Hi we're Natalie Meyers and Rick Johnson from the PresQT Team
Here’s a link to the google slides we’re presenting

[SHARE IN CHAT]

https://docs.google.com/presentation/d/1YcaLB-HDb81p9GT5LrRaRCKbmU8D-ISdDRL59-7NUHw/edit?usp=sharing

Or come back anytime using this URL
https://osf.io/3GNQD/
PresQT

An implementation grant and previous planning grant funded effort to address needs for preserving data and software. The goal is to collaboratively design, develop, and connect interoperable and repository agnostic Data and Software Preservation Quality Tools.

- Implementation grant: https://www.imls.gov/grants/awarded/LG-70-18-0082-18
- Planning grant: https://www.imls.gov/grants/awarded/lg-72-16-0122-16

Presqt is . . . An implementation grant and previous planning grant funded effort to address needs for preserving data and software.

The goal is to collaboratively design, develop, and connect interoperable and repository agnostic Data and Software Preservation Quality Tools.
We’re now in the implementation grant’s QA phase. We’ll describe the services and features we’ve built and how you and your organization can:

1. build an endpoint or nominate one,
2. learn to use the features that interest you,
3. and try our FAIR tests.

This is an open grant. All the planning phase resources gathered during stakeholder engagement are available to inform stakeholders’ future projects. Tool designs and final report will be actionable by anyone who wants to inform their planning through the priorities identified in PresQT.

Everything produced for the PresQT project is shared on github and the Open Science Framework (OSF)

- Open resources at [https://osf.io/d3jx7/](https://osf.io/d3jx7/) [SHARE IN CHAT]
- Open source software at [https://github.com/ndlib/presqt](https://github.com/ndlib/presqt) [SHARE IN CHAT]
PresQT Acknowledgements

Funded sub-awardees
- Sheridan Libraries, Johns Hopkins University
- National Data Service
- HUBzero team, Purdue University
- UC San Diego Library
- Yale University Library (EaaSI)

Collaborators and Testing Partners
- Members of Confederation of Open Access Repositories (COAR), Libraries at Amherst College, Fontbonne University, Tuskegee University
- Center for Open Science, CERN, Data Curation Network, Project Jupyter, Midwest Big Data Hub, ReproZip, RDA Groups, Science Gateways Community Institute, Software Preservation Network, URSSI (US Research Software Sustainability Institute)

Workshops:
Data Futures: Preserving Annotation with Peter Cornwell
SDSC: David Valentine and Ilya Vasilavsky
Mark Wilkinson: FAIR Evaluation Services
Daniel Clarke and Avi Ma’ayan FAIRShake, Assessment Rubrics

We’d like to acknowledge and thank PresQT Stakeholders including:

Funded sub-awardees
- Sheridan Libraries, Johns Hopkins University
- National Data Service
- HUBzero team, Purdue University
- UC San Diego Library
- Yale University Library (EaaSI)

Collaborators and Testing Partners
- Members of Confederation of Open Access Repositories (COAR), Libraries at Amherst College, Fontbonne University, Tuskegee University
- Center for Open Science, CERN, Data Curation Network, Project Jupyter, Midwest Big Data Hub, ReproZip, RDA Groups, Science Gateways Community Institute, Software Preservation Network, URSSI (US Research Software Sustainability Institute)

Workshops:
Data Futures: Preserving Annotation with Peter Cornwell
SDSC: David Valentine and Ilya Vasilavsky
PresQT’s Planning phase task was to identify Data and Software Preservation Tool gaps

Through Surveys and workshops we established an understanding of the concerns of:

- Different Repositories, Different Cultures of data management, processes, technology/tools, etc
- We heard the Switch cost between stand alone systems is [TOO] high for researchers;
- And that, Data flow/transactions between systems can be very difficult

Our analogy for this situation is a research data management archipelago - where stand alone systems and the data in them grow in isolation from one another and there are data islands.
PresQT Inter-Repository File Transfer

*Building Bridges Between Systems*

PresQT’s charge coming developing out of the planning phase was developing tools for connecting communities of practice and bringing preservation features together without moving files up and down to the desktop.

This way, Users and repository managers can move projects between systems and enhance their assets with fixity information, FAIR testing and better keywording along the way.

Data islands become connected by PresQT’s services.
Accessing PresQT Services

Web GUI: https://presqt-prod.crc.nd.edu/ui/


Feel free to stand up your own PresQT Services from the PresQT open source distribution: https://github.com/ndlib/presqt

You can visit the Web GUI at presqt-prod.crc.nd.edu/ui/ [SHARE IN CHAT]

and try the services described in today’s talk

It’s a stateless auth token based utility service prototype we built for enabling wide audiences to engage in testing the PresQT services

Or, You can access the PresQT endpoints directly through the API [SHARE IN CHAT] https://presqt.readthedocs.io/en/latest/api_endpoints.html

Or, You can stand up the PresQT service using the open source distribution on github https://github.com/ndlib/presqt [SHARE IN CHAT]
PresQT Service Features

**PresQT design:** pluggable, configurable architecture easily extendable for diverse systems under consideration of diverse structures

**Set of services:**
- Configure additional partner systems via JSON and Python functions
- Transfer in BagIt format
- Metadata in JSON
- Fixity checks
- Keyword enhancement
- EaaSI Emulation
- FAIR tests

https://presqt.readthedocs.io/en/latest/

PresQT (Preservation Quality Tool) is an open-source toolsuite with RESTful services to improve Preservation and Re-use of Research Data and Software

At (and between) each endpoint you can use PresQT for a variety of services

The PresQT services are documented at https://presqt.readthedocs.io/en/latest/

[SHARE IN CHAT]
Transfer between Repositories
presqt-prod.crc.nd.edu/ui/

And More: Establish and Check Fixity information, Enhance Keywords w/SciGraph, Do FAIR testing, Connect to Emulation services

You can use the GUI or API to
- transfer files between repositories,
- establish and check fixity information
- enhance keywords
- connect to emulation services
- or do FAIR testing.

Try it yourself at presqt-prod.crc.nd.edu/ui/

[SHARE IN CHAT]

The presqt system runs on token based authentication so you'll need to get access tokens from the service endpoints you are interested in

Don’t worry if you’ve never done this before, the documentation explains what tokens are and how to get them for the service endpoints you are interested in.
We connect to the services you see above

The services denoted with an asterisk are integrations in progress.

**If you want to plug in one of your own services you can work with us to connect via APIs**
• Here’s more detail about service features for each PresQT endpoint
During the resource upload process, fixity is checked in two locations:

1. When files are saved to the disk from the request.
2. After files are uploaded to the target.

Let's look at Fixity:

**First Fixity Check**
Resources are delivered in the POST request in BagIt format as a zip file. After unzipping the file and saving it to the server we validate the bag using BagIt's built in validator. If any files saved don't match the manifest originally given then the fixity has failed and the server will return an error.

**Generate New Hashes If Necessary**
We now know that the currently saved files are the same as what the user sent forward. Before uploading resources to the target we will make sure that there is a dictionary of hashes available generated by a hash algorithm supported by the target. If the target supports a hash algorithm provided by the resource's 'bag' then we will simply use those. If not, then we need to generate new hashes based on a target supported hash algorithm.

**Second (Post transfer Fixity Check)**
After resources are uploaded to the target, we compare the resources' hashes
brought back from the target to the hashes we captured before. If any hashes
don’t match then fixity fails. Since the resources have already been uploaded we
simply capture which resources’ fixity fails and pass that along the response
payload along with the message, ‘Upload successful but fixity failed’.
Keyword Enhancement

When transferring a resource you have the option of either manual or automatic keyword enhancement.

Acknowledgements: David Valentine & Ilya Vasilavsky, SDSC
SciGraph https://github.com/SciGraph/SciGraph

Keyword Assignment During Transfer

You and your organization can use PresQT services to auto-enhance Keywords via SciGraph thanks to workshop help from David Valentine and Ilya Zaslavsky at SDSC.

You can get more information at https://github.com/SciGraph/SciGraph [SHARE IN CHAT]

When transferring a resource you have the option of either manual or automatic keyword enhancement.

- **Manual enhancement** will only add source keywords and the keywords provided in the request body.
- **Automatic** will add all enhancements including any provided in the request body.
  - These can be set by setting `presqt-keyword-action` in the headers to either manual or automatic

Manual Keywords
If \texttt{presqt-keyword-action} is \texttt{manual} then PresQT will only add keywords found in the source target and keywords given in the body of the request. This means you need to get the possible enhancements before initiating a transfer.

**Automatic Keywords**

If \texttt{presqt-keyword-action} is \texttt{automatic} then PresQT will add keywords found in the source, keywords given in the request body, and any keyword enhancements found during the transfer process. The following steps occur during the transfer in this case:

1. Fetch all source keywords both in the target and in the FTS metadata file for the transferred resource.
2. Get enhancements with the given enhancer (Defaults to SciGraph for now).
3. Upload keyword enhancements to the Source Target and Destination Target.
4. Add the keyword enhancements to the FTS Metadata file that gets written to the Destination Target during the transfer.
5. Add the keyword enhancements to the FTS Metadata file that gets written to the Source Target during the transfer.
Practical Example/Demo of using PresQT for Keyword expansion
Keyword Enhancement w/ Transfer

Utilizing [github.com/SciGraph/SciGraph](https://github.com/SciGraph/SciGraph)

Or try the service on PresQT web GUI against your own endpoints
Keyword Enhancement can also be done without transferring.

1. Use the **Keyword Enhancement GET** endpoint to fetch the keywords from the resource.
2. Pass the keywords you want to enhance to the **Keyword Enhancement POST** endpoint.
3. Enhanced keywords will get uploaded to the target and a new action will get written to the FTS metadata file.
Practical Example of how keyword enhancement without transferring.

1. Use the Keyword Enhancement GET endpoint to fetch the keywords from the resource.
2. Pass the keywords you want to enhance to the Keyword Enhancement POST endpoint.
3. Enhanced keywords will get uploaded to the target and a new action will get written to the FTS metadata file.
The FAIR principles were introduced in 2016. You can see them on the left. The guiding principles allow many different approaches to rendering data and services Findable, Accessible, Interoperable, to serve the ultimate goal: the reuse of valuable research objects.

What FAIR is...

FAIR refers to a set of principles for humans or machines/software.

The FAIR principles emphasize that re-useful data (Data which actually can be reused) will be become as valuable as is possible.

From Cloudy, Increasingly FAIR written by FAIR’s originators and stakeholders emphasizes that the principles deliberately do not specify technical requirements, but are a set of guiding principles that provide for a continuum of increasing reusability, via many different implementations.

The principles describe characteristics and aspirations for systems and services to support the creation of valuable research outputs that could then be rigorously evaluated and extensively reused, with appropriate credit, to the benefit of both creator and user.
....and what FAIR is not

**FAIR is not a standard:** The FAIR guiding principles are sometimes incorrectly referred to as a ‘standard’, even though the original 2016 publication explicitly states they are not [25].

FAIR simply describes the qualities or behaviours required of data resources to achieve – possibly incrementally – their optimal discovery and scholarly reuse.

Standards are prescriptive, while guidelines are permissive.

FAIR originators and stakeholders suggest that a variety of follow-on valuable standards can and should be developed atop the FAIR principles, each of which can be guided by the FAIR Principles.
Here’s an example of that incremental approach. Here you see an infographic from ARDC Australian Research Data Commons that communicates essential elements of FAIRness in easy to understand chunks aligned with the principles.
Once funders like ARDC began to request FAIR data of their funded projects, the stakeholder community began pairing compliance aspects with the principles.

In turn, the growing maturity in FAIR digital objects and repositories began to create an information ecosystem where FAIRness could be tested.

But what should FAIR testing aim for?

This is tricky to implement because remember FAIR is for people and machines!

Early tests relied on surveys people would fill out answering questions about their data.

While FAIR testing today is increasingly automated and focused more on whether data is FAIR for "machines" / software.

To learn more about automating FAIR assessment, PresQT did workshops with Mark Wilkinson, Daniel Clarke and Avi Ma’ayan to learn their approaches to FAIR assessment.

Then we stood up FAIR testing using Mark’s evaluator service for
PresQT endpoints, let’s take a look:
The FAIR Maturity Indicator Authoring Group wanted to provide an objective, automated way of testing (meta)data resources against the FAIR Maturity Indicators. They defined maturity indicators and tests aligned with the FAIR principles, aligned with the metric they test for. You can browse them online. I’ll share them in the chat so you can have a look at:

https://fairsharing.github.io/FAIR-Evaluator-FrontEnd/#!/metrics

[SHARE IN CHAT]

You can read more about the indicators in the papers on the left. Mark Wilkinson and the group also created the FAIR Evaluator which runs as a demonstration service. You can see a screenshot of it on the right. I’ll share the URL in the chat:

https://fairsharing.github.io/FAIR-Evaluator-FrontEnd/#!/

[SHARE IN CHAT]

The Evaluator provides a registry and execution functions for:

- Maturity Indicator Tests
- Community-defined Collections of Maturity Indicator Tests
- Quantitative FAIRness evaluations of a Resource based on these Collections
Demo runs

FAIRshake is another web-based software toolkit that enables the assessment of compliance of biomedical digital research objects with the FAIR guiding principles.

I’ll paste a link to FAIRshake into the chat: [https://fairshake.cloud/](https://fairshake.cloud/) [SHARE IN CHAT]

FAIRshake also functions as a repository to store and serve FAIR assessments.

It provides FAIRness assessments of three types of digital objects:
1. Datasets
2. Tools,
3. and repositories/databases,

are based on answers to a set of questions

It has a concept of a test rubric that lets you group the tests you need together.

We’d like to thank the FAIRshake project for workshopping with PresQT particularly Daniel Clarke and Avi Ma’ayan, PhD Mount Sinai Center for Bioinformatics Icahn School of Medicine
We are working on integrating FAIRshake into PresQT testing.

Now let's take a look at what else is in progress:
Expanded services connecting:

- FAIRshake *(FAIR testing with rubrics)*
- EaaSI *(Emulation Nomination)*
- HUBzero *(Repository Connection)*
- WholeTale *(Repository Connection)*

In addition to the FAIRshake work:
- We created a demonstration integration with EaaSI. We'll show an example on the next slide.
- In 2021 Software Preservation Network Members have the opportunity to use a SPN hosted version of EaaSI - we anticipate another round of testing the PresQT prototype Emulation proposal launcher then to see how it can perform in the hosted environment.
- We also have Integration testing with WholeTale and Hubzero underway.
Let’s take a look at EaaSI:

- With the PresQT service now we can connect to the Emulation as Service aka EaaSI decentralized node network’s Universal Virtual interactor
- For that We’d like to thank
  - Klaus Rechert, and the OpenSLX team in Freiburg
  - the EaaSI team at Yale especially Ethan Gates, Seth Anderson and Euan Cochrane
  - And Jessica Meyerson at SPN
Like we mentioned, in 2021 SPN is partnering with EaaS! to offer scaled, high-quality emulation services to members beginning in January 2021 and run until December 2021.

Here’s a link to more information about the pilot that outlines the goals and structure for this exclusive, one year testing pilot of the EaaS! beta hosted service for participants in SPN member organizations.

https://www.softwarepreservationnetwork.org/hosted-emulation-services-pilot-summary/

[SHARE IN CHAT]

During this pilot, SPN would act as the first point of contact for EaaS! Hosted Emulation Services for its members, standardizing and overseeing EaaS! onboarding, and synthesizing feedback/data from members in order to inform the EaaS! development roadmap and service design.

SPN will standardize access to hosted emulation services by acting as a single “customer” and thereby streamlining invoicing, evaluation, onboarding, and documentation activities for SPN members.

Find out more at:

https://www.softwarepreservationnetwork.org/get-involved/
All this means here’s never been a better time to join SPN:

Membership is 5k, and joining is easy!

Just Download, complete, and return the SPN Membership Agreement to Jess Farrell
PresQT QA Opportunities to Connect

- **Sign up for our mail list** at presqt.crc.nd.edu by **Dec 4th, 2020**
- **Attend the PresQT community call** 1 PM EST Dec 7th, 2020
- **Contact us** at presqt-contact-list@nd.edu
  - Try the File Transfer QA Tests between systems you use like github & zenodo or figshare
  - Try the Keyword Enhancement QA Tests
  - Try the FAIR testing out on an object of your own at a PresQT endpoint

How else can you get involved with Data and Software Preservation?

- **Join the PresQT mail list before December 4th** and you’ll get an invitation to participate in QA
- You can join using form at presqt.crc.nd.edu [SHARE IN CHAT]
- **Then to onboard to PresQT’s QA** please encourage interested parties from your Organization to attend **PresQT community call 1PM on Monday Dec 7th** (afterward the calls will be offered twice monthly starting up again in 2021)
- **On your own:**
  - Try the File Transfer QA Tests between systems you use like github & zenodo or figshare
  - Try the Keyword Enhancement QA Tests
  - Try the FAIR testing out on an object of your own at a PresQT endpoint
- **Contact us and give feedback about the service or endpoints you want to add**
We also have some upcoming talks where you can learn more about PresQT:

You can find PresQT’s co-PI & Tech lead Sandra Gesing from Notre Dame’s Center for Research Computing this coming Monday at the Trusted CI Webinar. Here’s a link to the webinar: [https://blog.trustedci.org/2020/11/trustworthy-data-panel.html](https://blog.trustedci.org/2020/11/trustworthy-data-panel.html)

[SHARE IN CHAT]

At Trusted CI you can learn more about the efforts of the NSF’s Cybersecurity Center of Excellence and its Trustworthy Data Working Group (TDWG) in the broader context of related work by PresQT, NIST, and RDA-US.

And then a week from Wednesday Sandra will participating in the American Geophysical Union’s Session on “Best Practices and Realities of Research Data Repositories: Which One Should I Choose to Publish My Data?”

As you can imagine, the PresQT File Transfer service makes these decisions less about end-of-the-road one-size fits all solutions which we hope will interest you!
Thank You from the PresQT Team  
Justin Branco, Brett Fox, Sandra Gesing, Rick Johnson, Natalie Meyers, Noel Recla, Miranda VanNevel, and John Wang  

presqt.crc.nd.edu  

These slides: osf.io/3gnqd doi:10.17605/OSF.IO/3GNQD  

Right now you and your organization can: Add an Endpoint, Enhance your Keywords, Do FAIR testing!  

Together we are: Collaboratively designing, developing, and connecting interoperable and repository agnostic Data and Software Preservation Quality Tools  

An implementation grant (LG-70-18-0082-18) & previous planning grant (LG-72-16-0122-16) funded this effort to address needs for preserving data and software.