PresQT – How to achieve FAIRness in your gateway

Sandra Gesing
sandra.gesing@nd.edu
https://presqt.readthedocs.io/en/latest/
February 10, 2021

Science Gateways Community Institute Webinar

Hesburgh Libraries
FAIR Principles

- Findable
- Accessible
- Interoperable
- Reusable

PresQT uses FAIR Evaluator, which is a collaboration with FAIRsharing.

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

Who talks about FAIR?

“Scientists, scholars, funders and publishers all over the world are coming together to enhance the value of research data by making them FAIR”.

Turning FAIR into reality, https://publications.europa.eu/s/m9l4

https://eresearchnz.figshare.com/articles/presentation/Towards_FAIR_principles_for_research_software/11929617/1
Towards FAIR principles for research software

Issue title: FAIR Data, Systems and Analysis
Guest editors: Paul Groth and Michel Dumontier
Article type: Position Paper

Authors: Lamprecht, Anna-Lena\textsuperscript{a} | Garcia, Leyla\textsuperscript{b} | Kuzak, Mateusz\textsuperscript{c} | Martinez, Carlos\textsuperscript{d} | Arcila, Ricardo\textsuperscript{e} | Martin Del Pico, Eva\textsuperscript{f} | Domínguez Del Angel, Victoria\textsuperscript{g} | van de Sandt, Stephanie\textsuperscript{h} | Ison, Jon\textsuperscript{i} | Martinez, Paula Andrea\textsuperscript{j} | McQuilton, Peter\textsuperscript{k} | Valencia, Alfonso\textsuperscript{m,n} | Harrow, Jennifer\textsuperscript{l} | Psomopoulos, Fotis\textsuperscript{p} | Gelpi, Josep LL\textsuperscript{q} | Chue Hong, Neil\textsuperscript{t} | Goble, Carole\textsuperscript{u} | Capella-Gutierrez, Salvador\textsuperscript{v}

Affiliations: [a] Utrecht University, The Netherlands. E-mail: a.l.lamprecht@uu.nl | [b] ZBMED Information Centre for Life Sciences, Germany. E-mail: igarcia@zbmed.de | [c] Netherlands eScience Center, The Netherlands | [d] Dutch Techcentre for Life Sciences, The Netherlands. E-mail: m.kuzak@esciencecenter.nl | [e] Netherlands eScience Center, The Netherlands. E-mail: c.martinez@esciencecenter.nl | [f] EMBL-EBI, UK. E-mail: arcila@ebi.ac.uk | [g] Barcelona Supercomputing Center (BSC), Spain. E-mail: eva.martin@bsc.es | [h] L’Institut Français de Bioinformatique (IFB), France. E-mail: victoria.dominguez@france-bioinformatique.fr | [i] CERN, Switzerland. E-mail: stephanie.van.de.sandt@cern.ch | [j] National Life Science Supercomputing Center, Technical University of Denmark, Denmark. E-mail: jison@cbs.dtu.dk | [k] National Imaging Facility, Australia. E-mail: p.martinez@uq.edu.au | [l] Oxford e-Research Centre, UK. E-mail: peter.mcquilton@oerc.ox.ac.uk | [m] Barcelona Supercomputing Center (BSC), Spain | [n] Institución Catalana de Recerca i Estudis Avançats (ICREA), Spain. E-mail: alfonso.valencia@bsc.es | [o] ELIXIR Hub, UK. E-mail: harrow@ebi.ac.uk | [p] Institute of Applied Biosciences, CERTH, Greece. E-mail: fpsom@certh.gr | [q] Barcelona Supercomputing Center (BSC), Spain | [r] University of Barcelona, Spain. E-mail: gelpi@ub.edu | [s] Software Sustainability Institute, UK | [t] EPC, University of Edinburgh, UK. E-mail: n.chuehong@software.ac.uk | [u] University of Manchester, UK. E-mail: carole.goble@manchester.ac.uk | [v] Barcelona Supercomputing Center (BSC), Spain. E-mail: salvador.capella@bsc.es

“the current status of the debate around FAIR and software, as basis for the development of community-agreed principles for FAIR research software in the future”

Bridging the Gap to Data and Software Sharing

Researchers

“the local academic community struggles to effectively manage its assets which manifested itself in a number of challenges, and as for researchers, they lacked storage capacity and data curation processes, and the institution lacked standard metadata and indexing technologies, as well as tools that would support the whole research workflow” - Digital Asset Strategy Committee, DigitalND, 2011
Researchers

“the local academic community struggles to effectively manage its assets which manifested itself in a number of challenges, and as for researchers, they lacked storage capacity and data curation processes, and the institution lacked standard metadata and indexing technologies, as well as tools that would support the whole research workflow” - Digital Asset Strategy Committee, DigitalND, 2011

Libraries

Typically, data curation happens retroactively, and as a result data is either not captured at all or available metadata is incomplete.
Current Lifecycle of Research Projects

- **New project**
- **Selection/development of tools**
- **Data assembling/creating**
- **Preservation of Data**
- **Reports**
- **Funding ends**

Work-intensive and too late in the lifecycle
Target Lifecycle of Research Projects

New project

Selection/development of tools

Assure quality of data

Data assembling/creating

Assure quality of data

Reports

Preservation of Data

EASY STEP!!! (ideally)

Funding ends
Bridging the Gap to Data and Software Sharing

Researchers

“the local academic community struggles to effectively manage its assets which manifested itself in a number of challenges, and as for researchers, they lacked storage capacity and data curation processes, and the institution lacked standard metadata and indexing technologies, as well as tools that would support the whole research workflow” - Digital Asset Strategy Committee, DigitalND, 2011

Libraries

Typically, data curation happens retroactively, and as a result data is either not captured at all or available metadata is incomplete.

Pressures from the Outside

“...digitally formatted scientific data resulting from unclassified research supported wholly or in part should be stored and publicly accessible to search, retrieve, and analyze.” - White House OSTP Public Access Memo, Feb. 2013
PresQT Stakeholder Engagement
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 Acknowledgements

Funded Subawardees
• Sheridan Libraries, John Hopkins University
• NDS
• UC San Diego Library
• HUBzero team
• Yale University Library

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

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

Will it help my job performance?  
Is it USEFUL?  
Is it better than the old way?

Easy to use?  
Easy to learn?  
Time-consuming?  
Efficient?

Can I do this? Do I have knowledge? Support? Resources? Does it fit in with my work style?

Unified Theory of Acceptance and Use of Technology (UTAUT) Venkatesh et al, 2003
Collaborative Effort

Concept

- **not** standalone solutions
- partner systems and services easily integrable via RESTful APIs and services
- user-centered open design and collaborative development
Current PresQT Ecosystem
Current PresQT Ecosystem

Repositories

- Zenodo
- GitLab
- CurateND
- GitHub
- figshare

Users
Current PresQT Ecosystem

Science Gateways

- Zenodo
- GitLab
- CurateND
- GitHub
- figshare
- WHOLE TALE
- hubzero
- OSF

Hesburgh Libraries

University of Notre Dame
Current PresQT Ecosystem

- Zenodo
- GitLab
- WHOLE TALE
- hubzero
- CurateND
- GitHub
- figshare
- Users
- Services
- EaaSI
- SciGraph
- FAIRsharing.org
- FAIRshake

Hesburgh Libraries

UNIVERSITY OF NOTRE DAME

CRC CENTER FOR RESEARCH COMPUTING
Ecosystem without PresQT

zenodo
GitLab
WHOLE TALE
hubzero
OSF
CurateND
GitHub
figshare
EaaS
SciGraph
FAIRsharing.org
FAIRshake

Hesburgh Libraries

UNIVERSITY OF NOTRE DAME
CRC CENTER FOR RESEARCH COMPUTING
Ecosystem without PresQT

zenodo

GitLab

WHOLE TALE

hubzero

CurateND

GitHub

figshare

EaaS

SciGraph

FAIRsharing.org

FAIRshake

Hesburgh Libraries
Ecosystem without PresQT

zenodo

曲率ND

figshare

GitHub

GitLab

WHOLE TALE

hubzero

OSF

EaaSi

SciGraph

FAIRsharing.org

FAIRshake
Ecosystem without PresQT
PresQT Connects and Enhances

- Zenodo
- GitLab
- Whole Tale
- Hubzero
- OSF
- Curate ND
- GitHub
- Figshare
- EaaSI
- SciGraph
- FAIRshake
- FAIRsharing.org

Hesburgh Libraries
PresQT Set of Services

- Configure additional partner systems via JSON and Python functions
- Check fixity via hash algorithms
- Add metadata via JSON
- Transfer in BagIt format
- Enhance keywords
- Check available tools via EaaSI
- Check FAIRness of research objects
PresQT Keyword Enhancement
PresQT Keyword Enhancement

Github

SciGraph

[cat, dog, egg]

OSF

airplane

Target Internal Keywords

cat, dog

cat, feline, dog, canine, egg, scrambled

cat, feline, dog, canine, egg, scrambled

Final Keyword States

FTS Metadata Keywords

cat, dog, egg

cat, feline, dog, canine, egg, scrambled

cat, feline, dog, canine, egg, scrambled

airplane, cat, feline, dog, canine, egg, scrambled

jet, airplane, cat, feline, dog, canine, egg, scrambled

jet
# Current Integrations

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>OSF</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>[Github, CurateND, Zenodo, GitLab, FigShare]</td>
<td>[Github, Zenodo, GitLab, FigShare]</td>
<td>[sha256, md5]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>curateND (Fedora)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>TBA</td>
<td>TBA</td>
<td>[md5]</td>
<td>TBA</td>
<td></td>
</tr>
<tr>
<td>Github</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>[OSF, CurateND, Zenodo, GitLab, FigShare]</td>
<td>[OSF, Zenodo, GitLab, FigShare]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zenodo</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>[OSF, Github, CurateND, GitLab, FigShare]</td>
<td>[OSF, Github, GitLab, FigShare]</td>
<td>[md5]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GitLab</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>[OSF, Github, CurateND, Zenodo, FigShare]</td>
<td>[OSF, Github, Zenodo, FigShare]</td>
<td>[sha256]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FigShare</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>[OSF, Github, CurateND, Zenodo, GitLab]</td>
<td>[OSF, Github, GitLab, Zenodo]</td>
<td>[md5]</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

https://presqt.readthedocs.io/en/latest/
FAIR Principles for Science Gateways

PresQT’s approach currently – open for discussion!

1. Testing of FAIRness of source code
   Maturity indicator tests via source code in repository

2. Testing of FAIRness of data objects/tools in science gateways
   Maturity indicator tests via science gateway allowing access to data objects

Suggestions how to improve FAIRness based on connected services!
PresQT
Preservation Quality Tools

OSF Resources
Search OSF by Title

CREATE NEW PROJECT

Demonstrating PresQT Services for FAIR Software and Data Preservation

Resource Details

Kind
container

Id
3google

Title
Demonstrating PresQT Services for FAIR Software and Data Preservation

Date Created
2020-11-19T19:53:35Z

Date Modified
2020-11-19T20:13:55Z

Kind Name
project

Identifier
10.17605/OSF.IO/3GNQD

Hashes
md5: null
sha256: null

Available Connections

Development Partners

UNIVERSITY OF NOTRE DAME

Hesburgh Libraries

CRC
CENTER FOR RESEARCH COMPUTING

DEMO

presqt-prod.crc.nd.edu/ui/
Demonstrating PresQT Services for FAIR Software and Data Preservation

Resource Details

Kind
- container
- project

Identifier
- 10.17605/OSF.IO/9GNQD

Date Created
- 2020-11-19T19:53:35Z

Date Modified

Extra
- category: project
- fork: false
- current_user_is_contributor: true
- preprint: false
- current_user_can_delete: true
- custom_citation: Johnson, K. P., Moyer, N. K. "Demonstrating PresQT Services for FAIR Software and Data Preservation" CHI Fall 2020 Virtual Membership Meeting. (Dec 1 2020); doi:10.17605/OSF.IO/9GNQD, Available at osf.io/9gnqd
- collection: false
- public: true
- subjects: []
- registration: false
- current_user_can_comment: true
- wiki_enabled: true
- node_license: ['copyright_holders': ['", "year": 2020']]
- tags: []
- size: null
- hashes: 
  - md5: null
  - sha256: null

PresQT: Recap and Today’s Goals

PresQT: Workshop Goals - High Level Design - Timeline of the Project

Available Connections

Development Partners

presqt-prod.crc.nd.edu/ui/
Transfer Resource: PresQT: Workshop Goals - High Level Design - Timeline of the Project

1. Select destination target

2. Input destination target token

3. Select resource to transfer to or select nothing to transfer as a new project

4. Select the action to occur when a duplicate resource is found

5. Select the keyword action to occur

6. Keywords

7. FAIRshare Evaluator Opt-In

8. Transfer Agreement

9. Email Opt-In

10. Transfer Results
Some tests may be failing because the transferred resource URL is too new.

- FAIR Metrics Gen2: Unique Identifier
- Found an identifier of type 'uri'
- FAIR Metrics Gen2: Identifier Persistence
- The metadata GUID does not conform with any known permanent-URL system.
- FAIR Metrics Gen2: Structured Metadata
- FAIR Metrics Gen2: Grounded Metadata
- FAIR Metrics Gen2: Data Identifier Explicitly In Metadata
- FAIR Metrics Gen2: Metadata Identifier Explicitly In Metadata
- FAIR Metrics Gen2: Searchable in major search engine
- FAIR Metrics Gen2: Uses open free protocol for metadata retrieval
- FAIR Metrics Gen2: Metadata Persistence
- FAIR Metrics Gen2: Data Knowledge Representation Language (strong)
- FAIR Metrics Gen2: Metadata uses FAIR vocabularies (strong)
- FAIR Metrics Gen2: Metadata Includes License
Acknowledgements

Collaboration at ND between Hesburgh Libraries and CRC, especially
Co-Pis
• John Wang, Natalie Meyers, Rick Johnson
CRC
• Noel Recla, Brett Fox, Justin Branco
Hesburgh Libraries
• Miranda VanNevel, Don Brower, Mark Suhovecky, Ian Alford, Micala Narlock

Partners and collaborators mentioned above
Thank you!

presqt-contact-list@nd.edu
https://presqt.readthedocs.io/en/latest/
https://presqt-prod.crc.nd.edu/ui/