

The HEAP FAIR Toolbox

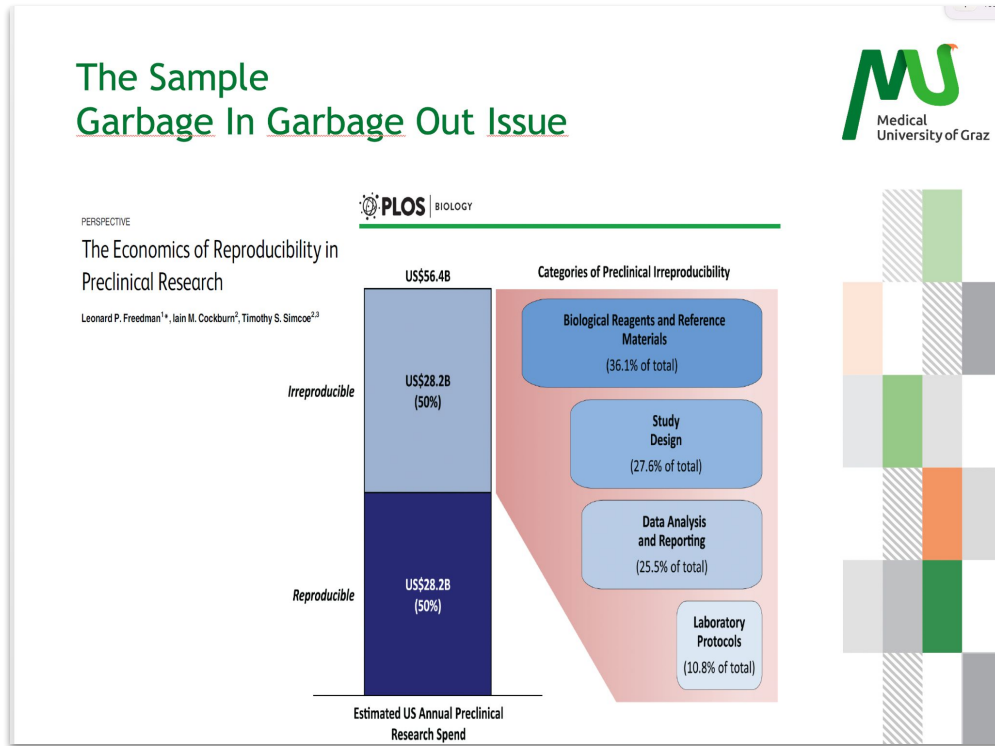
Heimo Müller - Medical University Graz



European Human **Exposome** ***NETWORK***



Need for High Quality Data



Research Data Reproducibility

Medical University of Graz

Reliability of 'new drug target' claims called into question

Bayer halts nearly two-thirds of its target validation projects because in-house experimental findings fail to match up with published literature claims, finds a first-of-a-kind analysis on data irreproducibility.

Asker Mufson
An outspoken industry risk analyst... that at least 50% of published studies from academic laboratories cannot be repeated in an industrial setting, says venture capitalist Bruce Booth in a recent blog post. A first-of-a-kind analysis of Bayer's internal efforts to validate 'new drug target' claims now not only supports this view but suggests that 50% may be an underestimate: the company's in-house experimental data do not match literature claims in 65% of deep questions about whether we can really believe the literature, or whether we have to go back and do everything on our own."

For the non-peer-reviewed analysis, Khursi Anasikh, head of Target Discovery at Bayer, and his colleagues looked back at 47 target validation projects, covering the majority of Bayer's work in oncology, women's health and cardiovascular medicine over the past 4 years. Of those, results from internal experiments matched up with the published findings in 31%.

These included inability to reproduce over-expression of certain genes in specific human types and decreased cell proliferation via functional inhibition of a target using DNA interference.

Irreproducibility was high both when Bayer scientists applied the same experimental procedures as the original researchers and when they adapted their approaches to internal needs (for example, by using different cell lines). High impact journals did not seem to be any more reliable than lower impact journals.

NATURE REVIEWS | DRUG DISCOVERY
VOLUME 10 | SEPTEMBER 2011 | 643

IS THERE A REPRODUCIBILITY CRISIS?
A Nature Reviews Drug Discovery analysis found that how researchers view the 'cost' of making science and what they think will help.

M Baker & D Penry 454 | NATURE | VOL 532 | 24 MAY 2016

The Economist
Britain's angry white men
How to do a nuclear deal with Iran
Investment tips from Nobel economists
Junk bonds are back
The meaning of Sachin Tendulkar

HOW SCIENCE GOES WRONG.
99
Einsteinium

Kurt Zatloukal, 01.06.2022



European Human Exposome **NETWORK**



G8 Science Ministers Statement (2013)

‘Open scientific research data should be easily discoverable, accessible, assessable, intelligible, useable, and wherever possible interoperable to specific quality standards.’

13 June 2013 <https://www.gov.uk/government/news/g8-science-ministers-statement>



European Human **Exposome NETWORK**



Some History

- Lorentz workshop Leiden, Data FAIRport initiative, 2014
- Wilkinson et al, (2016) 'The FAIR Guiding Principles for scientific data management and stewardship', Scientific Data 3:160018
- EOSC Initiative, 2015
- EOSC Association, 2020
- EOSC Tripartite Event, Prag 2022

European Commission, Directorate-General for Research and Innovation, *Turning FAIR into reality : final report and action plan from the European Commission expert group on FAIR data*, Publications Office, 2018, <https://data.europa.eu/doi/10.2777/1524>



European Human **Exposome NETWORK**



Reality Check Ahead

- High quality (meta) data needs a lot of effort
- It will only pay off in the (far) future



Push Factors

- Political
- Research funders
- Journals & publishers
- Science community
- Ressource Savings
- Team Support
- Visibility
- Early Pay-Off

Recomentations & Regulations

Support and Tools



Compiling (Open Source) Tools

- Data Management and Analysis
- Sample Management
- Data Repository Solutions

- Existing solutions **missing** mainly **interoperability** and **provenance** support

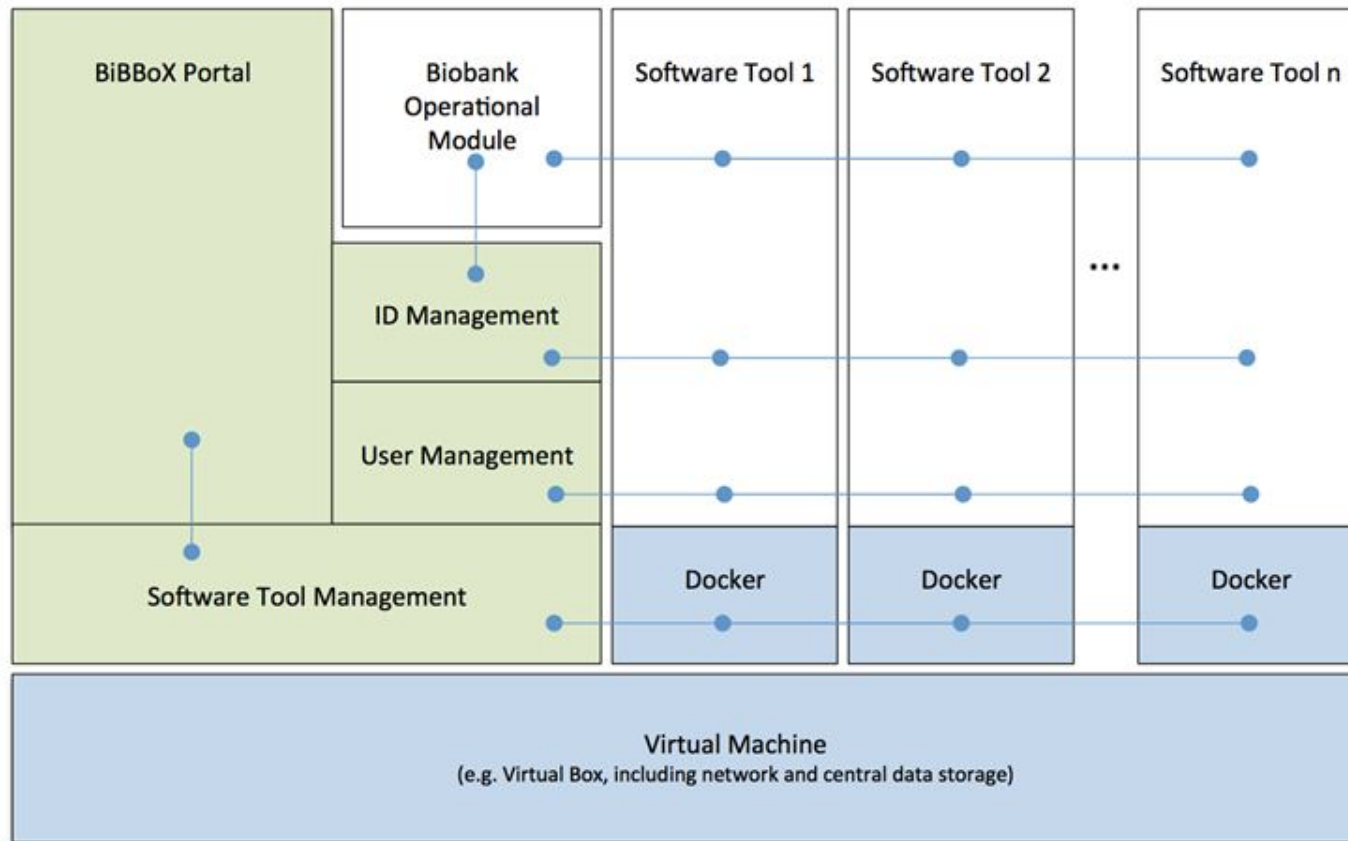


BIBBOX - The History of the FAIR Toolbox

- **Basic Infrastructure Building BOX**
- Development started as part of the B3Africa project
 - Backend: Liferay
 - Frontend: React



BIBBOX System Architecture



FAIR Toolbox Development

- FAIR Toolbox demanded complete redesign of the original BIBBOX
 - Core principle kept - complete rewritten code
 - Backend: Python (Flask)
 - Frontend: Angular
 - App Store & Docker Hub updated to Version 4.0
- Added **FDP** & **PAP** to each application



FDP - What is it?

- FAIR data point
- Resource level description (DCAT) of
 - Catalog
 - Dataset
 - Distribution
- Contact person, license, access link etc.
- Existing solution integrated in the FAIR Toolbox
<https://github.com/FAIRDataTeam/FAIRDataPoint.git>



PAP - What is it?

- Provenance Access Point is being developed within HEAP
- Based on the upcoming ISO Standard 23494 -
“Biotechnology – Provenance information model for biological specimen and data”
- Machine readable provenance information of a record

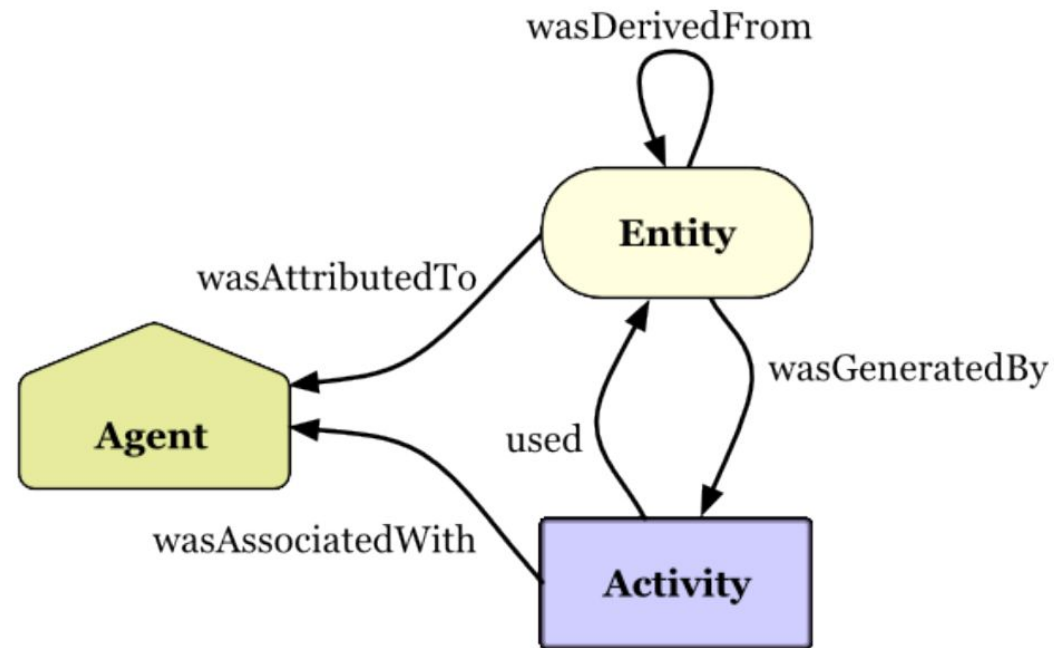
Rudolf Wittner, et al. ISO 23494: Biotechnology – Provenance Information Model for Biological Specimen And Data,
https://link.springer.com/chapter/10.1007/978-3-030-80960-7_16



European Human **Exposome NETWORK**



W3C PROV model

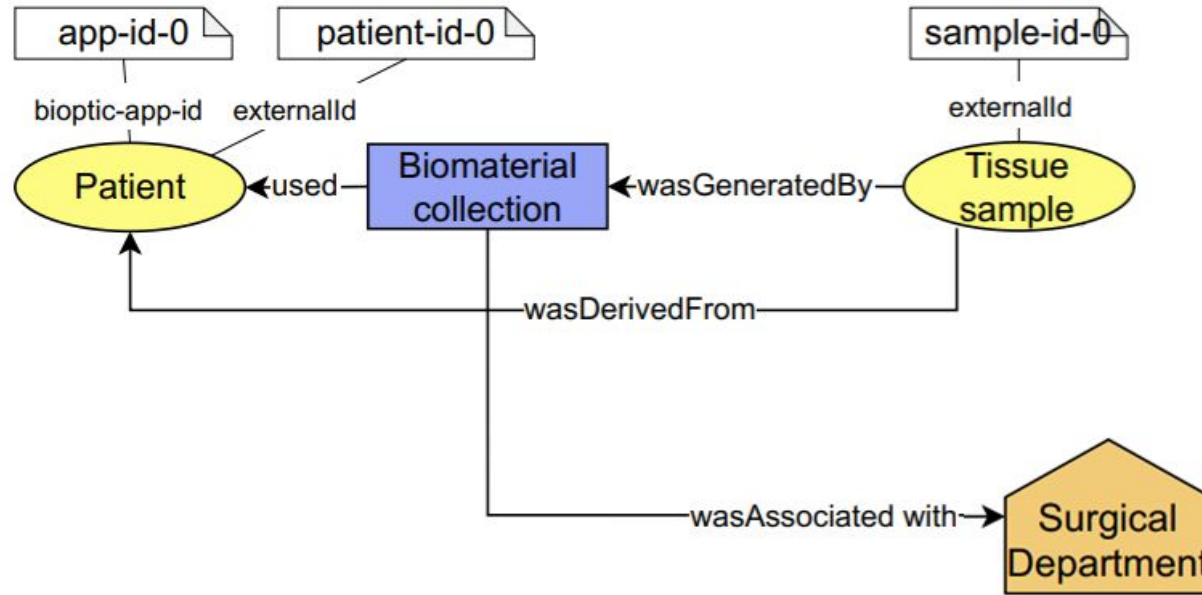


<https://www.w3.org/TR/prov-overview/>



European Human **Exposome NETWORK**

ISO 23494 example



```

:sampleSource
  a prov:Entity
  sprovdonorID "someDonorID";
  sprovdhealthStatusICD10 "C50.1";
  sprovdTreatmentDocumentation treatmentEntity;
  sprovdstartOfIchemia "2011-07-16T01:52:02Z"^^xsd:dateTime;

:Haahashtari
  a foaf:Person, prov:Agent, sprovdMedical;
  foaf:givenName "Haahashtari";
  foaf:mbox <mailto:haahashtari@example.org>;
  prov:actedOnBehalfOf :samplingContext;

:sampling
  a prov:Activity
  prov:wasAssociatedWith :Haahashtari;

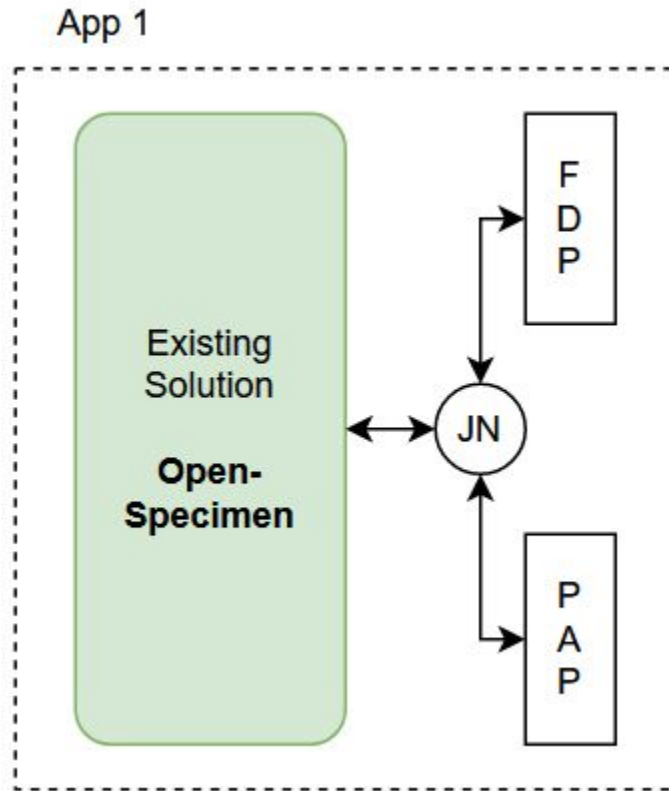
:primarySample
  a prov:Entity
  prov:generatedAtTime "2011-07-16T01:52:02Z"^^xsd:dateTime;
  prov:wasGeneratedBy :sampling;

:sampleSourceDescription
  a prov:Activity
  prov:generatedAtTime "2011-07-16T01:52:02Z"^^xsd:dateTime;
  prov:wasAssociatedWith :Haahashtari;

:samplingDocumentation
  a prov:Entity
  prov:wasGeneratedBy :sampleSourceDescription;
  sprovddocumentation :samplingDocumentationEntity_FHIR_link;
  
```



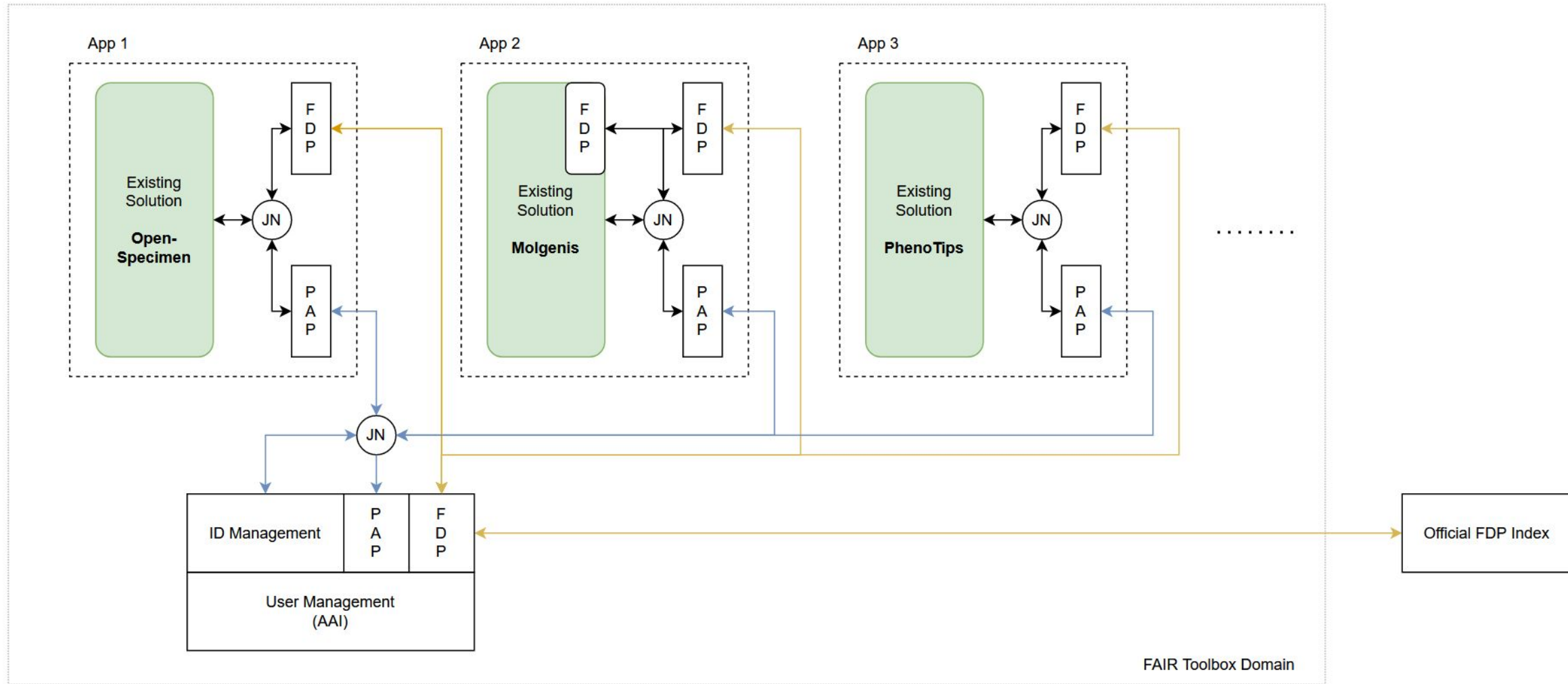
FAIRified App



- FAIR Data Point
 - Spring Boot
 - BlazeGraph DB
- Provenance Access Point
 - Neo4j DB
- Interface connecting App, FDP, PAP



FAIR Toolbox

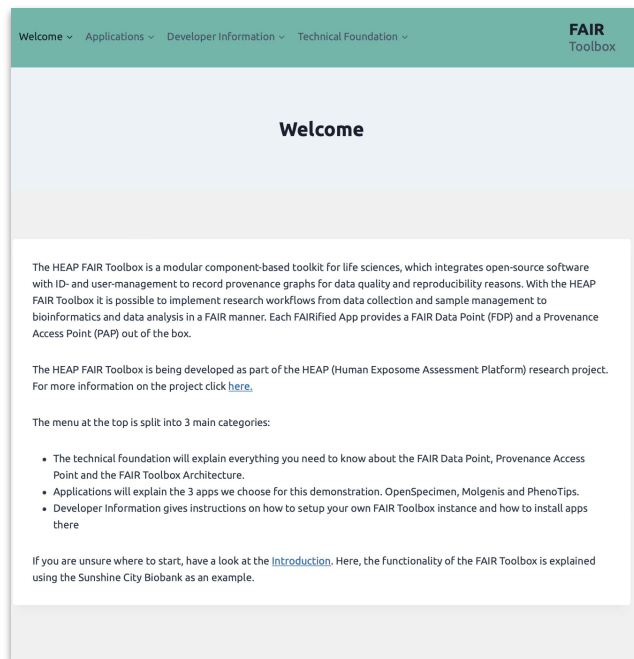


FAIR Toolbox DEMO

- 3 Apps for demonstration
 - Sample Management (OpenSpecimen)
 - Patient Registry and Pedigrees (Phenotios)
 - Omics Data Management (Molgenis)
- Master FDP
- <http://fairtoolbox.bibbox.org/>



FAIR Toolbox DEMO



Welcome

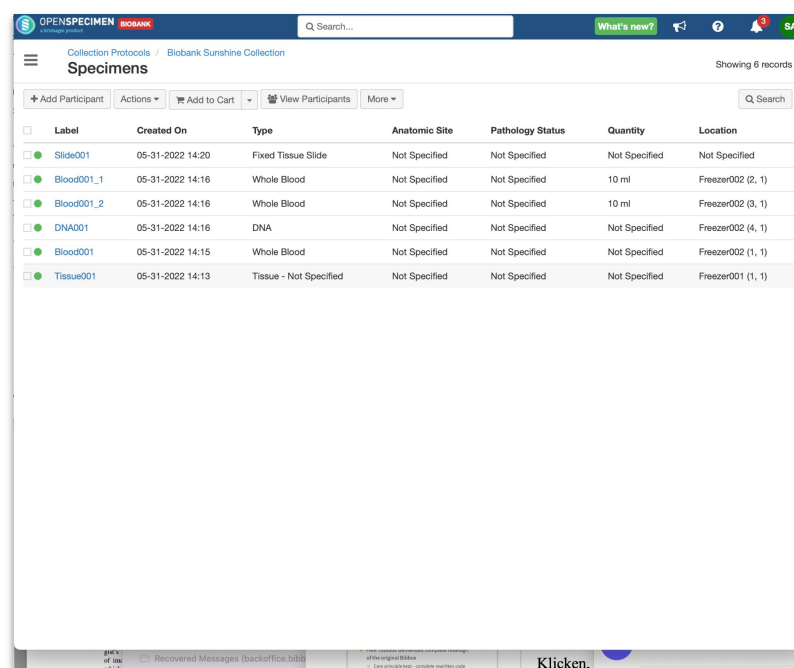
The HEAP FAIR Toolbox is a modular component-based toolkit for life sciences, which integrates open-source software with ID- and user-management to record provenance graphs for data quality and reproducibility reasons. With the HEAP FAIR Toolbox it is possible to implement research workflows from data collection and sample management to bioinformatics and data analysis in a FAIR manner. Each FAIRified App provides a FAIR Data Point (FDP) and a Provenance Access Point (PAP) out of the box.

The HEAP FAIR Toolbox is being developed as part of the HEAP (Human Exposome Assessment Platform) research project. For more information on the project click [here](#).

The menu at the top is split into 3 main categories:

- The technical foundation will explain everything you need to know about the FAIR Data Point, Provenance Access Point and the FAIR Toolbox Architecture.
- Applications will explain the 3 apps we choose for this demonstration. OpenSpecimen, Molgenis and PhenoTips.
- Developer information gives instructions on how to setup your own FAIR Toolbox instance and how to install apps there

If you are unsure where to start, have a look at the [Introduction](#). Here, the functionality of the FAIR Toolbox is explained using the Sunshine City Biobank as an example.

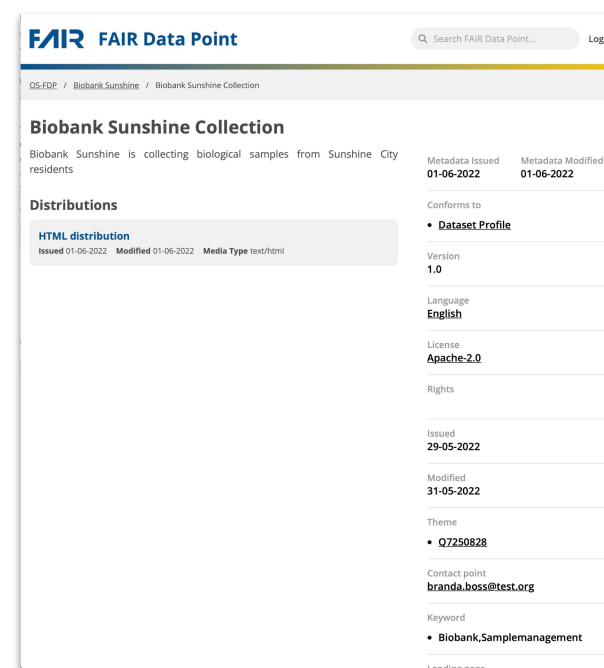


OPENSPECIMEN BIOBANK

Collection Protocols / Biobank Sunshine Collection

Showing 6 records

| Label | Created On | Type | Anatomic Site | Pathology Status | Quantity | Location |
|------------|------------------|------------------------|---------------|------------------|---------------|------------------|
| Slide001 | 05-31-2022 14:20 | Fixed Tissue Slide | Not Specified | Not Specified | Not Specified | Not Specified |
| Blood001_1 | 05-31-2022 14:16 | Whole Blood | Not Specified | Not Specified | 10 ml | Freezer02 (2, 1) |
| Blood001_2 | 05-31-2022 14:16 | Whole Blood | Not Specified | Not Specified | 10 ml | Freezer02 (3, 1) |
| DNA001 | 05-31-2022 14:16 | DNA | Not Specified | Not Specified | Not Specified | Freezer02 (4, 1) |
| Blood001 | 05-31-2022 14:15 | Whole Blood | Not Specified | Not Specified | Not Specified | Freezer02 (1, 1) |
| Tissue001 | 05-31-2022 14:13 | Tissue - Not Specified | Not Specified | Not Specified | Not Specified | Freezer01 (1, 1) |



FAIR FAIR Data Point

Biobank Sunshine Collection

Biobank Sunshine is collecting biological samples from Sunshine City residents

Metadata Issued: 01-06-2022
Metadata Modified: 01-06-2022

Distributions

HTML distribution
Issued 01-06-2022 Modified 01-06-2022 Media Type text/html

Conforms to

- Dataset Profile

Version: 1.0

Language: English

License: Apache-2.0

Rights

Issued: 29-05-2022

Modified: 31-05-2022

Theme: Q7250828

Contact point: branda.boss@test.org

Keyword: Biobank, Samplemanagement

<http://fairtoolbox.bibbox.org/>



European Human Exposome **NETWORK**



Acknowledgments



The HEAP project has received funding from the European Union's Horizon 2020 research programme under grant agreement No 874662.



European Human Exposome **NETWORK**

