

Metalnx 2.0 – The Future of Metalnx



Digital Transformation,
Innovation, & Strategy

Steve Worth

WorthWize Consulting

Terrell Russell

*iRODS Consortium
RENCI at UNC-Chapel Hill*

Mike Conway

NIEHS

6/7/2018

Step back in time to 2014....

- Ebola became a global epidemic
- We started the search for MH Flight 370
- Fleetwood Mac reunites
- Brazil hosts the world cup
- A robot lands on an asteroid
- Russia hosts the Olympics

- There was ~4 ZB of data in the world
- Most everyone thought their big data lake would look like this →

- I was asked to help investigate how metadata would evolve for data management



In 2018 it is estimated there
is over 25 ZB of data
(heading to 45+ ZB by 2020)

The sad reality is that many
data lakes resemble this



Automated metadata generation, ingest, index, and search help tremendously

```
usxxpaeglem1:ino paegle$ ls -l
total 0
-rw-r--r--  1 paegle  CORP\Domain Users  0 Sep 23 12:29 my.bam
-rw-r--r--  1 paegle  CORP\Domain Users  0 Sep 23 12:29 my.log
usxxpaeglem1:ino paegle$
```


STUDY_ID: NIH2015-UO-0001456
POPULATION: 185
SUBJECT_BASE_TYPE: Shell
SUBJECT_TYPE: Mollusk
BASE_COLOR: 15
PATTERN_TYPE: Radial

But still struggle with time
varying and complex
context metadata

Why?

Because we build systems,
ontologies, indexes,
workflows, etc. to fit how
the machine processes
information, not how we
want to process data.



Add to this the explosion of data growth and associated costs.
It leads to some interesting data management practices.

(This is not fiction, it is happening today.)



Eliminate data

(destroy intermediate data,
assume it is cheaper and
possible to recollect data)

Pay per use (copies /
transfer / or both)



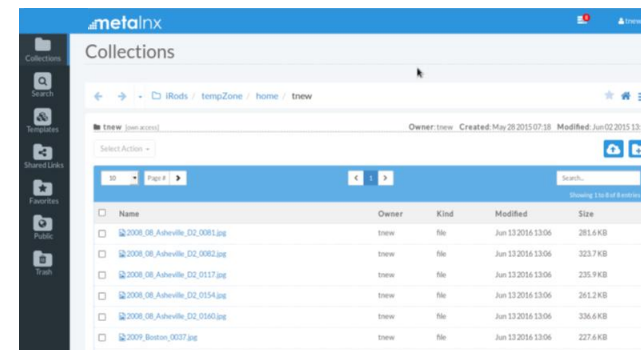
Keep minimal copies
(one)

Metalnx History

- 2014
 - Steve Worth & Sue Smith (EMC) studied how user defined metadata will evolve beyond basic indexing/searching/machine learning uses
 - Working with Sasha Paegle & Patrick Combes (EMC) about iRODS and its uses
 - During interviews with iRODS users and non-users we learned that non-programmers struggle with the power (& complexity) of the framework
 - Decided to gain insight by joining the Consortia and building Metalnx to hopefully draw in more commercial users and expand the use of user defined metadata
 - Brazil development team of Henrique Nogueira, Arthur Guerra, Milene Souto, and Men Yen Ng assembled at the Eldorado Research Institute

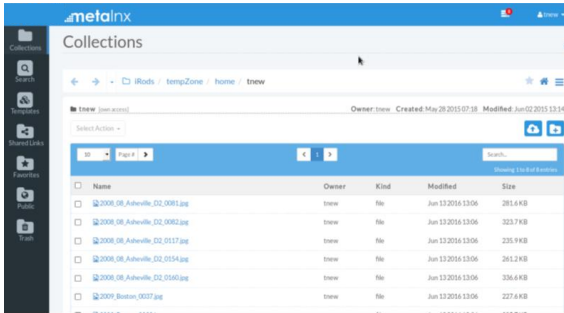
User commands (subset)

iinit
iput
iget
imkdir
ichmod
irm
ils



Metalnx History

- 2015
 - Revision 0.7 of Metalnx delivered a successful beta test with 3 user groups and received positive feedback at BioIT World 2015
 - Gave a Metalnx overview presentation and outlined a dynamic, on-demand, metadata mapping methodology
 - Revision 0.8 is demonstrated at the 2015 iRODS UGM with key features:



Graphical Collection Management



Automated metadata extraction framework



Metadata templates



Metadata searching

- Open Source approval process is initiated inside EMC

Metalnx History

- 2016
 - Awards!
 - Metalnx wins 2nd place overall in the EMC innovation competition
 - Metalnx is nominated for a BioIT World Best in Show award
 - Development continues, Rev 0.9 issued as a second limited customer beta test
- Dell acquisition of EMC results in an evaluation of all innovation programs for commercial potential
 - New business case study implemented per new corporate rules
- Work started to implement POCs (futures) as the review progressed:
 - user defined metadata type model
 - “on the fly” metadata typing (creating type structures as metadata is recorded)
 - Tiered metadata dictionaries (system, group, user) and priority models
 - Point to point transfer clients controlled by Metalnx
 - Metadata ontology validation
 - Metadata template chaining
- Resulting work demonstrated at an iRODS metadata working group meeting, with potential users, and internally across Dell EMC



Metalnx History

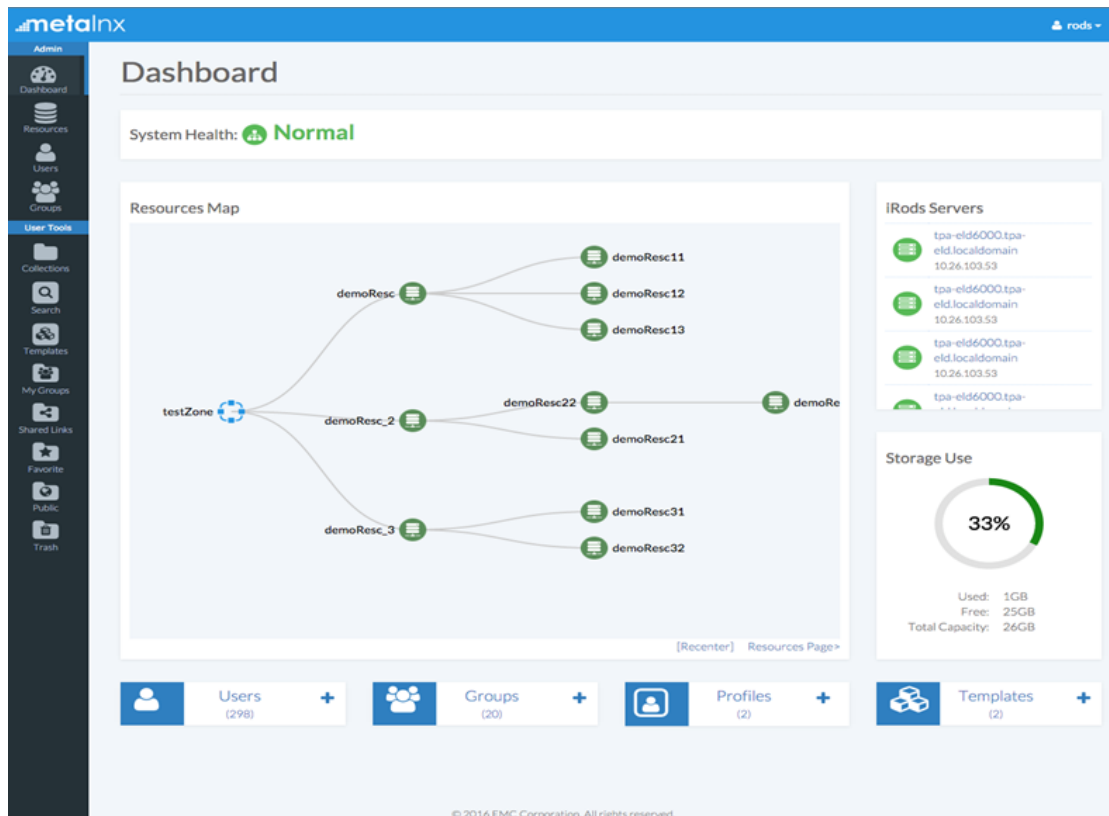


- 2017
 - Review decision reached to provide Metalnx as a Dell EMC open source tool
 - Rev 1.0 of Metalnx released on 3/10
 - Metalnx hosted on Github and on Code EMC sites
 - RPM and DEB packages hosted on bintray; Docker image hosted on Dockerhub
 - Rev 1.1 released 5/2:
 - Improved permissions handling, provide upload status tracking, adjusted SSL negotiation
 - Rev 1.2.1 release 6/13 coincident with 2016 iRODS UGM
 - Ticket support for iRODS and anonymous users
 - Rev 1.3.0 released 6/26:
 - Corrected problem were non-authenticated anonymous users could not see all tickets
 - Rev 1.4 released 7/19:
 - Add support for local rule file deployment across grid & fixed bugs

Dell EMC made decision to end Metalnx development and the project was shut down

Note: The logo designs scattered in this presentation are examples of logo designed that were considered, but not chosen. They added here for a bit of fun.

Key Features



- Basic rodsadmin (iadmin) support
- System status dashboard (& drill down)
- Granular permissions control
- Data replication at ingest
- Replication support
- Automated data extraction framework & tools for genomic data ingest
- Metadata templates
- Metadata & file property search
- Background uploading
- Favorites & shared links
- iRODS tickets
- Rule deployment across the grid
- ssl support
- Ldap integration tools
- Core microservice deployment verification

2018

- Dell EMC contributes Metalnx to the IRODS Consortium
 - License change to BSD3 approved
 - Metalnx Github repositories moved to irods-contrib
 - Metalnx domains(metalnx.com, metalnx.org, metalnx.info) transferred to the iRODS Consortium (for future use)
 - Metalnx working group is formed
- Work started to put Metalnx into CI
- NIEHS adopts and initiates changes to adjust Metalnx for internal needs
 - Changes provide better support for I18N, granular feature controls, skins, automated deployment, and changes to support a container deployment model
 - Conducted and passed a USGA Section 508 review
 - Modifications contributed back to the Consortia and integrated into the Metalnx master branch
 - This work becomes the foundation for Metalnx Rev 2.0

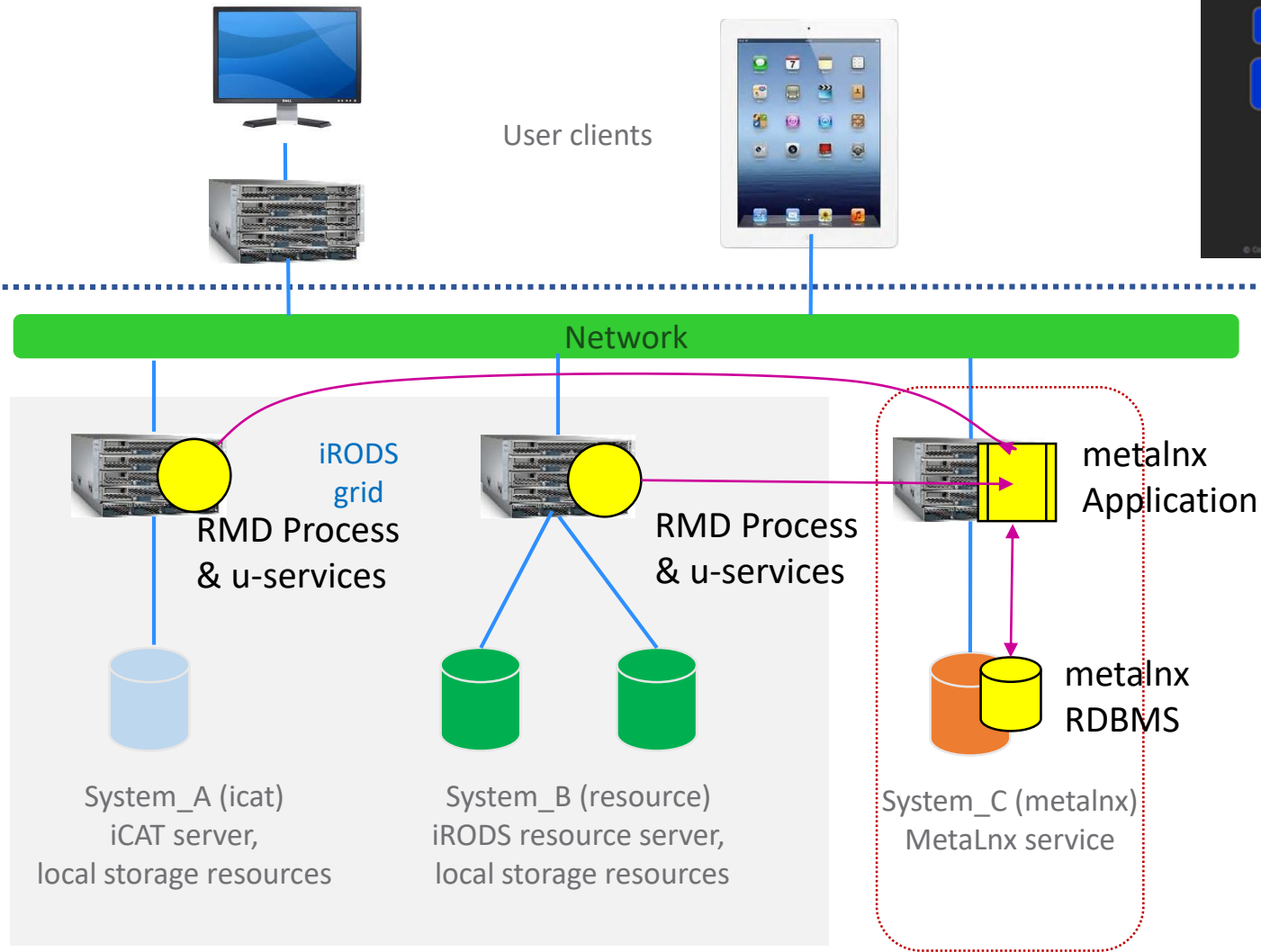
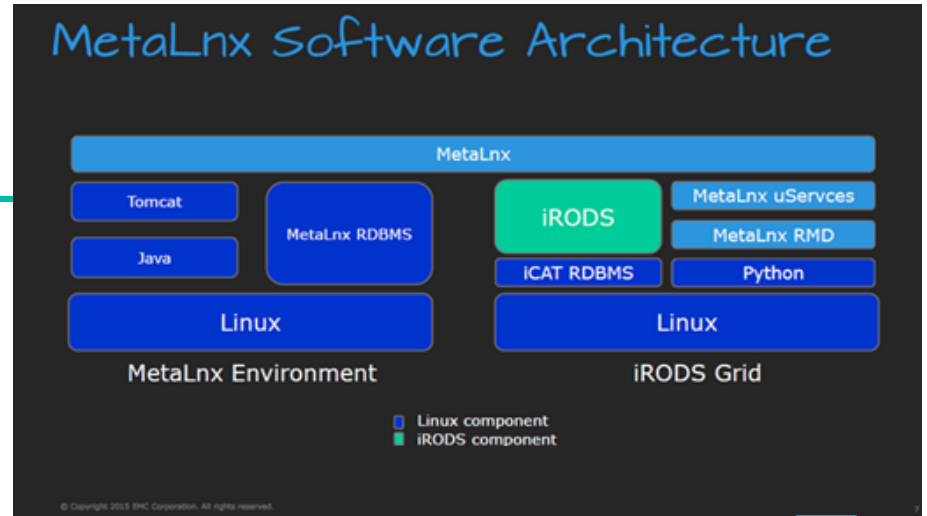


Where's Metalnx?

- Metalnx lives with the iRODS Consortium in irods-contrib!
 - metalnx-web
 - Core metalnx application
 - <https://github.com/irods-contrib/metalnx-web>
 - metalnx-rmd
 - Metalnx remote monitoring daemon
 - When installed on each node in an iRODS grid it will report to a rodsadmin detailed status information on the metalnx dashboard page (refreshes when dashboard page refreshed)
 - <https://github.com/irods-contrib/metalnx-rmd>
 - metalnx-msi
 - Collection of iRODS microservices used to extract genomic metadata and support rule deployment via the metalnx interface
 - <https://github.com/irods-contrib/metalnx-msi>
 - metalnx-irods-test-connection
 - Metalnx JAR used to test a connection to iRODS as part of a python install script.
 - Likely to be depreciated with a shift to a container deployment model
 - <https://github.com/irods-contrib/metalnx-irods-test-connection>



Metalnx Structural Model



Metalnx is designed to run alongside of iRODS or on an iRODS server.

Metalnx can scale for capacity by deployment of multiple instances.

Metalnx database can be Postgresql or MySQL.

Metalnx database can be on a separate database server or co-located with iRODS.

Metalnx R2.0

- Current version in master branch
- Key changes:
 - Docker image and/or .war file deployment going forward
 - (no more RPM / DEB packages)
 - Properties combined and moved into /etc/irods to help with skinning and Docker image release
 - Config & per-site property changes will go here
 - I18N support improvements
 - Change in login screen to support older browsers (IE8)
 - Upload speed improvements
 - Tickets / rule support can be turned on / off (off default)
- Software is beta, but near done:
 - Build & installation docs need updating
 - Metadata templates are the original metalnx model – will shift to the new model later
 - Metalnx-msi package must be updated for R4.2.3, work underway to correct
 - Tickets and rules are untested
 - Some bugs left to address



Metalnx – Consortium Adoption

- The process is underway, but to clear all hurdles...
 - Collect details on all Metalnx adopters and usage needs
 - Gain more metalnx users & community developers
 - Build a community based development model
 - Ensure future work is done in irods-contrib and to Consortium standards
 - Community Structure roles to be filled (positions can rotate)
 - Organizer (run monthly meeting, provide notes / updates, & guide assignments)
 - Working group (roadmap, alignment, and issues)
 - Coding volunteers (individuals or groups)
 - Code reviewers (includes test review)
 - Release coordinator (ensure builds, etc. work)
- In short – “We Need You!”



Metalnx – The Future

Old

Add Metadata

Attribute Name

Age

Type

Range

▼

Minimum value

10

Maximum value

15

Value

11

Unit

years

Cancel

Add

New

Add Metadata Attribute

Attribute Name (Required)

Value Set-up

Data Type

Enter the data type of the value

List

Data list

Type each option in the list on a separate line

MiSeq
NextSeq
NovaSeq

Default Value

This will be the default value when the template is applied

Select value from list

Hint Text

Enter helpful information about this metadata

Unit

☐ Required

Cancel

Add

- Metalnx & metadata working groups to be combined
- Add support for consortium metadata models, dictionaries, & ontologies
- Add support for iscan & ifsck
- Develop a model for workflows to provide notification of data completion steps
- Auto request for FAIR and other blocked data access
- Automated & semi-automated iRODS account requests from Metalnx login screen
- Quota support

And Now a Short Demo....

