

Interfaces Update

Mike Conway Data Systems Architect/Engineer Deep Patel Software Developer National Institute of Environmental Health Sciences



NIEHS Office of Data Science Developing an NIEHS Data Commons

Developing a Commons to manage research data, using iRODS as a platform for unifying and managing local and cloud resources.

Image: Stand Links of Health and Human Services (S Image: Stand Links of Health and Human Services (S Image: Stand Links of Health and Human Services (S Image: Stand Links of Health and Human Services (S) Image: Stand Links of Health and Human Services (S) Image: Stand Links of Health and Human Services (S) Image: Stand Links of Health and Human Services (S) Image: Stand Links of Health and Human Services (S) Image: Stand Links of Health and Human Services (S) Image: Stand Links of Health and Human Services (S) Image: Stand Links of Health and Human Services (S) Image: Stand Links of Health and Human Services (S) Image: Stand Links of Health and Human Services (S) Image: Stand Links of Health and Human Services (S) Image: Stand Links of Health and Human Services (S) Image: Stand Links of Health and Human Services (S) Image: Stand Links of Health and Human Services (S) Image: Stand Links of Health and Human Services (S) Image: Stand Links of Health And Human Services (S) Image: Stand Links of Health And Human Services (S) Image: Stand Links of Links of Links (S) Image: Stand Links of Links of Links (S) Image: Stand Links of Links (S) Image: Stand Links of Links (S) Image: Stand Lin	National Instances of Health and Personal Processing Con NIEHS Public Site of Sourch here such Tor Such For Suff For Suff For Suff For Suff Manager Werkerg Computer Ubrary News Policels & Purchasing Sufety Training Decisions Decisions Decision S Data Commons Public Site of Suff Processing Surch here Suff Training Decisions Decisions Resources F Image: Processing Procesting Processing Processing Processing Processing Proc		_													 _
The Junction for NIEHS Staff Go to NIEHS Public Site of Search here Search For Lab For Offlice Managers Working Computer Library News Policies & Purchasing Safety & Training Divisions Divisions > Data Commons > Safety & Training Divisions Safety & Training Divisions <t< th=""><th>The Junction for NIEHS Staff Got NIEHS Public Stafe Search Area Search Conguter Ubrary News Publices Purchasing Security Training Durisons Conserts Conguter Ubrary News Publices Purchasing Security Training Durisons Conserts Resources Resources <!--</th--><th></th><th></th><th></th><th></th><th></th><th></th><th></th><th></th><th></th><th></th><th></th><th></th><th></th><th></th><th></th></th></t<>	The Junction for NIEHS Staff Got NIEHS Public Stafe Search Area Search Conguter Ubrary News Publices Purchasing Security Training Durisons Conserts Conguter Ubrary News Publices Purchasing Security Training Durisons Conserts Resources Resources </th <th></th>															
For Lab For Offlice For Managers Working Computer Library News Policides & Purchasing Safety & Training Duisions Divisions + Data Commons + Divisions + Divisions + Divisions + Divisions Divisions Resources Image: Profile + + + commonsProdZone + home + conwaync Image: Profile +	Port and sources Port and sources <th< th=""><th></th><th></th><th colspan="6">The Junction for NIEHS Staff Go to NIEHS Public</th><th>c Site 🕫</th><th colspan="2">Site & Searc</th><th colspan="2">earch here</th><th>Search</th><th></th></th<>			The Junction for NIEHS Staff Go to NIEHS Public						c Site 🕫	Site & Searc		earch here		Search	
Rules CommonsProdZone > home > conwaymc Image: Control of the convergence Collections > Search Q Templates & in have spaces Image: Collections + in have spaces	Resources Image: Conserver of the second			For Lab Staff	For Office Staff	For Managers	Working Here	Computer Support			Policies & Ethics		Safety & Security			
Resources Rules Bules Users Groups Profiles Collections Search Date Mathematical Martin Search Date Date Shared Links Favorites DateComposeSpesions 01312018 off Exh 012018 15 Exh 0120	Resources = Rules = Users = Groups = Profiles = Collections = Search Q Template & Shared Links < Public ~ Trash =															
Resources Rules Rules Groups Groups Profiles Collections Search Q Templates Shared Links Favorites DataComposeSpesions 01312018 off DataComposeSpesions 01312018 off Exholice 10 10 10 10 10 10 10 10 10 10 10 10 10	Resource Rude Users Group Profile Collections Search Q Template Sharred Link Public Trash Collections Col															-
Rules • * commonProdUctre * home * conwayme Users Groups Profiles Collections • • • betal Marit2 2018 15 • • • • • • • • • • • • • • • • • • •	Rules Users Groups Profiles Collections Search Q Templates Shared Links Favorites Public P Trash	Resources														
Users Action Showing 0 to 0 of 0 entries Groups Action Showing 0 to 0 of 0 entries Profiles Image: Showing 0 to 0 of 0 entries Collections Image: Showing 0 to 0 of 0 entries Search Q Image: Showing 0 to 0 of 0 entries Templates Image: Showing 0 to 0 of 0 entries Shared Links Image: Showing 0 to 0 of 0 entries Favorites Image: Showing 0 to 0 of 0 entries DataComposeShesions 01312018 off Eeb 01 2018 15 Image: Showing 0 to 0 of 0 entries Image: Showing 0 to 0 of 0 entries	Users à Groups * Profiles à Collections * Search Q Templates & Shared Links < Pavorites * Public • Trash	Rules	Ð	# > commo	onsProdZone > h	ome » conwa	ymc							• 0	•	
Groups Showing 0 to 0 of 0 entries Profiles Image: Showing 0 to 0 of 0 entries Collections Image: Showing 0 to 0 of 0 entries Search Q Image: Showing 0 to 0 of 0 entries Templates & Image: Showing 0 to 0 of 0 entries Shared Links Image: Showing 0 to 0 of 0 entries Favorites * Image: DataComposeSpesions 01312018 off Eeb 01 2018 15 Image: DataComposeSpesions 01312018 off	Groups * Profiles - Collections * Search Q Templates & Favorites * Public * Trash *	Users	۵	Action	~											
Profiles Image: File Type Modified Size Action Collections Image: File Type Modified Size Action Search Q Image: File Type Mar 12 2018 15 - Image: Vew Info Templates Image: File Type Jun 25 2018 15 - Image: Vew Info Shared Links Image: File Type Image: File Type Image: Vew Info Image: Vew Info Favorites Image: File Type Image: File Type Image: Vew Info Image: Vew Info	Profiles Collections Search Complates Shared Links Favorites Public Transh Image: Search Collections Image: Search Image: Search <td>Groups</td> <td>쓭</td> <td></td> <td>- 10</td> <td></td> <td>_</td> <td>_</td> <td>_</td> <td>_</td> <td>_</td> <td>_</td> <td>Shov</td> <td>ving 0 to 0 of</td> <td>0 entries</td> <td></td>	Groups	쓭		- 10		_	_	_	_	_	_	Shov	ving 0 to 0 of	0 entries	
Collections Image: Collections Search Q Image: Collections Templates Image: Collections Shared Links Image: Collections Favorites Image: Collections Favorites Image: Collections Image: Collections	Collections Image: Control of the second	Profiles	4	— њ	File Type					It Mo	dified	+ Size	It Ac	tion	11	
Search Q Image: Search Q Image: Search Q Image: Search Q Image: Search Q Templates Image: Shared Links Image: Shared Links Image: Search Q Image: Search Q Image: Verwindo Image: Search Q Shared Links Image: Search Q Image: Search Q Image: Search Q Image: Verwindo Image: Search Q Favorities * Image: DataComponeShesions 01312018 ndf Feb 01 2018 18. 1 2 MB Image: Verwindo Image: Search Q	Search Q Imbedal Mar 12 2018 15 - Imbedal Templates & Jun 25 2018 15 - Imbedal Shared Links <	Collections	•	- +=						+1	*				+1	
Templates Image: Shared Links Jun 25 2018 15 - Image: Week Info Shared Links Image: Shared Links Image: Shared Links Image: Shared Links Image: Shared Links Favorites Image: Shared Links Image: Shared Links Image: Shared Links Image: Shared Links Favorites Image: Shared Links Image: Shared Links Image: Shared Links Image: Shared Links Favorites Image: Shared Links Image: Shared Links Image: Shared Links Image: Shared Links	Templates Image: Imag	Search	Q		hetal					Mar	12 2018 15		0	View Info		
Shared Links Image: Shared Links	Shared Links < Favorites * Public © Trash ©	Templates	æ		🖿 i have spa	ces				Jun	25 2018 15	-	0	View Info		
Favorites * DataCommonsDesions 01312018 rdf Eeb 01 2018 18 1 2 MB • View Infr	Public © Trash	Shared Links	~		testupload					Sep	19 2018 15		0	View Info		
	Public O Trash D	Favorites	*		DataCommo	nsDesigns_	01312018.pd	f		Feb	01 2018 18	. 1.2 MB	0	View Info		
Public •	Trash	Public	0													
Trash 🗅		Trash														
								_								

Data Commons integrated with processing pipelines and workflow systems.



Use Case:

- Data Commons as the hub for managing research projects in an ISA model
- Sample submission integrated with Clarity LIMS triggers NextFlow pipelines
- Data Commons as delivery mechanism gathering metadata and pipeline results

Setting future strategy anticipating move to cloud over time, with a hybrid of local research data, published artifacts and tiered storage in the cloud.

How can we develop strategies that work for cloud and local use cases?



GA4GH Cloud Work Stream APIs



O'Connor, Frian, and David Glazer. n.d. "20190319 - GA4GH Cloud Work Stream Overview - Google Slides." Accessed June 25, 2019. https://docs.google.com/presentation/d/1_MFTCw1uDrFNtbki2Nvyh2l2IYOIQKTHmrZgMTspdm4/edit#slide=id.g 54dc8a46d6_0_0.

National Institutes of Health U.S. Department of Health and Human Services



What interfaces?

- Web Interfaces
- REST API
- Standard protocols



Web interfaces have been a tough road

Hard to get a consensus Hard to find resources Hard to develop something that is generally accepted/useful



Metalnx

iRODS Web Interface

- Released into open source by Dell/EMC
- o <u>https://github.com/irods-contrib/metalnx-web</u>





Plusses and Minuses

+	-
It is a thing that finally has some small amount of consensus	It's not X,Y,Z
Out of the box it has lots of interesting functionality, especially in administrative tools and user functionality	It has an opinion on various administrative and user functions that are not well-defined in core iRODS
Has many standard constructs and frameworks common in the enterprise	Mixed approach (page rendering versus AJAX) that cause occasional inconsistencies
Has actually been super-stable	A bit hard to initially deploy and configure
Provides a good starting point	Has some amount of technical debt to manage



- Initial simplification and draw-down of scope
 - Hide some functionality for now (e.g. Dashboard)
 - Migrate functions specifically implemented as a Metalnx service with database support into iRODS core server facilities
 - E.g. looking at group and user management
 - Simplify app structure
 - Extract 'services' and recast/refactor as standard extensions interfaces
 - Make services more universal
 - Allow gradual refactoring underneath
 - Minimize circular dependencies



- Migrate existing features to more widely accepted/applicable features
 - Extract features one-by-one and re-assess through community processes
 - E.g. Metadata Templates from an internal Metalnx and database implementation to a formal working group and a Python rule engine focused service
 - Consider essential non-core functionality and refactor
 - Move towards OpenAPI and Microservices
 - Make features configurable and language independent (also alleviates classloader issues in core)



- Try to use Metalnx as a way to bring in community innovations and shape as widely applicable features
 - Encourage migration to a microservices architecture
 - OpenAPI and focus on multi-client
 - Make additions optional and configurable
 - E.g. Notifications



- Use MetaInx as a tangible basis for discussing new client capabilities and patterns
 - Identify common user expectations and how they can be mapped to new capabilities
 - e.g. new **Pluggable Search**
- Migrate defined abstract interfaces into REST endpoints
 - Go piece by piece, over time, surfaced in Metalnx



Developer Workflow Enhancements

- Collapse MetaInx codebase into a single project
- Extract common services out, begin folding into abstract interfaces, remove circular dependency problems
- Move towards a .war overlay for site-specific skinning and enhancements
- Migrate services into pluggable, adaptable microservices with a wider client base
 - o JWTs
 - OpenAPI Specs
 - o Containers
- Maximize capabilities as iRODS.THINGS and minimize MetaInx.THINGS, especially irods-ext database
- Move to Vue.js over time



Notifications

- Based on CyVerse API (<u>https://github.com/cyverse-de/cyverse-de/cyverse-de.github.io/blob/master/api/endpoints/notifications.md</u>)
- Two initial use cases:
 - Notification of data delivery via processing pipelines to PIs and data requesters
 - Notification of data access requests and approval/disapproval



Notifications

С

🖲 Swagger UI	× +					
i localhost:880	0/v2/swagger-ui.html#/Notifications		ବ୍ ★	🕑 🏫	0	R V
(\cdot)	swagger	default (/api-docs) 🗘	E	xplore	
Dat This is <u>Conta</u> Apact	a Commons Notification API e a notification microservice for NIEHS data commons. act the developer he 2.0					
Noti	fications	Show/Hide Lis	t Operations	Expand Op	erations	i.
DELET	/notification/delete			Delete noti	fications	
POST	/notification/messages		A	Add new no	tification	
GET	/notification/messages/{userId}			Get all noti	fications	
PUT	/notification/seen		Update n	otifications	as seen	
GET	/notification/unseen-count/{userId}		Get count of u	unseen noti	fications	
GET	/notification/unseen-messages/{userId}		Get all u	unseen noti	fications	,

[BASE URL: /V2 , API VERSION: 1.0.0]



Example from DE

CyVerse Discov	ery Environment					r 📩 🕹
Data						
-	Notifications	S				
Apps	All	•	Refresh Vark as Seen	Delete		_
		Category	Message		Created Date	_
Analyses		Analysis	DE_Word_Count_analysis1 running		2019-08-21 13:46:44	
		Analysis	DE_Word_Count_analysis1 submitt	ed	2019-08-21 13:46:40	
				Rows per page: 100▼	1-2 of 2 🛛 🕹 🔍	> >1
Section Notifications						

Metadata Templates

Add Commercial Advanced any field or value classes Add term Control of the study Control of the study Control of the study study in the stu	Droject Code		localhost			C Brolect Certer	
Affent Image: Support s	<u>CEDAR</u>		Study Template			Template Creator 📃	
Image: Result VALUE	Add Element					a text	
Study RELE STUDY ASSAY FACTOR VALUE STUDY TIME Cardinality STUDY TIME Cardinality STUDY TIME Cardinality STUDY TIME Cardinality STUDY TIME Cardinality Study Time Study Time <td>RESULT VALUE</td> <td></td> <td>Enter Field Title</td> <td></td> <td></td> <td></td>	RESULT VALUE		Enter Field Title				
Intension of the study	DATA FILE		Study Title Enter Field Description (Help Text)				
FACTOR VALUE STUDY TIME TOUR SELEMENTS Caltrem Add Item Caltrems Caltrems Construction of the population Study Support St	STUDY ASSAY		The title of the study				
Image: Study time MORE ELEMENTS Add tem Image: Study time	FACTOR VALUE		Advanced	Cardinality		• •	
MORE ELEMENTS Set Our Add Item Construction ALITEMS Construction Chicked Study Class Details Chicked Study Chicked Study Studenter Class Details Chicked Study NAME CLINICAL STUDY ONTOLOGY Study Support Study Support Study Support	STUDY TIME	,	fou have not added any field or value class	es			
MORE ELEMENT Statue Acd Item Image: Statue in the statue in t				FIELDS	_		
Add Item ALITEMS Class Details Ontology Details Control Study Port Study Analysis Scientific Investigation Study Stoppet Study Stoppet Study Stoppet Risk Assessment Study Title Risk Assessment Study Title Risk Review Risk Review Risk Assessment Study Title Risk Review Risk Assessment Study Title Risk Review Risk Assessment Study Title Study	MORE ELEMENTS	~	State (
ALLITEMS Class Obtails Ontology Tree Class Details Ontology Outails Image: Clinical Study Ima	Add Item		Search study title			Q	
Cincial Study Pors-Study Analysis Scientific Investigation Production Predict Population Predint Population Predict Population Predict Population	ALL ITEMS	~	← Ontology Tree	Class	Details	Ontology Details	
Acronym CSO Patient Population Butty Study Support Butty Study Study Study Butty Study Study Study Butty Study Study Study Study Study Study Study Study Study Butty Study			Clinical Study B Post-Study Analysis	NAME	CLINICAL	STUDY ONTOLOGY	
Budy Support BRO Average IRB Average BRO Aver			 Scientific Investigation Patient Population 	Acronym	CS0		
IBB Renewal Bescription Ontologic describe general clinical studies. IBD Number Risk Assessment. # of Classes 37 IBB Expiration IBB Expiration ADD STUDY TITLE			 Study Support IRB Oversight 	@id	http://data	bioontology.org/ontologies/CSO	
ADD STUDY TITLE			 IRB Renewal IRB Approval 	Description	Ontology t studies.	o describe general clinical	
ADD STUDY TITLE			 IRB Number Risk Assessment 	# of Classes	37		
ADD STUDY TITLE			 Study Title IRB Review IRB Expiration 				
ADD STUDY TITLE							
AUDSTUDY TILLE							
			- *	ADD STUDY TI	ITLE		

- Define templates as JSON Schema
- Bind templates (required or suggested) to collections
- Present curation tools for templates
- Store metadata as AVU
- Validate and manage exceptions
- Retrieve as merged templates for editing and display
- Base new search tools on templates



Search

- Directly leverages the now-developing indexing capabilities
- Allow custom search schema (initial target ElasticSearch schema) to be plugged into MetaInx and other iRODS clients, even iRODS itself
 Inspired by PubMed Interfaces



PubMed Model

Select a provisioned search schema





PubMed Model

Schema defines applicable attributes for text search or advanced search builder

S NCBI Res	sources 🕑 How To 🗵	Sign in to NCBI
PubMed Hom	ne More Resources 🔻 Help	
PubMed Adv	vanced Search Builder	You Tube Tutorial
	Use the builder below to create your search	
Ē	Edit	Clear
E (Builder Affiliation ✓ All Fields Author AND ◆ Author - Corporate Author - First Author - First Searc Author - Full Author - Identifier Author - Last Book Conflict of Interest Statements	Show index list



PubMed Model

 Search can have a familiar 'Googleish' result format that can be customized by schema

SNCBI Resources 🖸	How To 🕑	
Bookshelf	Books	glioblastoma
		Browse lities Create alert Advanced
Previous versions or editions	Display Settings:	Summary, 20 per page, Sorted by Relevance Send to: -
Include in the search		
result	Results: 1 to 2	0 of 142 books (1035 items)
date		
1 year		<< First < Prev Page 1 of 8 Next > Last >>
5 years 10 years Custom range	1. D	<u>Glioblastoma [Internet].</u> De Vleeschouwer S, editor. Brisbane (AU): Codon Publications; 2017 Sep 27.
Added to Bookshelf		Top results in this book Table of Contents
60 days 1 year 5 years Custom range	2. MICAD B	Molecular Imaging and Contrast Agent Database (MICAD) [Internet]. Bethesda (MD): National Center for Biotechnology Information (US); 2004-2013. Top results in this book Table of Contents
Resource types Report Book Collection Documentation	3. J	Brain tumours (primary) and brain metastases in adults. National Guideline Alliance (UK). ondon: National Institute for Health and Care Excellence (UK); 2018 Jul. (NICE Buideline, No. 99.) ▶ Top results in this book Table of Contents



Search API

See <u>https://app.swaggerhub.com/apis/michael-conway/data-grid-search-api/1.0.0-oas3</u>

	/v1/openapi.json	Explore
Data Grid Search Morenapijson This is and extensible, pluggable search API for Contact the developer Apache 2.0	TAPI 1.0.0-cas3 (AS3) for iRODS focused on pluggable Elastic Search indexes.	
Servers http://localhost:8080/v1 - Local dev serve	и у	Authorize 🔒
Info Endpoints for describing search cap	pabilities and options	\sim
GET /attributes/{index_n	name) Find search attribute terms for a specific index	â
GET /indexes Find index types s	supported by this api	
Search Search on index		~
POST /search Generic search on o	one or all available indexes	a
Schemas		~
index-schema-description >		



Search API implementation responsibilities...

- Describe the schema supported at this endpoint
- Describe the attributes supported by this schema
- Accept searches in the following formats:
 - Free text "glioblastoma cell gene"
 - Free text with attributes "CELLLINE:xb4-124 ASSAY:RNA-Seq"
 - Builder style attributes



Search API implementation responsibilities...

- Interpret various searches and translate into appropriate query (E.g. ElasticSearch DSL)
 - Search hypotheses and project description as free text with fuzzy matching, search SampleID as an exact match
- Return queries in one of two formats
 - Google like result
 - Virtual Collection (ILS Listing)



Configuration in Metalnx

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

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

enable pluggable search globally and show the search GUI components pluggablesearch.enabled=true

JWT subject claim used to access search endpoint for data gathering. User searches will utilize the name of the individual

pluggablesearch.endpointAccessSubject=

timeout for info/attribute gathering, set to 0 for no timeout

pluggablesearch.info.timeout=0

timeout for actual search, set to 0 for no timeout

pluggablesearch.search.timeout=0



Metalnx Responsibilities

- Based on configuration, initialize and ping each endpoint to gather registration information (schema/attributes)
- Cache schema info to provision search interfaces
- Call search and present results
 - Google like result will include links to objects/collections
 - Virtual collection row/column format can be used as a familiar collection display



Example Implementation

- <u>https://github.com/michael-conway/grid-search-service</u> is an example
- There are many opportunities to add frameworks and abstract class support for development!
- Jargon-extensions-if contains support for adding search to other types of clients, REST API:
 - <u>https://github.com/DICE-UNC/jargon-extensions-if/tree/30-indexing/src/main/java/org/irods/jargon/extensions/searchplugin</u>



Streaming

- REST
- Metalnx
- Cyberduck
- NFS
- WebDav
- CIFS
- GA4GH DRS

All rely on streaming i/o

National Institutes of Health U.S. Department of Health and Human Services



The most important clientsupporting move the Consortium needs to make is a dedicated effort to <u>extract maximum performance</u> <u>out of streaming i/o!</u>

Maybe even more important that pluggable multipart transport, or could even be a key enabler!



Thank You!

Mike Conway Deep Patel NIH/NIEHS Office of Data Science

https://www.niehs.nih.gov/research/atniehs/dntp/osim/index.cfm

mike.conway@nih.gov deep.patel@nih.gov

National Institutes of Health U.S. Department of Health and Human Services