Simdal

Madrid Interop - ESAC - 2014 May, 22

David Languignon

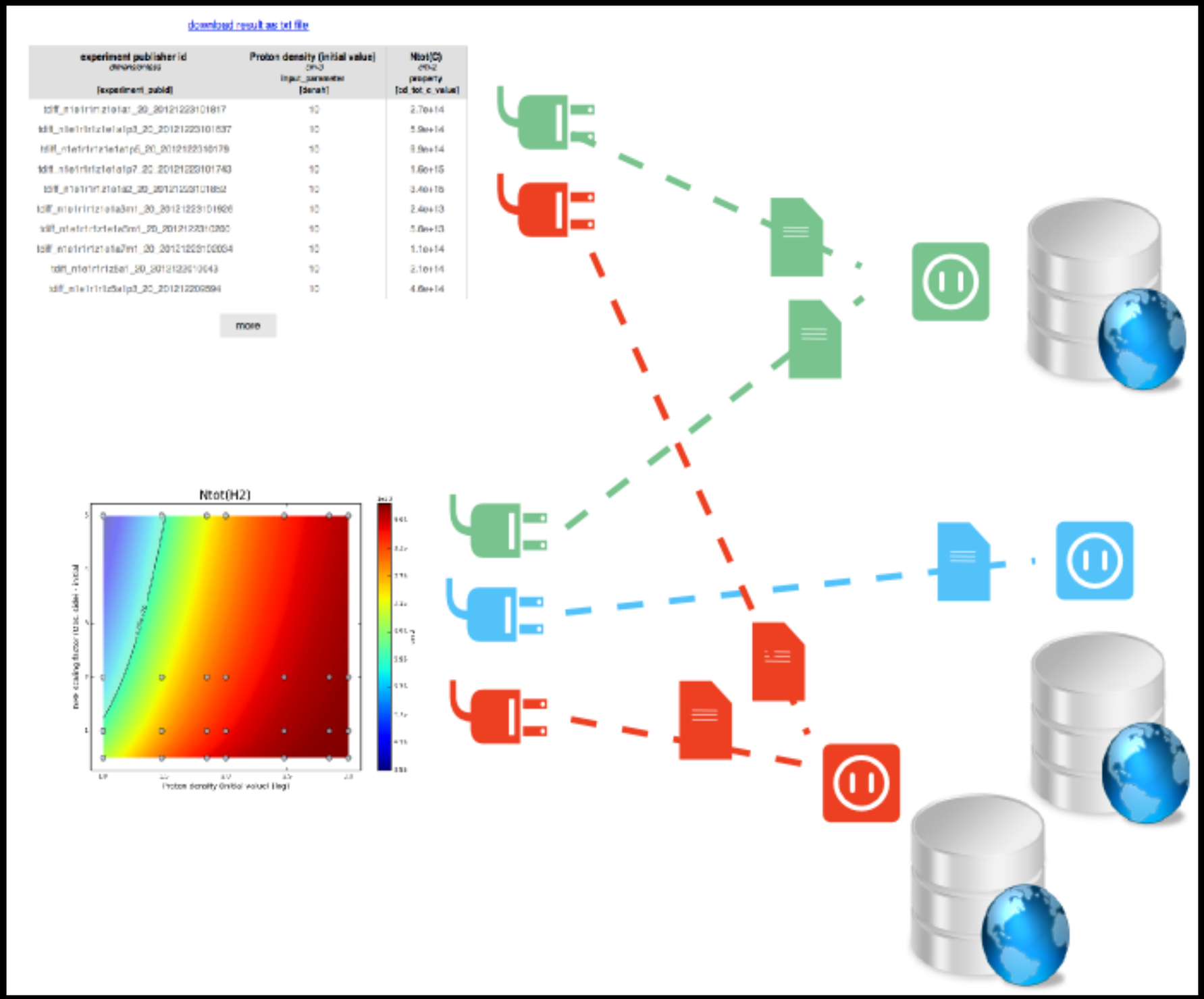
A numerical model exists for your observation.

Ask SimDAL

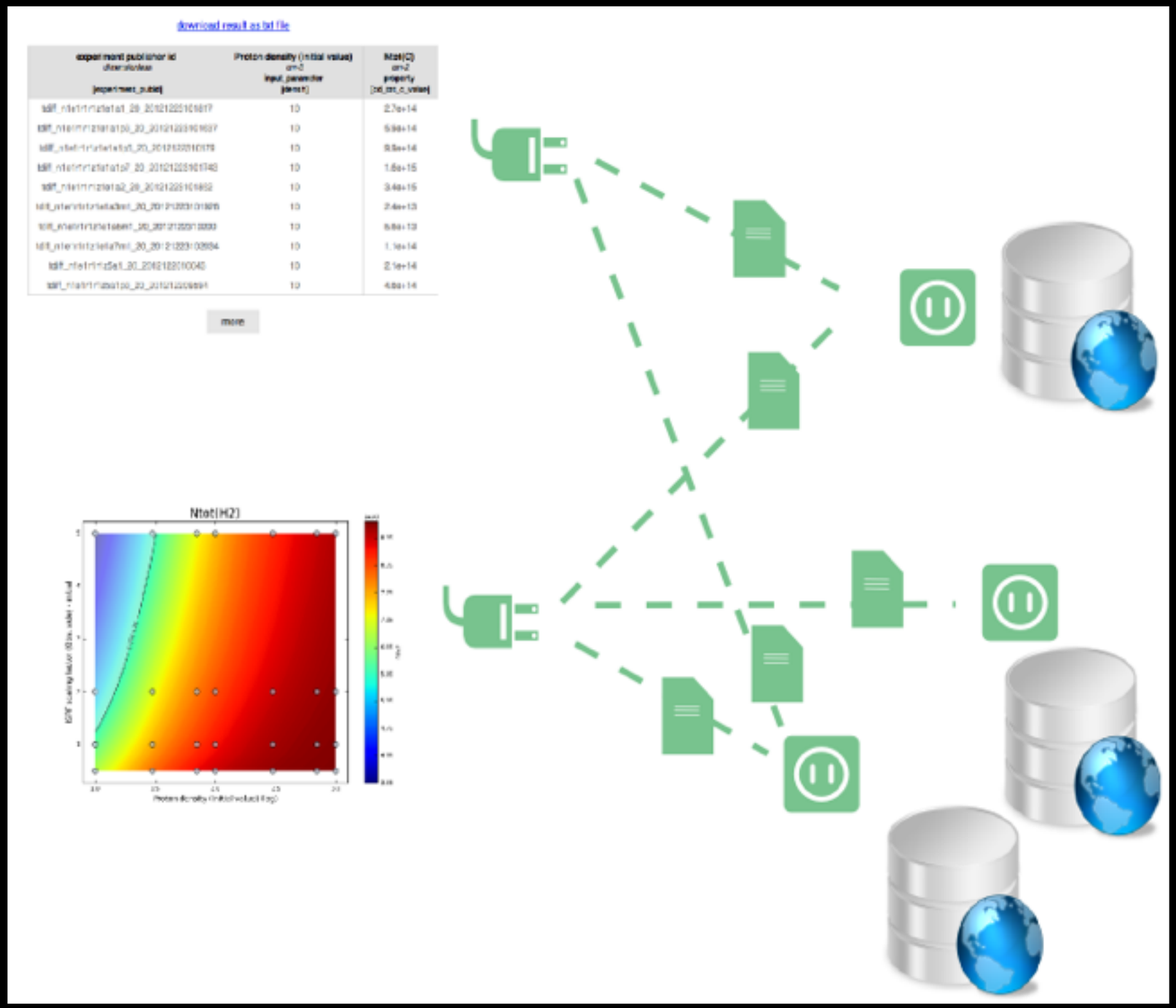
You need help

- Help me to find **what I want to search for**
  - Simulation Registry (~SimDB / VO Registry)
- Help me to **search for & get the values**
  - Simulation Search
- Help me to get **the model raw data**
  - Simulation Data Access / Cutout

But why ?


download result as txt file
experiment publisher id
dimensionless
[experiment_pubid]
Proton density (initial value)
cm-3
input_parameter
[1e+xx]
Ntot(C)
cm-2
property
[cd_tot_c_value]
more
Ntot(H2)

Because you want to understand what I tell you



Simulations Registry

Simulations Registry

- Where **you find what you want to search for**
  - astronomical object : *halos, cloud, star*
  - simulation code : *ramses, gadget, pdr*
  - physical quantity : *mass, proton density*
  - scientist : *Gerard Lemson, Einstein*

How does Registry help ?

- **FullText search** on Project / Service / Protocol

Think Google bar

- Provides **SimDM XML serialisations**
  - **pivot format**
  - with / without nested (large) collections
  - **streaming access** to nested collections

API example

```
/search?q=proton density&results=20

/services?project=pdr-grid-201403

/projects?authority=ism.obspm

/protocols?project=par-grid-201403

/input_parameters?protocol=pdr161

/properties?protocol=pdr161&object=cloud
```

Simulations Search

Simulations Search

- Help you **get / constraint the values**
- Help you **identify a dataset to cutout**
- **Provide** the SimDM **Experiment XML serialisation**

How does Search help ?

- **Defines a virtual (flat, big) table**
  - input parameters, statistical summaries, … as “columns”
- **Describes the columns** (values, ranges, …)
- **Defines a way to query the columns**

Why a virtual table ?

- The **end user doesn’t necessarily know SimDM**
- Just present him a (virtual) **table with columns.**
- Whatever they are SimDM input parameter, property…

From table to Nd cube

A flat table is actually a coord system

table -> coord system

column -> axis

row -> point / vector

Why not a standard table ?

- We need to define a **standard NOT COUPLED with implementations details**
- Which let us scale in the future

NO TAP (adql => RDBMS)

but **virtual columns ~ parameters ~ axis**

Interop Napoli (2011), Pune (2011), Heidelberg (2013)

Describe axis

- What ?
- basic **values meta** (ranges, etc…, votable like)
- pointers towards **SimDM protocol file** for more meta (~ virtual TAP_SCHEMA)

Describe axis

- How ?
  - self description **/axis-metadata** endpoint
    - S3 like self-description
  - pointer towards **SimDM protocol file** for more meta (~ **virtual TAP_SCHEMA**)

/axis-metadata

protocol publisherDID

protocol.xml

```
<PARAM ID="temp" name="INPUT:temp" unit="Gyr"
    utype="simdm://resources/protocol/input_parameter.publisherDID"/>
  <DESCRIPTION>Age</DESCRIPTION>
  <LINK content-role="OPERATORS" value="O_LT_EQ"/>
  <VALUES type="actual">
    <OPTION value="0.001"/>
    <OPTION value="0.002"/>
    <OPTION value="0.005"/>
    <OPTION value="0.006"/>
  </VALUES>
</PARAM>
```

Discuss

- self description
- standard votable output
- utype towards protocol.xml

How to query ?

- Think in **tables, they are easy to understand**
  - **close to VO semantic** (adql / tap / votable)
  - **conceptual subset of SQL for single flat tables** “simple SQL”

select *<columns>* ***where*** *<filters>*

select *temp* ***where*** *temp = 0.005*

```
<PARAM ID="temp" name="INPUT:temp" unit="Gyr"
    utype="simdm://resources/protocol/input_parameter.publisherDID"/>
  <DESCRIPTION>Age</DESCRIPTION>
  <LINK content-role="OPERATORS" value="O_LT_EQ"/>
  <VALUES type="actual">
    <OPTION value="0.001"/>
    <OPTION value="0.002"/>
    <OPTION value="0.005"/>
    <OPTION value="0.006"/>
  </VALUES>
</PARAM>
```

available operators

```
<       : O_LT
>       : O_GT
=       : O_EQ
< & = : O_LT_EQ
> & = : O_GT_EQ
< & > : O_LT_GT
~       : O_REGEX (for string axis)
```

because we don’t always have enough information
to filter < AND > (ex: min/max stats only)

Query request example

- /query?select=p1,p4&filters=p7<2e23,p5~"density*"
- POST

```
{
    "select"  : ["p1", "p4"],
    "filters" : [
        {"axis" : "p7", "op" : "<", "value" : "2e23"},
        {"axis" : "p5", "op" : "~", "value" : "density*"}
    ]
}
```

Query response example

- votable
- groups used to define input parameters and property statistics

```
<GROUP ID="group_temp" utype="simdm:resource/experiment/parameter_setting">
  <PARAM name=""
         utype="simdm:/resource/experiment/parameter_setting.input_parameter"
         value="densh"/>
  <FIELDRef ref="temp_numeric_value"/>
</GROUP>

<FIELD name="" ID="temp_numeric_value" ucd=""
       ref="group_temp"
       utype="simdm:/resource/experiment/parameter_setting.numeric_value.value"
       datatype="float" unit="d"/>
```

Optional languages

- ADQL (back compatibility with existing TAP services)
- Simple SQL with **arithmetic**
  - /query?select=p4&filters=(p2 + p7)<2e23”

Discuss

- how to pass parameters to /query ?
- query language
- post / get

API examples

```
/experiments?protocol=pdr161

/parameter_settings?experiment=tdiff_2e3&results=5&page=2

/output_datasets?experiment=tdiff_2e3

/statistical_summaries?output_dataset=tdiff_2e3_d1
```

Data Access - Cutout

Data Access

- (repeat) **tables are too specific**, let's generalise

  (virtual) tables -> coord system

  columns -> axis

- **SSD & Data Access can be defined the same way**

Interop Pune (2011), Heidelberg (2013)

An unique API remains

- /axis-metadata
- /query

Query operations

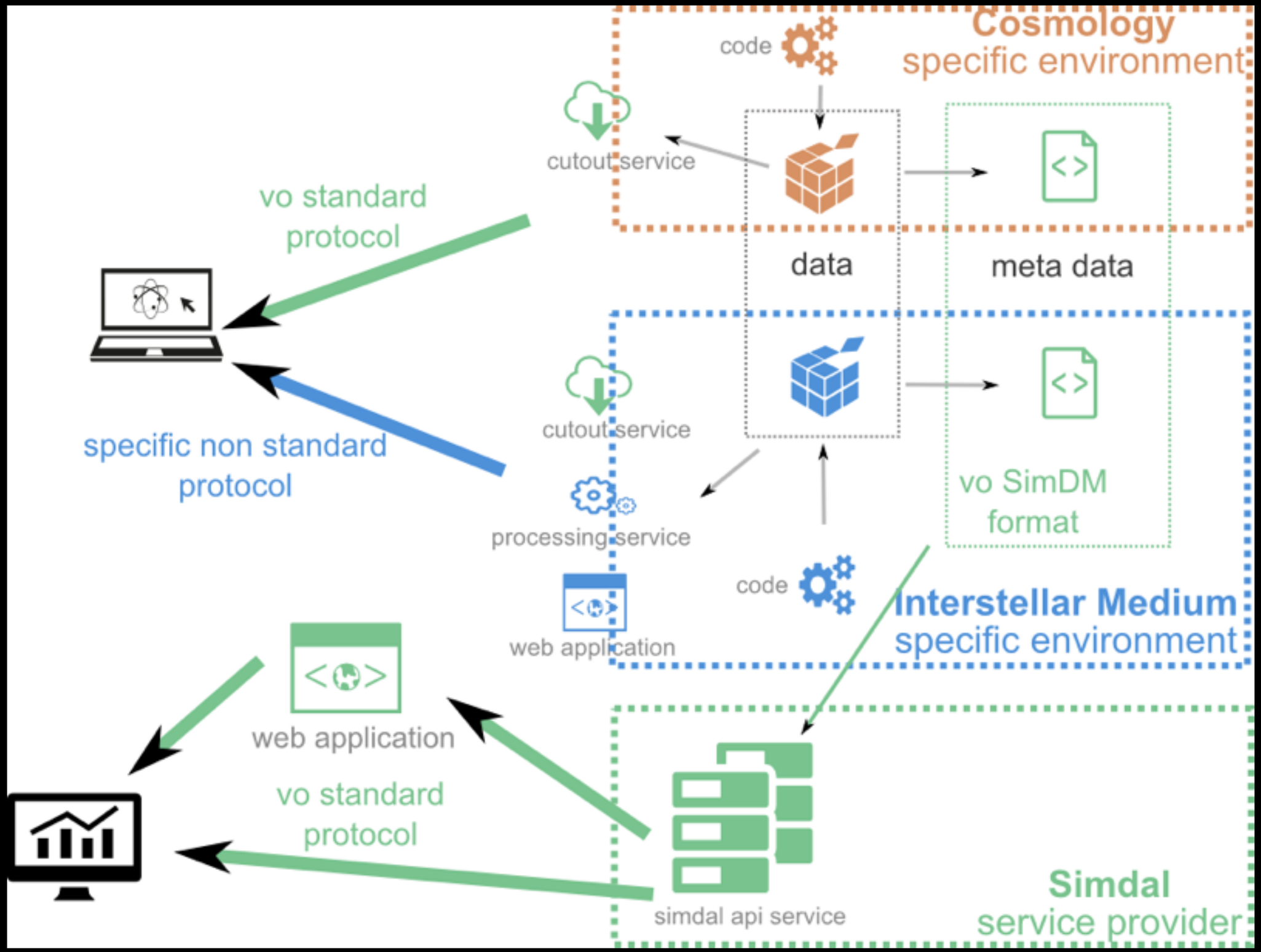
- axis based service allows **composite filters**

axis1 / (axis1 + axis2) < 3e12

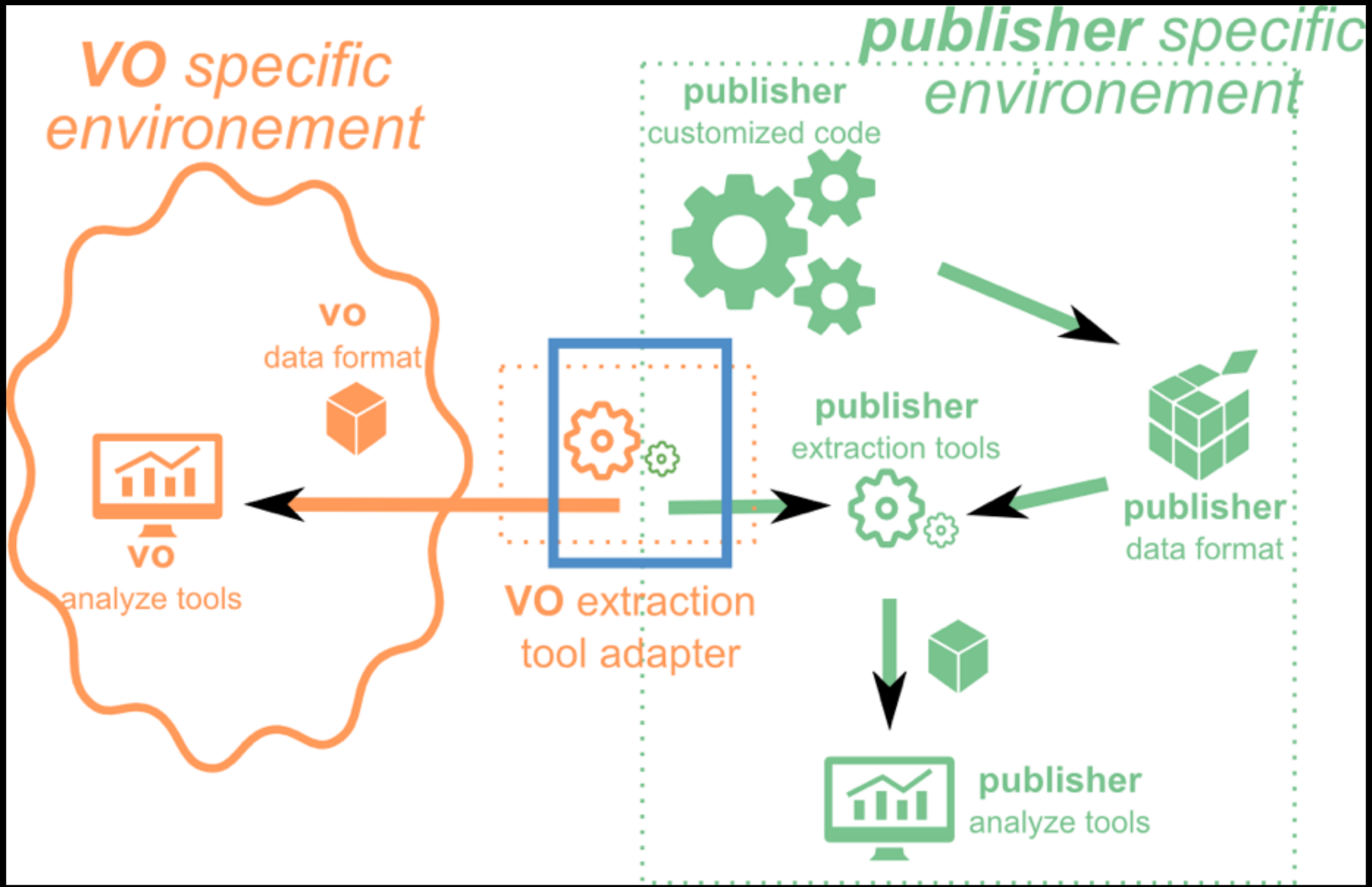
What goes where ?

- Registry
  - Centralized : Simulation Repository/Registry
- Search and Discovery
  - Centralized OR publisher managed
- Cutout
  - REST Api wrapper on top of publisher software

Ok, here is the sugar



How much work for publisher ?



Discuss

- Who goes where ?

Take-out

- Simulation **Registry** (Centralized Registry)
  - Help me find what I want to search for
- Simulation **Search** (Centralized / Distributed)
  - Help me search for the values
- Simulation **Data Access / Cutout** (Publisher service)
  - Help me access the raw data