We have access to:

Data repository from different set of experiment and data (phenotypic data, engrafted mouse, expression dataset...)

Resources:

• PDX Finder

• Mouse Tumor Biology (MTB) Database

Leverage on existing datasets to improve FAIRNESS of the dataset in order to answer a use case.

General USE CASE:
Identifying mutations that are lost through passages associate those lost mutations with protein domains if present

Specifically look at one gene:
Given a Gene of interest, (e.g. TP53), retrieve all the patients (models) that have variations present at this genomic location and with this genomic range obtain the protein domain information

Hackathon Plan

- FAIR evaluation Run purple polar bear to test FAIRness

- Datasets Extraction

- Meta-data annotation

- Connect the Datasets

- Use case application

- Mash-up with cBioPortal (for next time)

Average Score

How FAIR is your dataset?

26%

Findable	**Accessible**	**Interoperable**	**Reusable**
37%	**50%**	**0%**	**16%**
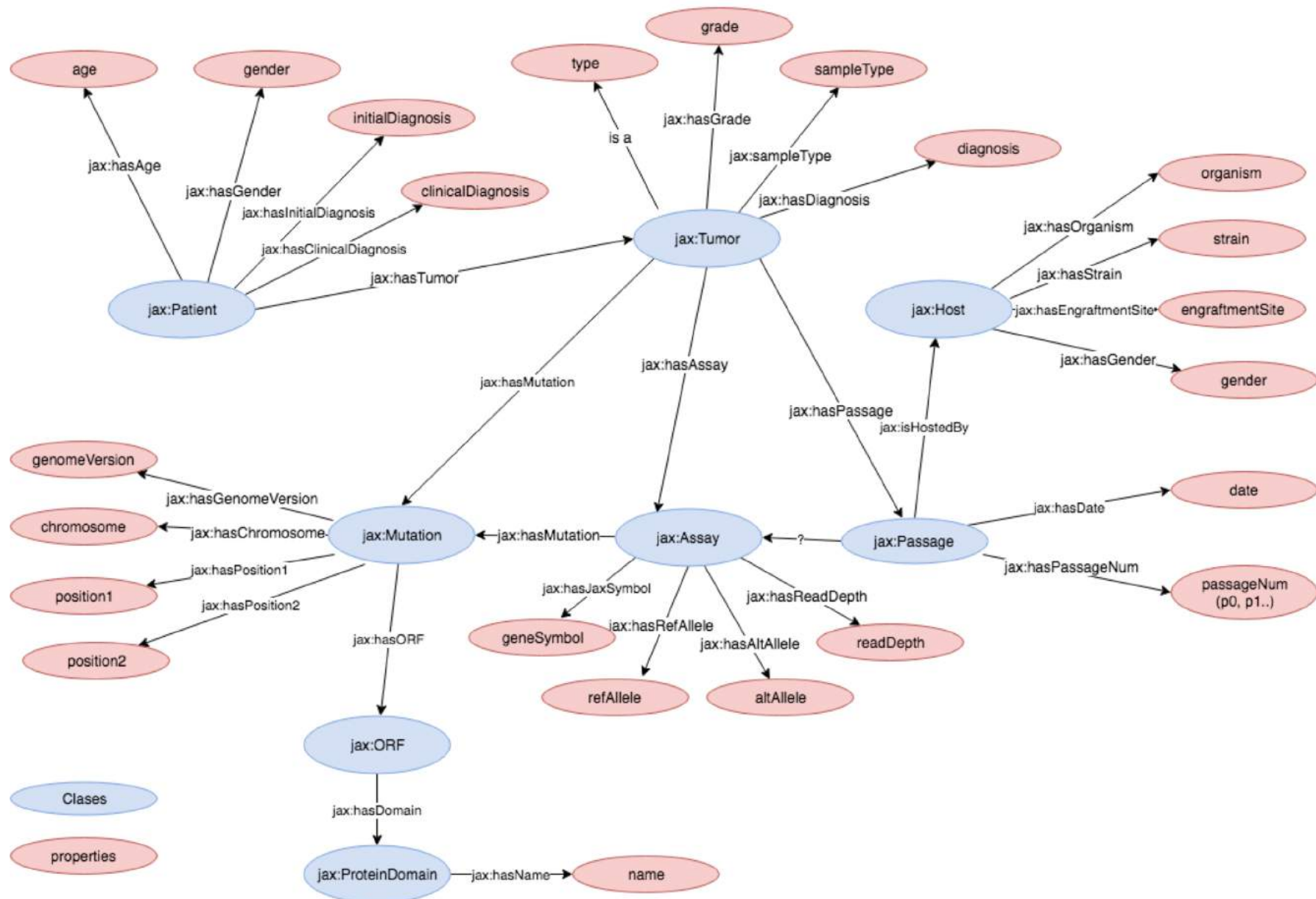
BioIT World 2018 May 14-15 Hackathon
JAX Data Team

Base URL for PDX JSON web services

Current Services:
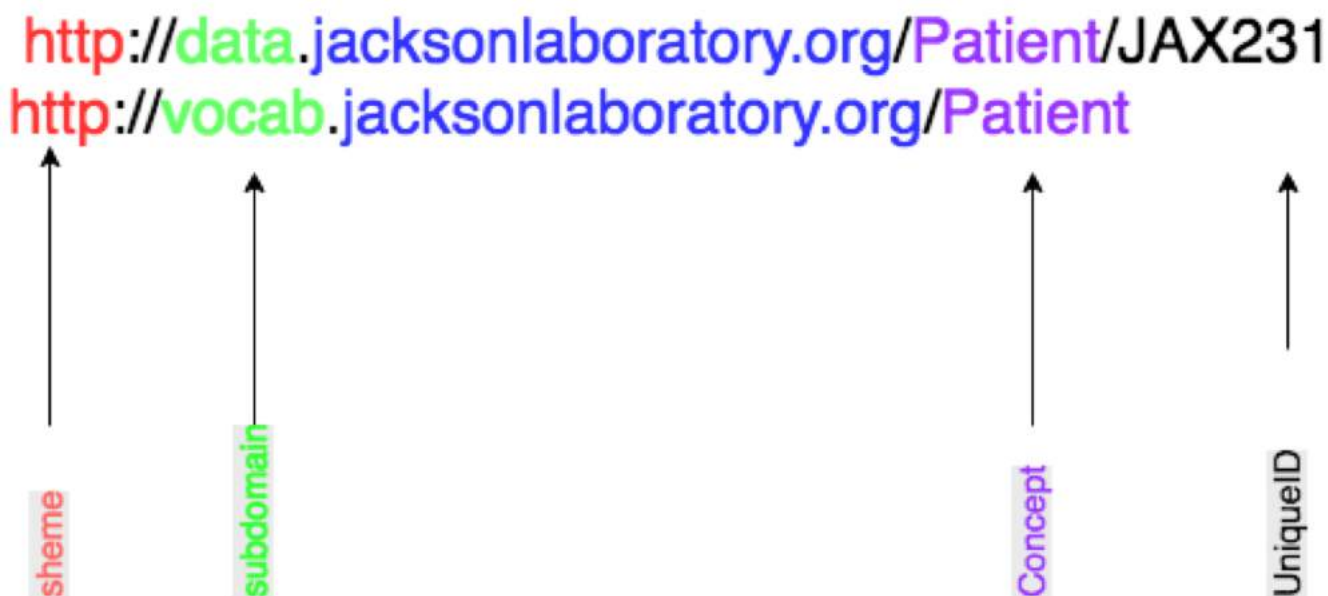/JSONData.do?allModels -- > all
/JSONData.do?modelVariation=<modelID>
/JSONData.do?modelHistology=<modelID>
/JSONData.do?modelCNV=<modelID>
/JSONData.do?modelExpression=<modelID>

Global Identifiers @JAX

http://data.jacksonlaboratory.org/Patient/JAX231
http://vocab.jacksonlaboratory.org/Patient

sheme   subdomain   Concept   UniqueID

F1 Q1: Do you have a URL to a registered scheme that defines the globally-unique structure of the identifier(s) for your digital resource?
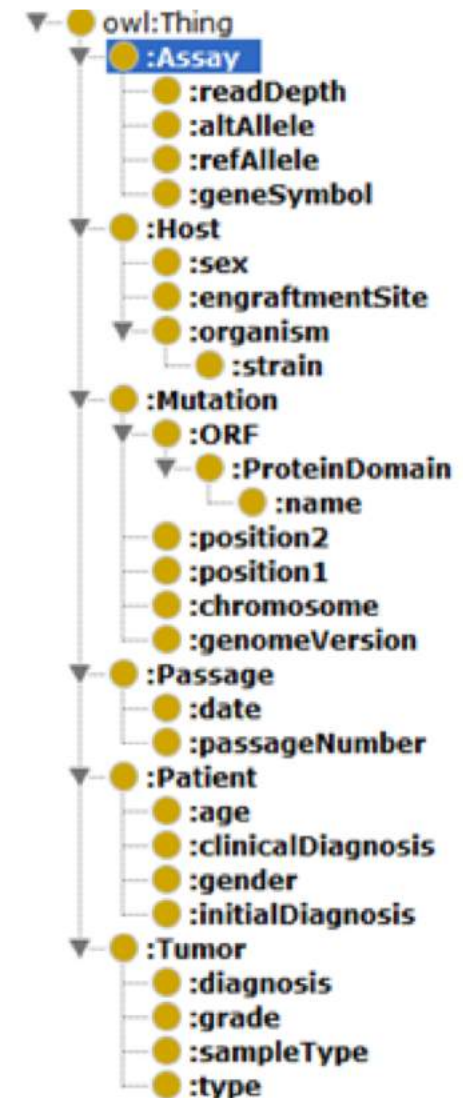
OWL model for schema – someone trying to use the data could look at that OWL definition

F1 Q2: Do you have a URL to a document that defines the persistence policy of your identifier(s)?
Action: JAX puts up a webpage or word document.

F4 Q8: Is there a query that when executed will discover your RESOURCE ID (found in the first page of the search)?
Action: make this a RESTful service to enable machine discovery

```
▼ owl:Thing
  ▼ :Assay
      :readDepth
      :altAllele
      :refAllele
      :geneSymbol
  ▼ :Host
      :sex
      :engraftmentSite
    ▼ :organism
        :strain
  ▼ :Mutation
    ▼ :ORF
      ▼ :ProteinDomain
          :name
      :position2
      :position1
      :chromosome
      :genomeVersion
  ▼ :Passage
      :date
      :passageNumber
  ▼ :Patient
      :age
      :clinicalDiagnosis
      :gender
      :initialDiagnosis
  ▼ :Tumor
      :diagnosis
      :grade
      :sampleType
      :type
```

A2 Q14: Do you have a URL to a metadata longevity plan?
Action: Develop and publish a metadata longevity plan.

I1 Q15: Do you have a URL to the specification of a knowledge representation language?

We have defined the structure for a URI for the use case based on the data model in RDF. JAX has a namespace with a unique identifier that can map to the representation of the data.

I2 Q16: Do you have one or more (max 3) URLs representing the vocabularies used within the (meta)data that is returned by resolving the RESOURCE ID?

Action: e.g. vocab.jacksonlabs.org

I3 Q17: Do you have a URL to a formal Linkset (defined at: https://www.w3.org/TR/void/#linkset) that describes at least a portion of the content at RESOURCE ID?

Action: figure out what this means

R1.2 Q19: Do you have URLs (maximum 3) to the vocabularies being used to describe the provenance of the content resolved from RESOURCE ID (be that data, or metadata; e.g. http://purl.org/dc/terms/)?

Provenance available, but not resolved by RESOURCE ID.

Action: Create a service that returns this information based on a RESOURCE ID query.

R1.2 Q20: Do you have a URL (maximum 3) for the vocabularies being used to describe the domain information of the content resolved from RESOURCE ID (be that data, or metadata)?

Vocabulary creation as above.
Action: Create a service that makes the vocabulary discoverable by RESOURCE ID.

R1.3 Q21: Do you have a URL that represents the certification from a recognized authority in your community or domain, indicating that the content of RESOURCE ID is compliant with the standards of your community?

Action: To be investigated

Improved Score by more than 2X!

Average Score

How FAIR is your dataset?

54%

Findable
83%

Accessible
66%

Interoperable
33%

Reusable
33%

<u>Links to Hackathon code</u>

https://github.com/BioITHackathons/JaxTeam

https://groups.google.com/forum/#!forum/jaxteambioit

Thank you!

BioIT World 2018 May 14-15 Hackathon
JAX Data Team

Team Members:

Olaitan Awe: laitanawe@gmail.com
John Wright: john.wright@jax.org
Marjorie Wilke: margiewilke@gmail.com
Rajan Desai: rajan1106@yahoo.com
Anne Deslattes Mays: adeslat@jax.org
Chris Evelo: chris.evelo@maastrichtuniversity.nl
Kjiersten Fagnan: kmfagnan@lbl.gov
Berenice Wulbrecht: berenice@ontoforce.com

Thank you!