Time to REST: The University of Stirling's REST API based integration between Worktribe and DSpace

DuraSpace-euroCRIS joint Repository/CRIS Workshop @OR2019
June 10th, 2019

Michael White
Senior Developer
Business Applications & Integrations
Information Services
University of Stirling

michael.white@stir.ac.uk
History/Background

STORRE: Stirling Online Research Repository
  • Mature, Full text only DSpace based Institutional Repository

Converis Research Management System (RMS)
  • Live 2012
  • “Out of the box" integration with DSpace - had “issues”

Worktribe selected to replace Converis (Aut 2016)
  • Cloud based/SaaS CRIS Supplier – number of “Modules”:
    ➢ Pre/Post Award, Outputs, REF, Repository ....
  • No existing DSpace integration
    ➢ Not CERIF Compliant
    ➢ Extensive REST API

Implementation Project Jan 2017 - May 2018
Worktribe REST API & “Action Subscription Queues”

Standard REST endpoints for each WT Entity:
• to Create (POST), Update (PUT), Get (GET), and Delete (DELETE)

WT API/Admin Users can create **Worktribe Action Subscription Queues**:
• API Endpoints for accessing Entity records based on Workflow or Edit events
• Snapshot of Entity added to Action Subscription Queue

2 types of Action Subscription Queue
• **PUSH Queue**: WT pushes relevant modified records to user supplied REST API endpoint
• **PULL Queue**: WT adds relevant modified records to Queue which user can poll on a schedule of their choice

The WT-STORRE integration makes extensive use of PUSH Queues – one each for:
• Output, OutputAuthor, OutputContributor, OutputFile, OutputFunder
STORRE Overview

Built on DSpace v6
  - Had to upgrade to get required REST functionality

DSpace REST API is based around Communities, Collections, Items, and Bitstreams
  - GET, POST (create), PUT (update), DELETE
  - “?expand=“ → Child/Parent objects

STORRE Community Structure based on University Structure
  - 1 Top Level Community per Faculty/Service Area
    - Some Faculties also have a sub-Community for each Division (“department”)

Collections based on Output Type
  - Each (sub)Community has 1 Collection for each Output Type deposited by their Faculty/Division

The WT-STORRE Integration has to work out where to deposit Outputs!
The API Layer Overview

• Provides REST endpoints to accept PUT requests from WT PUSH Queue
• Processes those requests and adds Outputs to STORRE Queue if appropriate
• Provides tools for Repository Librarian to manage items on the Queue
  • And manage mappings, create Collections etc
• Exports Queued Outputs to STORRE as required
• Writes Item’s handle back to WT
• Logs results & sends email report at end of export
Architecture Diagram

Worktribe

- Worktribe Push Queue sends Outputs for addition to STORRE Queue
- API Layer requests additional data relating to Output and related entities
- API Layer writes STORRE Handle back to Output record

BAI API Layer
worktribe-stirling Java/JAX-RS 2 application

- API Layer adds appropriate Outputs to STORRE Export Queue and Processes Queued Outputs
- API Layer inserts new Outputs/updates existing Outputs in STORRE

WT API DB

Library staff manage STORRE Export Queue via JSP pages

Workstation

STORRE

BE THE DIFFERENCE
Database Objects

STORRE Queue
- Output ID, Title, Last User, File Count, First Added Date, Last Added Date, Processed Date

Output Export Records – one per Output
- Last Modified Date, DSpace First/Last Export Date, DSpace Handle, DSpace Collection, Embargo End Date, Submit flag, Submit Date
- New Output Export records added for new Outputs via PUSH Queue

Mappings
- OrgUnit → DSpace Community
- WT Type → DSpace Type (→ DSpace Collection Names)
- DSpace Collection Names → Collection Handles
- WT OutputFile → DSpace Bitstream
Adding to STORRE Queue

- Every time an Output, OutputAuthor, OutputContributor, OutputFile, or OutputFunder is created, edited, or deleted:
  - Worktribe calls API Layer PUSH Queue endpoint

- API Layer (gets full Output record from WT and) checks Output:
  - is at an appropriate workflow status (Discoverable);
  - is of Type that is accepted in STORRE;
  - has been modified since it was last exported to STORRE;
  - has files for STORRE deposit attached (i.e. either Discoverable or Embargoed files)

- If output is eligible for deposit/update in STORRE, add to Queue
- Update Output Export record as appropriate
The STORRE Queue

### STORRE Export Queue Admin

**Migrate next N Worktribe Outputs to STORRE**

Number of Worktribe Outputs to export: [Export N Outputs to STORRE]

---

### STORRE Export Queue

**STORRE Export Queue key:**

<table>
<thead>
<tr>
<th>WT D</th>
<th>Title</th>
<th>Last Modified</th>
<th>First Export Date</th>
<th>Last Export Date</th>
<th>Collection</th>
<th>Handle</th>
<th>Embargo End Date</th>
<th>Submit Flag</th>
<th>Submit Date</th>
<th>Last Editor</th>
<th>File Count</th>
<th>Actions</th>
</tr>
</thead>
<tbody>
<tr>
<td>532284</td>
<td>Can Survey Participation Alter Household Saving Behaviour?</td>
<td>2019-06-04 09:43:17</td>
<td>2017-03-30 00:00</td>
<td>2018-01-09 12:32:59</td>
<td>null</td>
<td>1893/232</td>
<td>1893/2329</td>
<td>null</td>
<td>Yes</td>
<td>2017-03-10</td>
<td>Claire Allan (email)</td>
<td>1</td>
</tr>
<tr>
<td>136509</td>
<td>Social Identity Moderates the Effects of Team-Referent Attributions on Collective Efficacy but Not Dividends (Forthcoming)</td>
<td>2019-06-03 17:29:22</td>
<td>-</td>
<td>-</td>
<td>null</td>
<td>1893/521</td>
<td>null</td>
<td>Yes</td>
<td>2019-06-03</td>
<td>Claire Allan (email)</td>
<td>1</td>
<td>Update Record</td>
</tr>
<tr>
<td>534582</td>
<td>Saving the World’s Territorial Megafauna</td>
<td>2019-06-03 11:48:20</td>
<td>2016-07-28 00:00</td>
<td>2018-01-20 02:29:14</td>
<td>null</td>
<td>1893/128</td>
<td>1893/12684</td>
<td>null</td>
<td>No</td>
<td>2016-07-28</td>
<td>Tania Struemel (9522)</td>
<td>1</td>
</tr>
<tr>
<td>1261370</td>
<td>Responsible participation and housing: restoring democratic theory to the scene</td>
<td>2019-04-05 11:20:10</td>
<td>-</td>
<td>-</td>
<td>null</td>
<td>1893/232</td>
<td>null</td>
<td>Yes</td>
<td>2019-03-13</td>
<td>Elizabeth Fuller (email)</td>
<td>1</td>
<td>Update Record</td>
</tr>
</tbody>
</table>
Processing Queue records

Inserting:

- Checks OutputExport record:
  - Submit Flag, Submit Date
- GETs Output and related data from WT via WT’s REST API
- Transforms WT Output (etc) Attributes to Dublin Core metadata
  - Complicated by number of types, required richness of our metadata, ...
- Determines Community/Collection (if not already/explicitly set)
  - OrgUnit → Community/OutputType → Collection
- POSTs/PUTs Dublin Core and Files to STORRE
  - By default, only Admin has Anonymous READ Policy on Files
- Determines any Embargoes that need to be applied (complex spec!)
  - Sets Anonymous READ Policies as required
    - With or without Start Date

Updating:

- Checks if Metadata, Files, Collection actually require updating!
<table>
<thead>
<tr>
<th>Timestamp</th>
<th>WT ID</th>
<th>Action</th>
<th>Result</th>
<th>Item Handle</th>
<th>Title</th>
<th>Collection Handle</th>
<th>Files</th>
</tr>
</thead>
<tbody>
<tr>
<td>2019-06-04T01:00:00</td>
<td>1229916</td>
<td>Update</td>
<td>SUCCESS</td>
<td>1893/28845</td>
<td>Leading up 1893/236</td>
<td>[No Change: Stevens et al. (in press).pdf]</td>
<td>Item does not require updating in DS</td>
</tr>
<tr>
<td>2019-06-04T01:08</td>
<td>1263496</td>
<td>Update</td>
<td>SUCCESS</td>
<td>1893/25520</td>
<td>An incubation 1893/123</td>
<td>[No Change: map_23_08.pdf]</td>
<td></td>
</tr>
<tr>
<td>2019-06-04T01:10</td>
<td>1263533</td>
<td>Update</td>
<td>SUCCESS</td>
<td>1893/29250</td>
<td>Rhizosphere 1893/123</td>
<td>[No Change: Revised PLoS manuscript with Figures.pdf (2021-04-03)]</td>
<td></td>
</tr>
<tr>
<td>2019-06-04T01:12</td>
<td>1064675</td>
<td>Update</td>
<td>SUCCESS</td>
<td>1893/23845</td>
<td>Health-rel 1893/236</td>
<td>[Update: SheS_HRB_PAPER_FINAL_FINAL_PRE-PRINT.pdf (2019-12-10)]</td>
<td></td>
</tr>
<tr>
<td>2019-06-04T01:33</td>
<td>554582</td>
<td>Update</td>
<td>ERROR</td>
<td>1893/23934</td>
<td>Saving the 1893/123</td>
<td>[Update: Ripple et al 2016 Saving the World’s Terrestrial Megaflora: No Bitstream UUID found for WT Out</td>
<td></td>
</tr>
<tr>
<td>2019-06-04T01:33</td>
<td>1383003</td>
<td>Insert</td>
<td>SUCCESS</td>
<td>1893/29664</td>
<td>Do school 1893/213</td>
<td>[Insert: Curriculum_narrowing_ShapiroPriestley.pdf (2022-12-01)]</td>
<td></td>
</tr>
</tbody>
</table>
Thanks for listening!