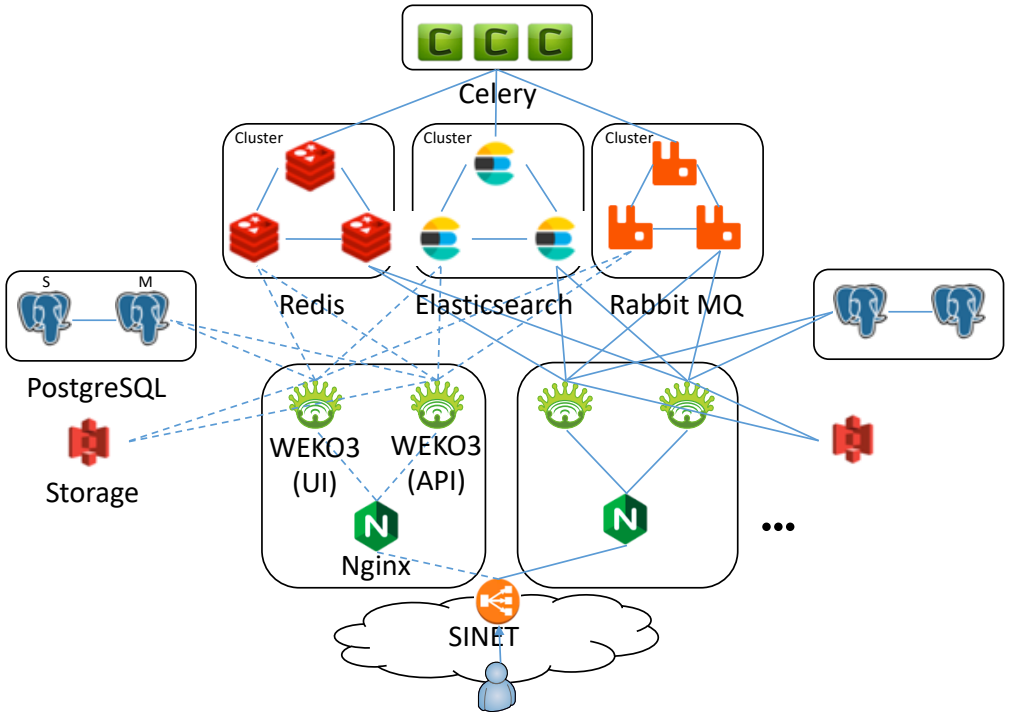
- Forked modules of Invenio v3
  - invenio-files-rest
  - invenio-s3
  - invenio-oaiharvester
  - invenio-previewer
  - invenio-stats
  - invenio-records-rest
  - invenio-communities
  - invenio-mail
  - invenio-oaiserver
- Extension modules of Invenio v3
  - weko-theme
  - weko-accounts
  - weko-admin
  - weko-logging
  - weko-groups
  - weko-search-ui
  - weko-records
  - weko-records-ui
  - weko-user-profiles
  - weko-sitemap
  - weko-deposit
- Original modules
  - weko-schema-ui
  - weko-index-tree
  - weko-authors
  - weko-workflow
  - weko-indextree-journal
  - weko-bulkupdate
  - weko-gridlayout
  - weko-itemtypes-ui
  - weko-items-ui
  - weko-items-autofill

Improving coding skills and moving to test-driven development are the challenges for our team.

Toward next version of JAIRO Cloud

Current version of JAIRO Cloud provides
over 500 institutional repository services on 25 servers.

Server spec per instance:
vCPU: 6
Mem: 16GB
SDD(OS): 20GB
HDD(DATA): 250GB or 500GB



Optimization of computational resources and
stable operation are important.
And the biggest challenge is data migration.

Summary and future plan

- Summary
  - We are developing Invenio v3 based repository software,WEKO3 that targets the next version of JAIRO Cloud.
  - WEKO3 has the same functions of the previous version of WEKO and implemented in sophisticated Invenio v3 architecture.
  - We shown some of the characteristic features of WEKO3.
- Future plan
  - We are planning to release an initial version of WEKO3 in the fall of this year.
  - A new version of JAIRO Cloud based WEKO3 is scheduled to start in FY2020.
  - Also we would like to implement the functions recommended by COAR NGR.
    - ResourceSync, SWORDv3

Thank you

https://github.com/RCOSDP/weko