

FAIR Data APIs in the FISH (FAIR in Vivo Data Sharing) Platform

SEMANTiCS Vienna 2022 - Onto4FAIR Workshop

Felix Schwagereit, Martin Romacker, Fabien Richard,
Robert Trypuz, Thomas Liener, Olivier Roche

Roche Pharma Research and Early Development, Data & Analytics
Roche Innovation Center Basel, Switzerland

F.A.I.R - what does it mean ?

FAIR: Stands for

- **F**indable: To be Findable any Data Object should be uniquely and persistently identifiable
- **A**ccessible: Data is Accessible in that it can be always obtained by machines and humans
- **I**nteroperable: To be machine readable and based on shared metadata
- **R**eusable: To be Reusable Data Objects should be well described to be automatically linked with other data and refer to their sources

Why do we care about FAIR?



- Efficient organization, storage, retrieval and access of data is crucial to our business success as a Pharma and Diagnostics company
 - Enabler for generation of new insights from large data sets
 - Estimated [1] €10.2 billion is lost every year in research alone due to insufficient data management
- As a globally operating organization, we created or internalized a large amount of data in the past
- Every year our data output has been growing exponentially in terms of volume and variety.
 - Capability to properly master our value-driven data management processes is necessary for our future business success evolving towards a data-centric company.
- Expected cost savings and gains in productivity enabled by proper data management are crucial to realize Roche's 10-years-ambition - to create at least twice the benefits for patients while reducing costs to society by 50% or more

[1] Cost of not having FAIR research data <https://doi.org/10.9933>

Approach

- Apply best data practices in order to build FAIR(er) applications:
example FISH
- Implement best practices in a tool (RTS) that
 - Integrates them by leveraging on dependencies and existing information
 - Hosts reference data
 - Hides technical details behind a user interface → lowers the hurdle for practical implementation of FAIR practices
- Main functionalities
 - URI namespace service
 - Terminology management
 - Reference Model management
 - Application-specific model management
 - ISON LD interface specification for FAIR data APIs

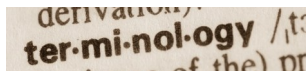
RTS FAIR - Informatics & Organizational Capabilities

Findability

Accessibility

Interoperability

Reusability



Terminology



Identifier Resolution



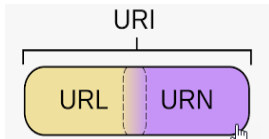
Data Catalog



Semantic Search



Standard API



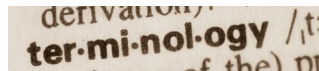
Resources (URI)



Metadata (DCAT)



Data Sharing/
Data Protection



Terminology



Metadata Registry



Business Glossary



Data Standards



Data Quality



Minimal Models



Data Validation

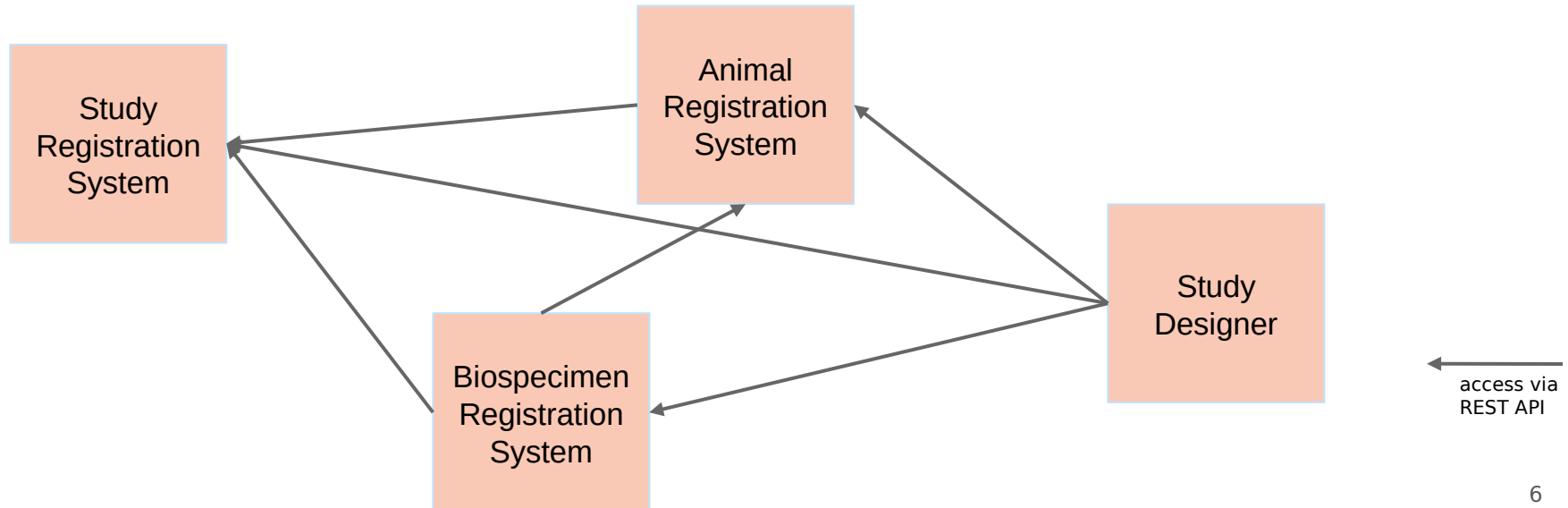


Data Governance

FAIR in vivo data sharing platform (FISH)

System topology

- Component-based Architecture
- RESTful APIs
- Relational DBs per component



Use Open Standards

Recommendations for content and code

- Open and well-defined standards for file formats and knowledge representation
- Eliminates risk of a vendor lock-in
- Secures future reusability and interoperability
- Notable Standards
 - HTTPS as secure content transfer protocol and RESTful API
 - OpenAPI for API documentation
 - JSON / JSON-LD for API file format
 - RDF/OWL/SHACL for metadata model definition
 - SPARQL as query language
 - Basic Formal Ontology (BFO)
 - Dublin core vocabulary for generic metadata
 - RTS for terminology management and as model repository
 - ...

Examples for FAIR URIs

Taken from pREDi FISH Study Registration System

Digital Object	Scheme	Example
Study	https://id.roche.com/a2/{studyId}	https://id.roche.com/a2/32
Study Document Protocol	https://id.roche.com/a2/{studyId}/document	https://id.roche.com/a2/32/document
Local Descriptor	https://id.roche.com/a2/{studyId}/localDescriptor/{descriptorId}	https://id.roche.com/a2/32/localDescriptor/2

Registration System: id.roche.com

- Stable URIs by decoupling URIs from application
- Implementation of best practises and learnings from past projects
- Rolled out Q3 2021
- short URIs
- uses opaque id part as combination of two characters/digits from Base 33
- Self registration for solution architects and software engineers
- lean registration process due to no need to governan speaking ids → no need to align between all similar systems on a

pREDi FISH Component	Namespace
Study Registration System	https://id.roche.com/a2/
Animal Registration System	https://id.roche.com/a3/
Biospecimen Registration System	https://id.roche.com/a4/
Formulation Registration System	https://id.roche.com/a5/
Study Designer	https://id.roche.com/a6/

RTS as registration system for terminologies

Recommendation for Metadata guidelines and conventions

- Inventory of harmonized terminologies
- Supports frictionless data integration
- Is prerequisite for a fully harmonized and interoperable data landscape
- Prospectively align application specific terminologies with existing terminologies and concepts
- Formal and machine-readable terminology representation scales better than documentation only in Github or Confluence

Terminology Management Application

RTS Curation Client

The screenshot displays the Terminology Management Application interface, divided into two main sections: the **Application Navigator** on the left and the **Application Concept Entity Properties** on the right.

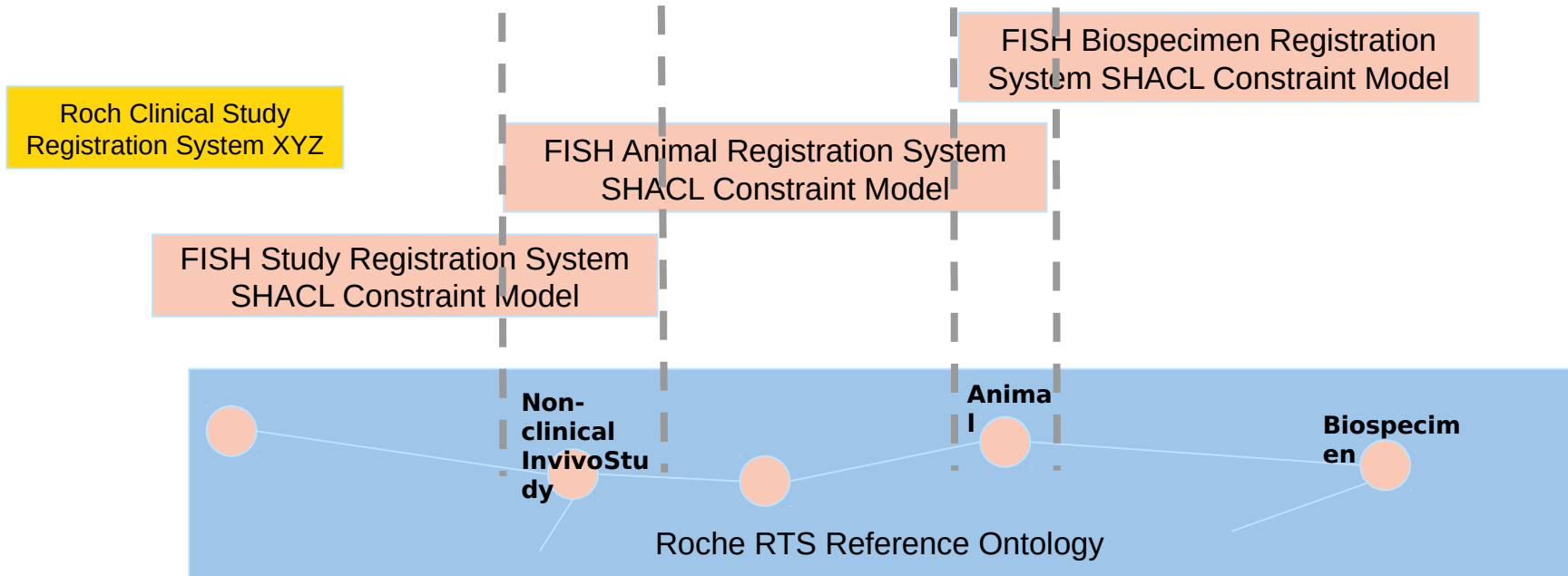
Application Navigator: A tree view showing the hierarchy of concepts. The **Published** status is highlighted under the **Data record status App Term** category. A callout box labeled **Concept** points to the **Published** status.

Application Concept Entity Properties: This section displays the details for a selected concept. It includes:

- Information:** A form with fields for Label (Published), Identifier (ROX1446033494572), Local Status (Active), and Status (Active). A callout box labeled **GUPRI** points to the Identifier field.
- Definition:** A text area containing the definition: "Data record that is released and available to a large community (RTS)."
- Comment:** An empty text area for additional notes.
- Mapping Label:** A callout box labeled **Mapping Label** points to the **Mapping Label** tab in the navigation bar.
- Navigation:** Tabs for References, Relations, Mapping Label, and Landing Page are visible.
- Table:** A table showing the relationship between the concept and its terminology.

Concept	Terminology	Link
Published data record	Qualifier Descriptor	

Model alignment between Applications through the Reference Ontology



Ontology and Constraints Model in integrated RTS Tool

Roche Terminology System v3.9.0 **PRODUCTION**

FAIR Metrics

robert.trypuz@roche.com [Logout](#)

Terminology Application Variable Model Curation Administration Information

Search

Reference Model Navigator

- Reference Model
 - Class
 - Entity
 - Continuant
 - Generically dependent continuant
 - Information content entity
 - Collection
 - Data item
 - Decision
 - Directive information entity
 - Formulation actual numeric set condition
 - Nominal agent dose
 - Nominal numeric set condition
 - Plan specification
 - Algorithm
 - Plan of activity
 - Protocol
 - Study design
 - Substance mixture description
 - Portion of substance description
 - Document
 - Document part
 - Enumeration
 - Identifier
 - Study theme
 - Independent continuant
 - Specifically dependent continuant
 - Occurrent

Model Navigator

- Application
 - Productive
 - Under development
 - FISH Biospecimen
 - FISH Formulation
 - FISH Study Design
 - Cell and Virus
 - FISH SD Execu
 - FISH SD Model
 - Collected I
 - Formulation
 - Local descriptor
 - Model animal
 - Molecular entity type
 - Nominal agent dose
 - Non clinical in vivo study
 - Person
 - Plan of activity
 - Plan of study activity sequence
 - Produced unmixed substance
 - Produced unmixed substance type
 - Study design
 - Study subject
 - Study subject group
 - Substance mixture type

SHACL Constraint Models

- is planned for
- is primary topic of
- is reviewed by
- specifies study subject group

Application Model Property Entity View

Properties [Edit](#) [Refresh](#)

r:	ROX38338272444006757
el:	is planned for
e:	ObjectProperty
st:	is planned for
id:	
Reference URI:	
Inverse of:	

Local Usage Properties

Type:	ObjectProperty
Used at class:	Study design
Target class:	Non clinical in vivo study
Multiplicity:	1..1
Local Technical Key:	plannedFor

Properties

Definition	Comment
indicates a planned process that is to be executed according to the plan	

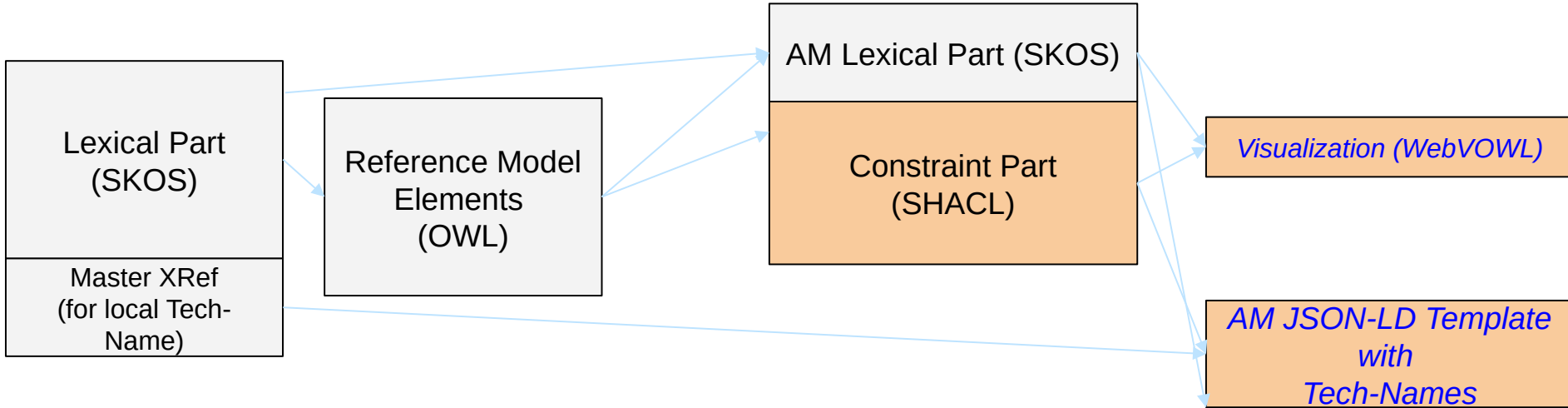
Relationships between layers

Terminology

Ontology

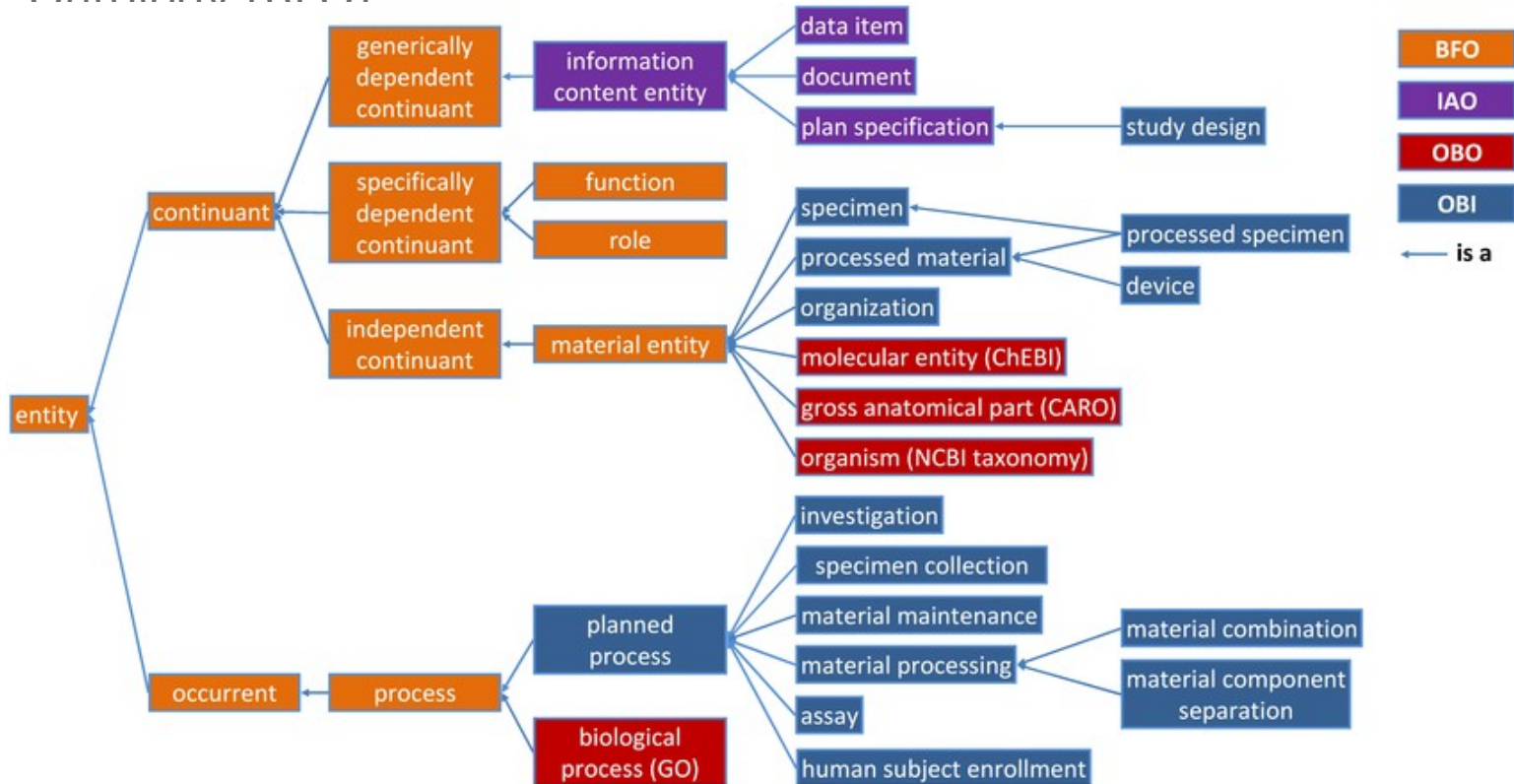
Application Model (AM)

generated



Foundational Ontologies

Ontology of Biomedical Investigations (OBI) & Basic Formal Ontology (BFO)



Application Specific Model in SHACL

RTS Curation Client

The screenshot displays the SHACL Curation Client interface. On the left, the **Model Navigator** shows a tree structure under **FISH SRS Model**. The **Study** class is expanded, listing properties such as **has study scientific title**. An orange callout box labeled **NodeShape** points to the **Study** class, and another labeled **Property Shape** points to the **has study scientific title** property.

The main area shows the **Entity View** for the **Study** class. The **Model Global Properties** section includes:

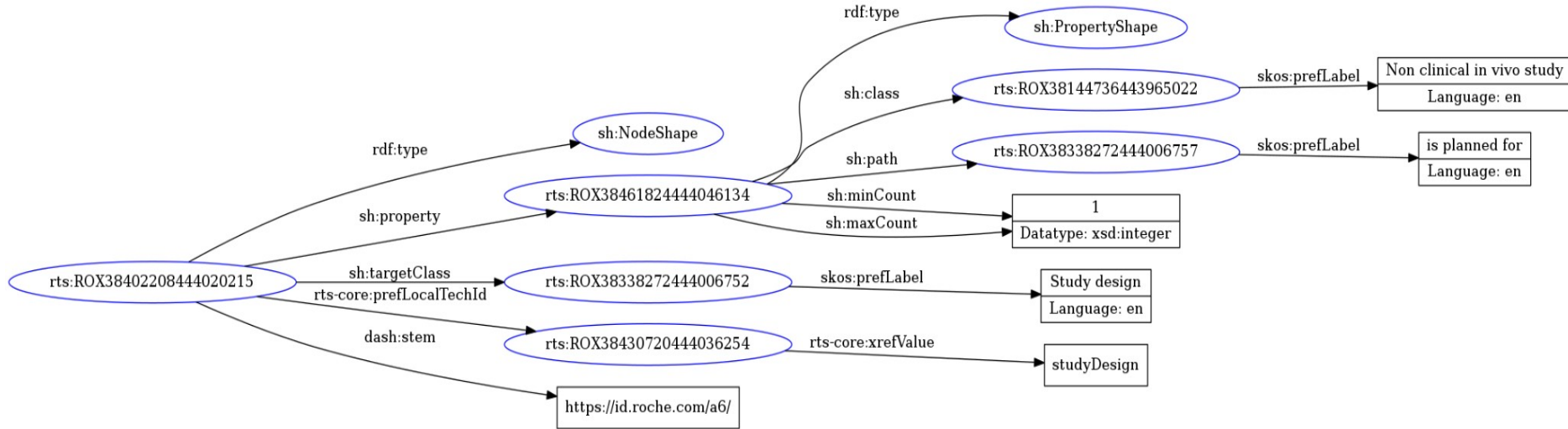
- Master Concept Identifier: ROX37664352443822935
- Preferred Label Identifier: has study scientific title
- Local Technical Key: studyScientificTitle
- Preferred Reference URI: (empty)

The **Local Usage Properties** section includes:

- Used at class: Study
- Target class: (empty)
- Inverse of: (empty)
- Data type: String
- Multiplicity: 1..1

An orange callout box labeled **Property Constraints** points to the **Multiplicity** field. Below this, the **Properties** section contains a **Definition** field with the text: "Indicates a comprehensive summary of study design and objectives, aimed at scientific audience. Scientific" and a **Comment** field.

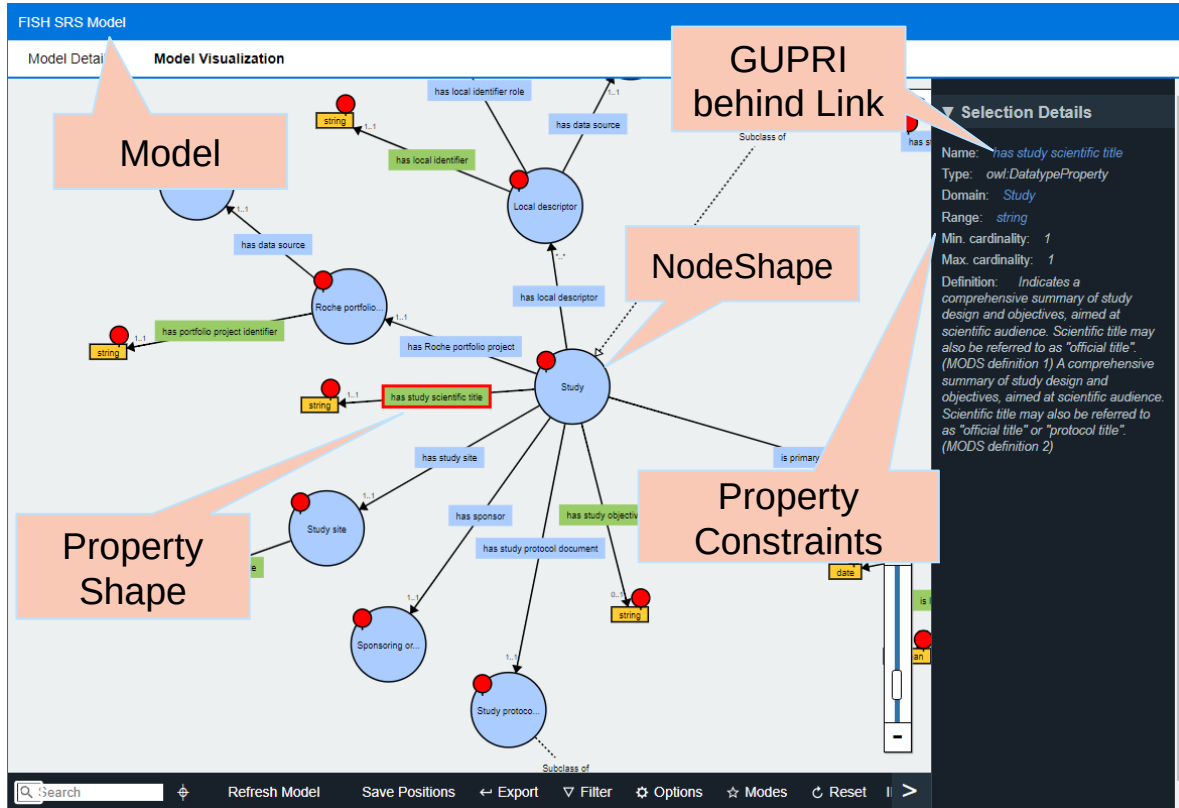
SHACL Constraint Model Example



Namespaces:
 rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
 xsd: <http://www.w3.org/2001/XMLSchema#>
 rts-core: <http://ontology.roche.com/model/rts/core/>
 rts: <http://ontology.roche.com/>
 skos: <http://www.w3.org/2004/02/skos/core#>
 skos-xl: <http://www.w3.org/2008/05/skos-xl#>
 sh: <http://www.w3.org/ns/shacl#>
 dash: <http://datashapes.org/dash#>

Visualizing the SHACL Model - like an Ontology

RTS Model Browser



Programmatic Access to Application Model

Public REST API

Application Model RTS Application Model API

GET `/int/pre/appmodels/{mid}` Export Application Model by given id.

Parameters Try it out

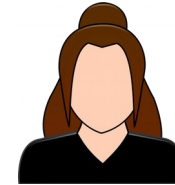
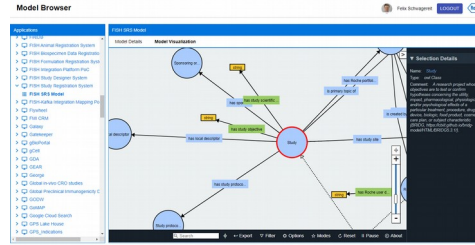
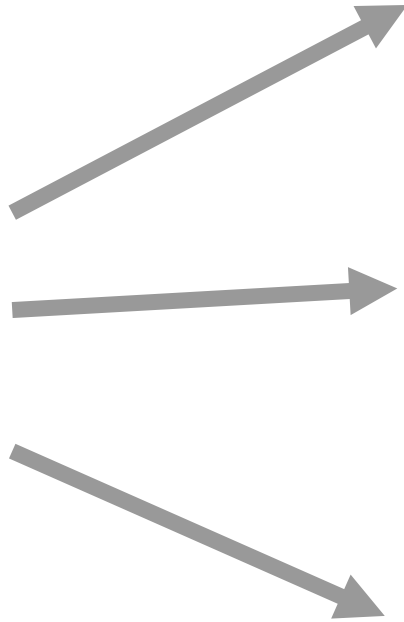
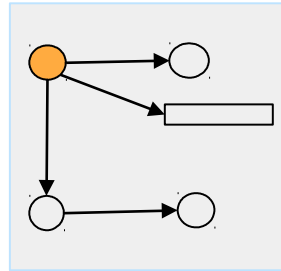
Name	Description
X-Auth-Token	
string <i>(header)</i>	

```

"property": [
  {
    "id": "ROX38461824444047818",
    "type": "PropertyShape",
    "datatype": "xsd:string",
    "maxCount": "1",
    "minCount": "1",
    "path": {
      "id": "ROX37664352443822935",
      "type": "Class",
      "prefLabel": "has study scientific
title"
    }
  }
],

```

Usages for the Reference Model + Application SHACL Model



Visualization
Domain experts, DB developers, Data scientists, ...

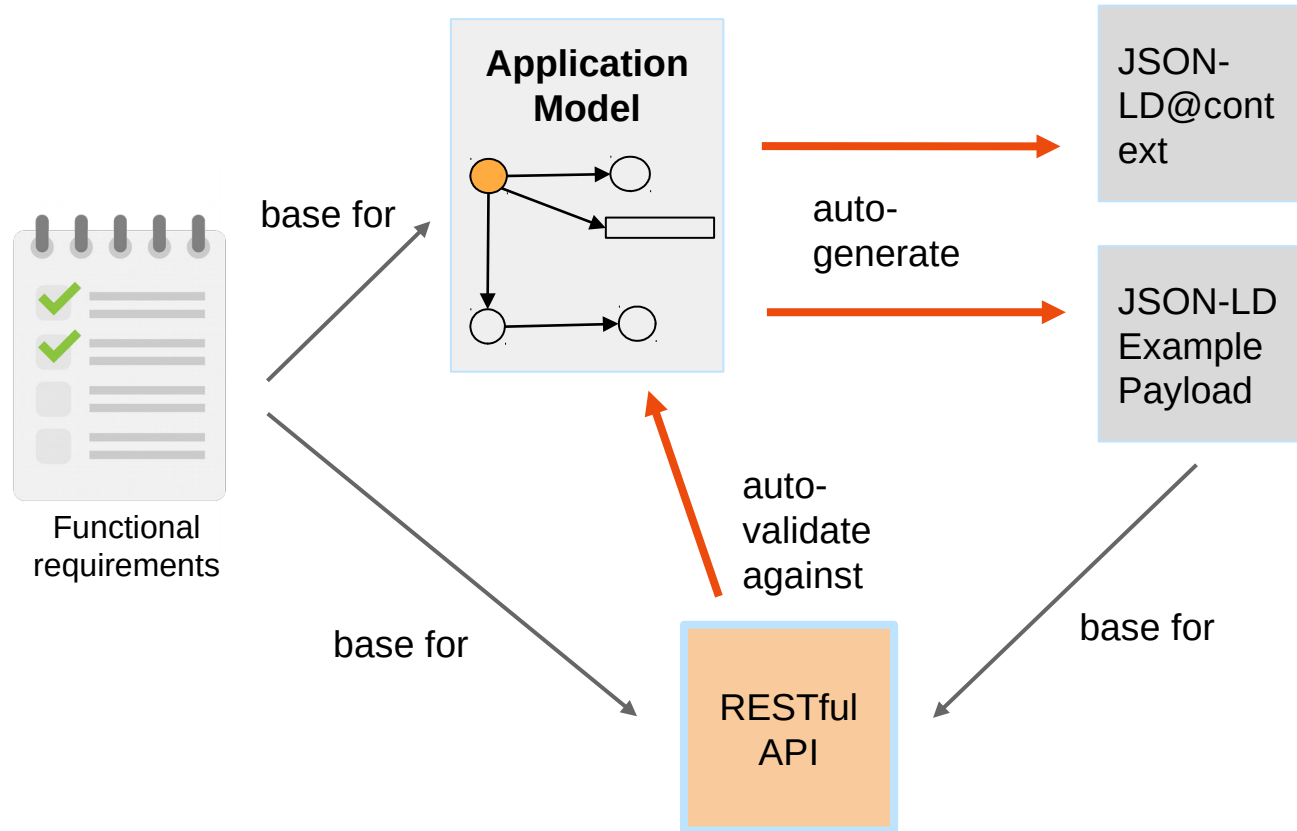


Semantic Database
data constraints, semantics of data



API
Semantic description of payload, structure of payload

Model-supported JSON-LD REST API Design



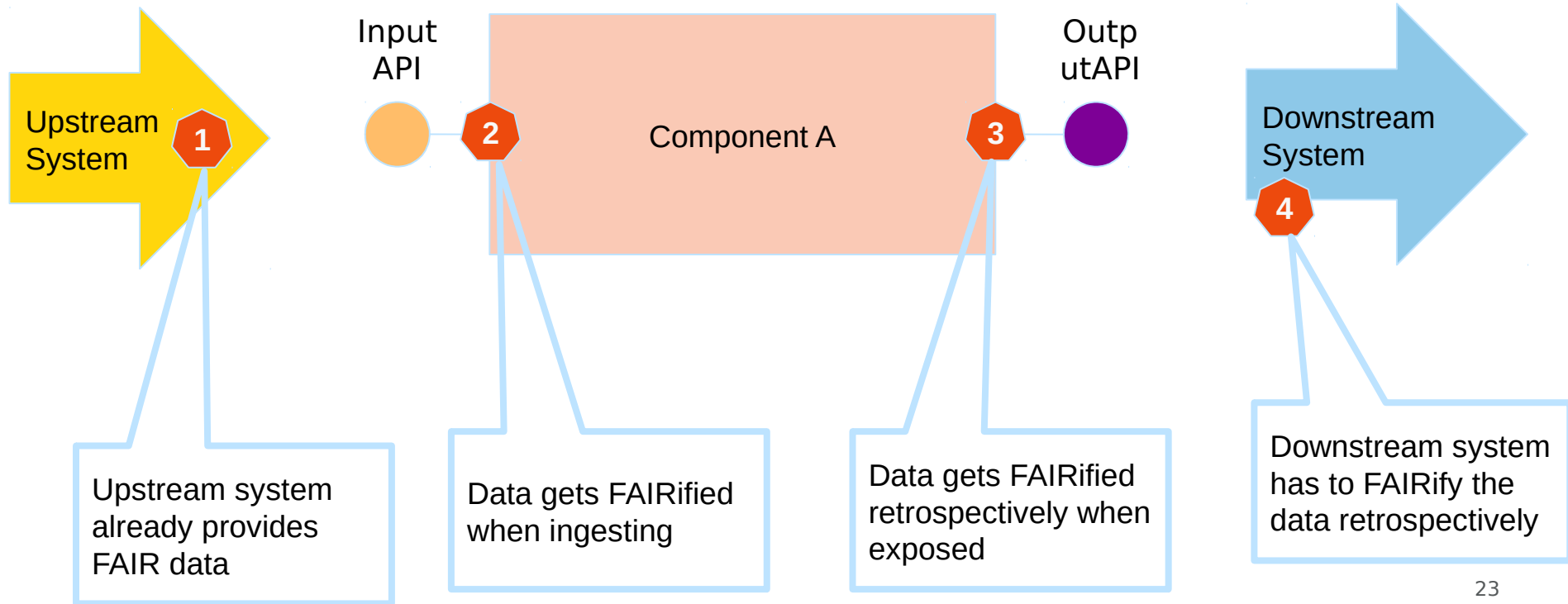
Apply a FAIR API Design with JSON-LD as payload format

Recommendation for Standards and code

- JSON (JavaScript Object Notation)
 - de facto standard for Web Services based on REST
- JSON-LD (JavaScript Object Notation for Linked Data)
 - Valid JSON !
 - Keys can be still defined locally as part of API definition
 - Standardized (W3C Recommendation)
 - Adds semantic annotations
 - is also a valid semantic graph (RDF)
 - maps payload to registered model
 - Reference Implementation of JSON-LD processor in Java: Apache Jena

Alternative places where to FAIRify data

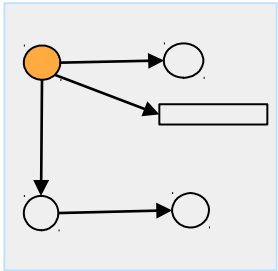
For the goal of having FAIR data in the downstream system



Example: Study Design Object as JSON-LD

```
{
  "@context":{
    "@import":"https://ontology-service.roche.com/rts2-api/v3/appmodels/
ROX38389248444017485/context?version=2022-06-08T09%3A05%3A30.000Z"
  },
  "@id":"https://id.roche.com/a6/1",
  "@type":"StudyDesign",
  "plannedFor":{
    "@id":"https://id.roche.com/a2/1",
    "@type":"NonClinicalInVivoStudy"
  },
  "numberOfFirstStudyDay":1,
  "comment":"some comment",
  "reviewedBy":{
    "@id":"https://id.roche.com/xyz/fishcur1",
    "@type":"Employee",
    "userName":"GLO FISH CURATOR"
  },
}
```


Generate API payload format from the model



Generate



RTS

GET /int/prj/appmodels/{mid}/example Get Application Model Example

Parameters Cancel

Name	Description
X-Auth-Token <small>string (header)</small>	OAuth header <input type="text" value="X-Auth-Token - OAuth header"/>
Authorization <small>string (header)</small>	Basic Auth header <input type="text" value="Authorization - Basic Auth header"/>
mid * required <small>string (path)</small>	Application Model Id <input type="text" value="ROX3840134444019991"/>
focalClassId * required <small>string (query)</small>	Focal Class Id <input type="text" value="ROX3814473644396502"/>

Server response

Code	Details
200	<pre style="border: 2px solid orange; padding: 5px;"> { "@context": "https://ontology-services.roche.com/rt2-api/1/int/prj/appmodels/ROX3840134444019991/context", "@graph": [{ "@type": "foaf:Individual", "ianaLabel": "farAuthTestLicense", "@type": "ianaLabel:FarAuthTestLicenseClass", "ianaLabel": "farAuthTestLicenseIdentifier": "string", "ianaLabel": "farAuthTestLicenseIdentifier": "string", "@type": "ianaLabel:FarAuthTestLicenseClass" }], "externalContact": "string", "pipFlag": true, "authScopePortfolioProject": { "@type": "scopePortfolioProject", "idSource": "id", "id": "rt2:ROX3782864443893996", "prefLabel": "pipOp10000", "portfolioProjectIdentifier": "string" }, "hasStudySite": { "@type": "StudySite", "external": true }, "internalContact": { "@type": "Employee", "user": "string", "username": "string" }, "laboratoryFacility": { "@type": "MetaentRecord", "createdAt": { "@type": "Employer", "user": "string", "username": "string" } }, "createdAt": "2021-09-09", "idSource": { "@type": "rt2:ROX38039328443948504", "prefLabel": "Safety Data Integration" } } </pre>

Implement



FISH SRS API - TEST

GET /api/study/{studyId} Find studies by the global study id

Parameters Cancel

Name	Description
studyId * required <small>string (path)</small>	Study Id <input type="text" value="1"/>

Execute Clear

Responses

Code	Details
200	<pre style="border: 2px solid orange; padding: 5px;"> { "@context": "https://ontology-services.roche.com/rt2-api/1/int/prj/appmodels/ROX3840134444019991/context", "@graph": [{ "@type": "foaf:Individual", "ianaLabel": "farAuthTestLicense", "@type": "ianaLabel:FarAuthTestLicenseClass", "ianaLabel": "farAuthTestLicenseIdentifier": "string", "ianaLabel": "farAuthTestLicenseIdentifier": "string", "@type": "ianaLabel:FarAuthTestLicenseClass" }], "externalContact": "string", "pipFlag": true, "authScopePortfolioProject": { "@type": "scopePortfolioProject", "idSource": "id", "id": "rt2:ROX3782864443893996", "prefLabel": "pipOp10000", "portfolioProjectIdentifier": "string" }, "hasStudySite": { "@type": "StudySite", "external": true }, "internalContact": { "@type": "Employee", "user": "string", "username": "string" }, "laboratoryFacility": { "@type": "MetaentRecord", "createdAt": { "@type": "Employer", "user": "string", "username": "string" } }, "createdAt": "2021-09-09", "idSource": { "@type": "rt2:ROX38039328443948504", "prefLabel": "Safety Data Integration" } } </pre>

Generation of JSON-LD Example Payload

RTS Model Browser PRO.D 3.15.0 Felix Schwagererit LOGOUT Roche

Applications

- ▼ Applications
 - ▼ Productive
 - >
 - >
 - >
 - >
 - 🔰 Cell and Virus exposure
 - 🔰 FISH SD Execution Model
 - 🔰 FISH SD Model
 - 🔰 **Merged Model**
 - 🔰 proposed exposure activity
 - 🔰 SD Substance-Molecule model
 - 🔰 Substance exposure activity
 - 🔰 Substance reference model
 - >
 - >
 - >
 - >
 - >
 - >
 - >
 - >
 - >
 - >
 - >
 - >
 - >
 - >
 - >
 - >

Merged Model

Model Details Model version: 2022-08-24T16:13:01Z Model Visualization Context Payload Example JSON-LD Validation

Focal Class / Shape

Expand to prefLabels

RTS as Base

Avoid special characters

Depth

GENERATE
COPY TO CLIPBOARD
DOWNLOAD

Payload Example

```

{
  "@context": {
    "@import": "https://ontology-services.roche.com/rts2-api/v3/appmodels/ROX38521440444104515/context?version=2022-08-24T16:13:01Z"
  },
  "@graph": {
    "@id": "https://id.roche.com/a6/studyDesign/1",
    "@type": "studyDesign",
    "comment": "string",
    "containsCollectedBiospecimenGroupSpecification": [
      {
        "@id": "https://id.roche.com/a6/collectedBiospecimenGroupSpecification/1",
        "@type": "collectedBiospecimenGroupSpecification",
        "biospecimenType": {
          "@id": "rts:ROX1308059566380",
          "@type": "collectedBiospecimenTypeATEnum",
          "prefLabel": "Thymus"
        },
        "denotes": {
          "@id": "https://id.roche.com/a6/collectedBiospecimenGroup/1",

```

Example: Study Design Object as Generated from SHACL

```
"@context": "https://ontology-services.roche.com/rts2-api/v3/appmodels/R0X38389248444017485/context?version=2022-04-07T14%3A46%3A15.000Z",
"@graph": {
  "@id": "https://id.roche.com/a6/1",
  "@type": "StudyDesign",
  "comment": "string",
  "plannedFor": {
    "@type": "NonClinicalInvivoStudy",
    "hasParticipant": [
      {
        "@type": "Subject",
        "@id": "https://id.roche.com/a6/1/subjects/1"
      },
      ...
    ]
  }
}
```

Example: JSON-LD Context for Study Design Object

```

{
  "@context": {
    "id": "@id",
    "type": "@type",
    "rts": "http://ontology.roche.com/",
    "prefLabel":
"http://www.w3.org/2004/02/skos/core#prefLabel",
    "StudyDesign": "rts:R0X3833827244400675",
    "plannedFor": {
      "@id": "rts:R0X3833827244400675",
      "@type": "@id"
    },
    "title": "rts:R0X37664352443822935",
    "sponsoringOrganization": "rts:R0X38092032443955602"
  }
}

```

Look up "title" in RTS for detailed semantics
<http://ontology.roche.com/R0X37664352443822935>
 → "Study scientific title" : "Indicates a comprehensive summary [...]"

Validation of JSON-LD Payload

RTS Model Browser PROD 3.15.0

Felix Schwagerer

LOGOUT



The screenshot displays the 'Merged Model' interface with the 'JSON-LD Validation' tab selected. On the left, a tree view shows the application structure, with 'Merged Model' highlighted. The main area contains a file selection interface with a 'Datei auswählen' button and 'Keine ausgewählt' text. Below this is a large text area for pasting JSON-LD content. The 'JSON-LD Validation Result' section shows a list of errors:

- 2. The JSON-LD payload contains Classes not defined in the ApplicationModel: [collectedBiospecimenTypeEnum, recordStatusAEnum, studySubjectGroupCategoryAEnum]
- 3. Class "recordStatusAEnum" contains properties that are not defined in Application Model Class: [prefLabel]
- 3. Class "collectedBiospecimenTypeEnum" contains properties that are not defined in Application Model Class: [prefLabel]
- 3. Class "collectedBiospecimenTypeEnum" contains properties that are not defined in Application Model Class: [prefLabel]
- 3. Class "studySubjectGroupCategoryAEnum" contains properties that are not defined in Application Model Class: [prefLabel]
- 3. Class "studySubjectGroupCategoryAEnum" contains properties that are not defined in Application Model Class: [prefLabel]
- 4. The range of the "biospecimenType" property is below the minimal count of the said range defined by the application model
- 4. The range of the "membersDerivedFromGroup" property is below the minimal count of the said range defined by the application model
- 4. The range of the "localIdentifier" property is below the minimal count of the said range defined by the application model
- 4. The range of the "memberOf" property is below the minimal count of the said range defined by the application model
- 4. The range of the "assignedFrom" property is below the minimal count of the said range defined by the application model
- 4. The range of the "participatesIn" property is below the minimal count of the said range defined by the application model
- 4. The range of the "name" property is below the minimal count of the said range defined by the application model

F.A.I.R Metrics Assessment



Findability

- ★ F1. Metadata are assigned globally unique and persistent identifiers:
- ★ F2. Data are described with rich metadata (structured and grounded)
- ★ F3. Metadata clearly and explicitly include the identifier of the data they describe
- ★ F4. Metadata are registered or indexed in a searchable resource



Accessibility

- ★ A1. Metadata are retrievable by their identifier using a standardised communication protocol
 - A1.1. The protocol is open, free and universally implementable
 - A1.2. The protocol allows for an authentication and authorisation where necessary
- ★ A2. Metadata is accessible even when the data is no longer available



Interoperability

- ★ I1. Metadata use a formal, accessible, shared, and broadly applicable language for knowledge representation
- ★ I2. Metadata use vocabularies that follow the FAIR principles
- ★ I3. Metadata include qualified references to other metadata



Reusability

- ★ R1. Metadata are richly described with a plurality of accurate and relevant attributes
 - ★ R1.2. Metadata are associated with detailed provenance
 - ★ R1.3. Metadata meet domain-relevant community standards

Required for FAIR Data APIs

- F1
- F2
- F3
- A2
- I2
- I3

Supported by FAIR Data APIs

- F4
- A1 (A1.1 A1.2)
- I1

Acknowledgements



Authors: Felix Schwagereit, Martin Romacker, Fabien Richard, Robert Trypuz, Thomas Liener, Olivier Roche

RTS Team and FISH Team Topic Collaborators: Adam Zawada, Anastasios Moresis, Anke Eisenmann, Christian Blumenröhr, James Hinchliffe, Jaroslaw Gorski, Joachim Rupp, Karol Broniec, Krzysztof Majewski, Michal Bielak, Pratihtha Duhan, Rama Balakrishnan, Shima Dastgheib, Silvia Jimenez

Doing now what patients need next

Slide title

Subtitle goes here but is not mandatory