Next Generation Repositories
Invenio

Kazutsuna Yamaji, NII
Lars Holm Nielsen, CERN
Jose Benito Gonzalez Lopez, CERN
An Overview of **Invenio**

- **Flexible** – Invenio is a framework for building large scale digital repositories: RDM, Media Archive, Institutional Repository, Search Engine, ILS, Brokers.

- **Scalable** – Handles 100M records and petabytes of files and integrates with any storage system.

- **Powerful search** – with the full capabilities of Elasticsearch such as geospatial queries.

- **Born for the web** – REST APIs everywhere.

- **Data model agnostic** – Natively based on JSON from storage to indexing.

- **Best of Open Source** - Built on top of the best: Elasticsearch, PostgreSQL/MySQL, Flask & Python.
Invenio Release Roadmap

• V3.0: June 2018 (OR2018)
  • Authentication & role-based access control.
  • Metadata storage, indexing and searching.
  • Persistent identifier management.
  • OAuth 2.0 client and provider: e.g. ORCID login.

• V3.1: November 2018
  • Files management – REST API, Previews and IIIF Image API.
  • Internationalization (I18N): e.g. support for Japanese.
  • Statistics.

• Note: Most modules are already running in large production systems for more than 1.5 year.
Long-term Vision: Flavours

• Invenio v3 Integrated Library System
  • RERO & CERN (in progress)

• Invenio v3 Research Data Management (RDM)
  • CERN & … (based on Zenodo and B2SHARE)

• Invenio v3 Videos and Photo Archive (MA)
  • CERN (based on videos.cern.ch)

• Invenio v3 Search
  • CERN

• WEKO3
  • NII & CERN

• ... actively looking for development partners....
WEKO3

• A flavor of Invenio v3 for Institutional Repository (IR) and Its Cloud Service

• Mainly features of WEKO3 (including feature plan)
  • Integrate WEKO2 Functions for IR based on Practical Requests from JAIRO Cloud Users
  • **Flexible metadata editor** and OAI-PMH provider feature
  • **Flexible workflow feature** for ingest workflow
  • **Cloud storage integration** for storing research data
  • **Multitenancy features** for providing perspective on needs of institution and the JAIRO Cloud service
Basic Concept for NGR Implementation

• COAR Next Generation Repository
  • To position repositories as the foundation for a distributed, globally networked infrastructure for scholarly communication

• Our approach:
  • To develop the NGR functions with users requirements and make the use cases of them
  • To develop the NGR functions based on the context of resource centric
  • To combine suitable services (e.g. repository, discovery and so on) for realizing NGR functionality
## Plan for Implementation of NGR on Invenio

<table>
<thead>
<tr>
<th>CERN is developing</th>
<th>Technologies</th>
</tr>
</thead>
<tbody>
<tr>
<td>ETag</td>
<td></td>
</tr>
<tr>
<td>ORCID</td>
<td></td>
</tr>
<tr>
<td>Sitemap</td>
<td></td>
</tr>
<tr>
<td>IIIF (Image API)</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>CERN will develop</th>
<th>Technologies</th>
</tr>
</thead>
<tbody>
<tr>
<td>Signposting</td>
<td></td>
</tr>
<tr>
<td>COUNTER</td>
<td></td>
</tr>
<tr>
<td>SUSHI</td>
<td></td>
</tr>
<tr>
<td>WebSub</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>NII will develop</th>
<th>Technologies</th>
</tr>
</thead>
<tbody>
<tr>
<td>ResourceSync</td>
<td></td>
</tr>
<tr>
<td>SWORD v3</td>
<td></td>
</tr>
</tbody>
</table>
# Plan for Implementation of NGR on Invenio

<table>
<thead>
<tr>
<th>Technologies</th>
<th>Release plan</th>
</tr>
</thead>
<tbody>
<tr>
<td>ETag</td>
<td>June 2018</td>
</tr>
<tr>
<td>ORCID</td>
<td>June 2018</td>
</tr>
<tr>
<td>Sitemap</td>
<td>June 2018</td>
</tr>
<tr>
<td>IIIF (Image API)</td>
<td>November 2018</td>
</tr>
<tr>
<td>Signposting</td>
<td>Early in 2019</td>
</tr>
<tr>
<td>COUNTER</td>
<td>Middle in 2019</td>
</tr>
<tr>
<td>SUSHI</td>
<td>Middle in 2019</td>
</tr>
<tr>
<td>WebSub</td>
<td>Later in 2019</td>
</tr>
<tr>
<td>ResourceSync</td>
<td>December 2018</td>
</tr>
<tr>
<td>SWORD v3</td>
<td>December 2018</td>
</tr>
</tbody>
</table>
Expected case study in CERN

• Interaction
  • IIIF: Search, preview and zoom images (e.g. 180k+ biodiversity images)

• Discovery:
  • Sitemaps: Google Indexing
  • Signposting: e.g. CrossRef/DataCite Event Data will know underlying DOI.

• Usage metrics counting and exposure:
  • COUNTER, SUSHI

• Scholarly link brokers (Asclepias Project):
  • WebSub: Real-time subscription and publishing of e.g. citation events in a distributed network of hubs.

• NGR:
  • Scholix for scholarly link exchange.
  • JSON-LD (schema.org) in landing pages.
  • Data Metrics (TODO: real name?).
Expected case study in NII

Resource Transfer

System

IRs in Japan

JAIRO Cloud

A Univ.
B Univ.
C Univ.

541 IRs

D Univ.
E Univ.

270 IRs

Discovery

Domestic

National Diet Library (NDL)
Doctoral Thesis Collection System

Japan Link Center (JaLC)
DOI Registration System

NII CiNii Articles
Discovery Service

International

OpenAIRE

BASE

Resource Transfer

ResourceSync

SWORD v3

GakuNin RDM

Research Data Management System based on OSF
Conclusion

• Invenio is a Flexible and Scalable framework for building large scale digital repositories

• Our Developing plan for NGR functions
  • Coming soon: ETag, ORCID, Sitemap, IIIF
  • CERN will develop: Signposting, COUNTER, SUSHI, WebSub
  • NII will develop: ResourceSync, SWORD v3

• We will develop NGR functions on user needs driven and will realize concept of NGR