# Fair-EASE: Data Provider around ERDDAP

Maxence Couppey

: Already Done

: Will be done by the end of September

: Could be great to be done by the end of September

## **History System**

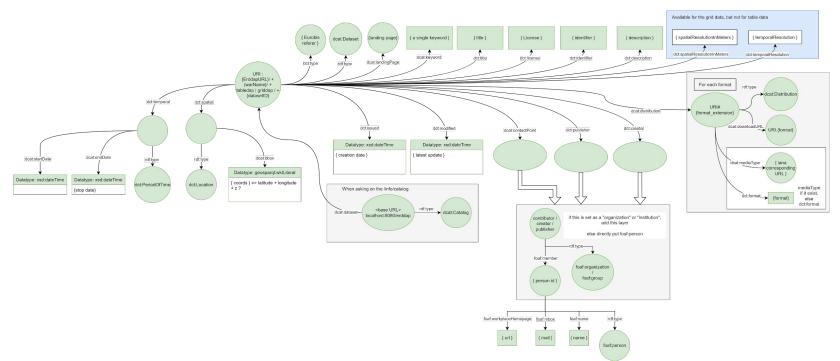
- -> There's no in-built way to track every changes that happened to a dataset
- -> On every ERDDAP restart, the RSS date is updated

Store all updates description of each dataset in a persistent File

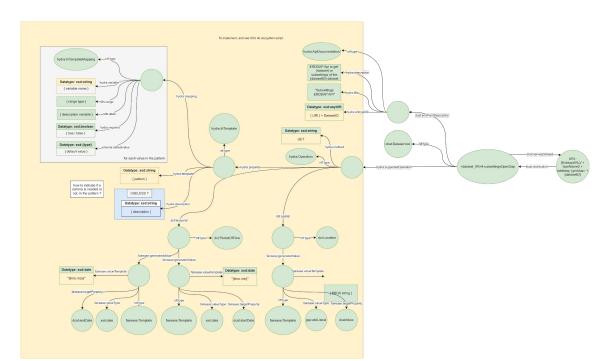
/{bigParentDirectory}/history.json

```
"test_maj": [
  "date": "2023-07-27T07:52:09Z",
  "description": "A combinedGlobalAttribute changed:",
  "old": " old line #12=" infoUrl=???".".
  "new": " new line #12=" infoUrl=https://vliz.be"."
  "date": "2023-07-27T07:51:29Z".
  "description": "The number of variables changed:".
  "old": " old=1.".
  "new": " new=2."
  "date": "2023-07-27T07:50:39Z".
  "description": "The number of variables changed:",
  "old": " old=2.".
  "new": " new=1."
  "date": "2023-07-27T07:48:54Z".
  "description": "The destinationName for dataVariable #0=temp changed:",
  "old": " old=temperature.".
  "new": " new=temp.".
  "combined": "The combinedAttribute for dataVariable #0=temp changed:",
  "oldCombined": " old line #1=" ioos_category=Temperature",",
  "newCombined": " new line #1=" _FillValue=NaNd":"
  "date": "2023-07-27T06:59:26Z".
  "description": "Initial init test_maj",
```

#### (Pre) Definition of a RDF Model

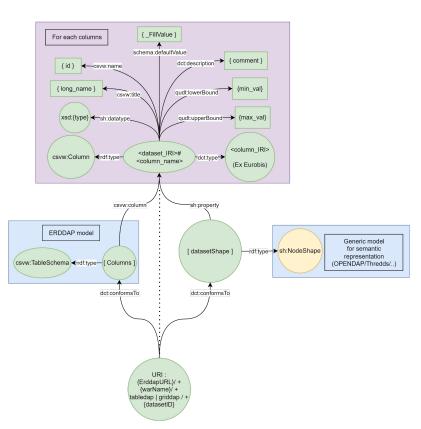


#### (Pre) Definition of a RDF Model



#### (Pre) Definition of a RDF Model

content Description:



#### **Add File Formats**

#### File type: (more information)

```
IntmlTable - View a UTF-8 .html web page with the data in a table. Times are ISO 8601 strings.

IsonId - View the JSON-LD RDF serialization representation.

In 3 - View the Notation3 RDF serialization representation.

In 4 - View the N-Triples RDF serialization representation.

In 5 - View the N-Quads RDF serialization representation.

In 6 - View the N-Quads RDF serialization representation.

It 7 - View the RDF-XML RDF serialization representation.

It 7 - View the Turtle RDF serialization representation.

It 7 - View the Turtle RDF serialization representation.
```

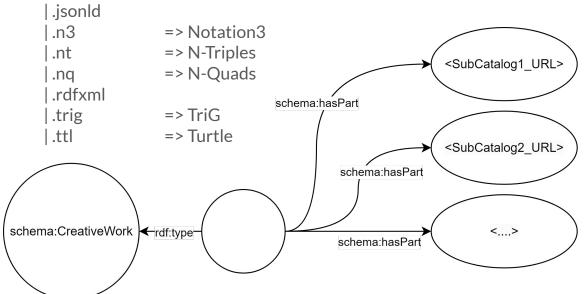
#### **Sub-Catalog Access**

localhost:8080/erddap/info/catalog.{format}?page=1&itemsPerPage=5 .jsonld <dataset1 URL> <RDF MODEL> .n3 => Notation3 dcat:dataset => N-Triples .nt => N-Quads .nq .rdfxml .trig => TriG <RDF MODEL> <dataset2 URL> .ttl => Turtle Sort by the dcat:dataset latestModificationDate <erddap Base <RDF MODEL> dcat:Catalog ← rdf:type-<....> dcat:dataset URL>

 Doesn't return the private dataset (if you don't have rights to read them)

### **Catalog Access**

localhost:8080/erddap/info/catalog.{format}



#### **Content Negotiation**

```
\label{localhost:8080/erddap/info/index.html} \begin{tabular}{l} => & localhost:8080/erddap/info/index.html \\ & Header: "*/*" \end{tabular}
```

```
localhost:8080/erddap/info/index.html => localhost:8080/erddap/info/catalog.ttl Header: "text/turtle"
```

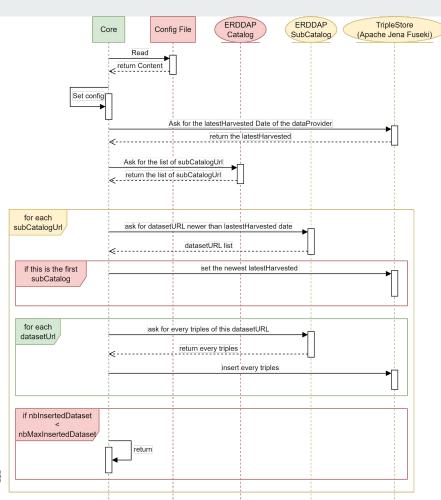
#### kind of a Harvester

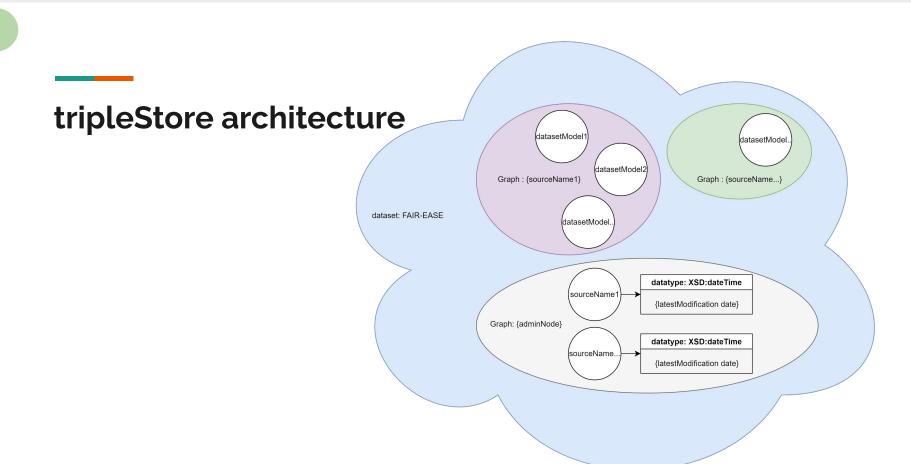
#### Does:

- "smart" inserting of the Dataset (does not drop every dataset in the source & re-insert them all)
- Follow the main path (doesn't know that this is an ERDDAP Server, can work with everything)
- Can easily be threadable for each dataProvider and use different harvesting timer

#### Question:

- Depending on the RDF library that'll be used :
  - If the RDF lib makes query on <u>cached</u> data (from file), ok
  - If not, we'll request the subCatalog for every Dataset, overloading/unused information => use the {dataset}.{format}





#### kind of a DAL

from a 'user', based on a request with parameters

- return a list of matching dataset with:
  - base url
  - subsetting URL who fit the request => will probably be rounded to be easily cacheable
  - influence of each parameter of the request

#### Will be done:

- Create documentations (for ERDDAP from ERDDAP, General architecture documentations for FAIR-EASE a more use)
- Improve history
- Improve harvester (Delete dataset in tripleStore on update / append it to the LDES, remove old updateDate, thread it to get multiple source in input and query them every source-defined time)

## Hopefully, will be done

- LDES feed to track changes on dataset
- "DAL" => After a request, with 2 or + corresponding dataset => merge them into a single one

## Thanks!

Let's talk