Application of iRODS to NIEHS Data Management

Mike Conway, Deep Patel
Office of Data Science
National Institute of Environmental Health Sciences
What is keeping us awake at night?

Here are two of the things that there’s time to mention

https://flic.kr/p/97oo5F
Maintaining Relevance in Key Platforms/Standards
Maintaining Relevance in Key Platforms/Standards
Looking for Pathway to Play Together Nicely

- DRS is part of a suite of standards that support distributed execution of tasks, distributed data, and standard workflow execution environments, our “Compute to Data” story
- Gen3 is building DRS support into its platform
- Make iRODS a DRS platform


https://github.com/michael-conway/irods-ga4gh-dos
What’s Keeping us Awake at Night

- Handling metadata
  - Curation and getting beyond AVUs
  - Mechanics of ingest of data + metadata
  - Bolting SKOS and Synaptaic Graphite to our Commons
  - Indexing (on demand and near real-time)
  - I have an index, how can I search it without polluting community codebases?
  - I can search it, is it useable by relevant communities? How can I micro-target search?
Structuring Metadata, Metadata Models

Metadata Templates! Working Group making slow but visible progress, this is important!

Flexible Semantic Data Models and how they relate to our Commons
Vocabulary and Metadata Management

- How do we incorporate standard terms/labels in templates?
- How can we leverage templates and provide extensible search options and collection formation?
Pluggable Search

- Indexing Capability (iRODS Capability)
- Metadata Templates (MDT WG)
- Search Interface/Virtual Collection
- Pluggable Search For a Persona
Search Plugins follow simple OpenAPI Spec
Add endpoints in metalnx.properties

# Pluggable search configuration. Turn on and off pluggable search globally, and configure search endpoints.

# N.B. pluggable search also requires provisioning of the jwt.* information above

# configured endpoints, comma delimited in form https://host.com/v1

# enable pluggable search globally and show the search GUI components
pluggablesearch.enabled=true
Schema Plugins are interrogated and represented

**Data Grid Search API**

This is an extensible, pluggable search API for iRODS focused on pluggable Elastic Search indexes.

Contact the developer
Apache 2.0

<table>
<thead>
<tr>
<th>Servers</th>
</tr>
</thead>
<tbody>
<tr>
<td><a href="http://localhost:8082/v1">http://localhost:8082/v1</a> - Local dev server</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Info</th>
</tr>
</thead>
<tbody>
<tr>
<td>Endpoints for describing search capabilities and options</td>
</tr>
</tbody>
</table>

- **GET** /attributes/{index_name}  Find search attribute terms for a specific index
- **GET** /indexes  Find index types supported by this api
- **Search**  Search on index

- **POST** /search  Generic search on one or all available indexes
Plugins Advertise Supported Attributes in a Little Language

- Text entry in familiar ‘advanced query’ form to start
- Builder queries with autocomplete to be supported
Classic Search Result (Plugin can format in interesting ways, including sublinks)
ILS Type File Listing (WIP)