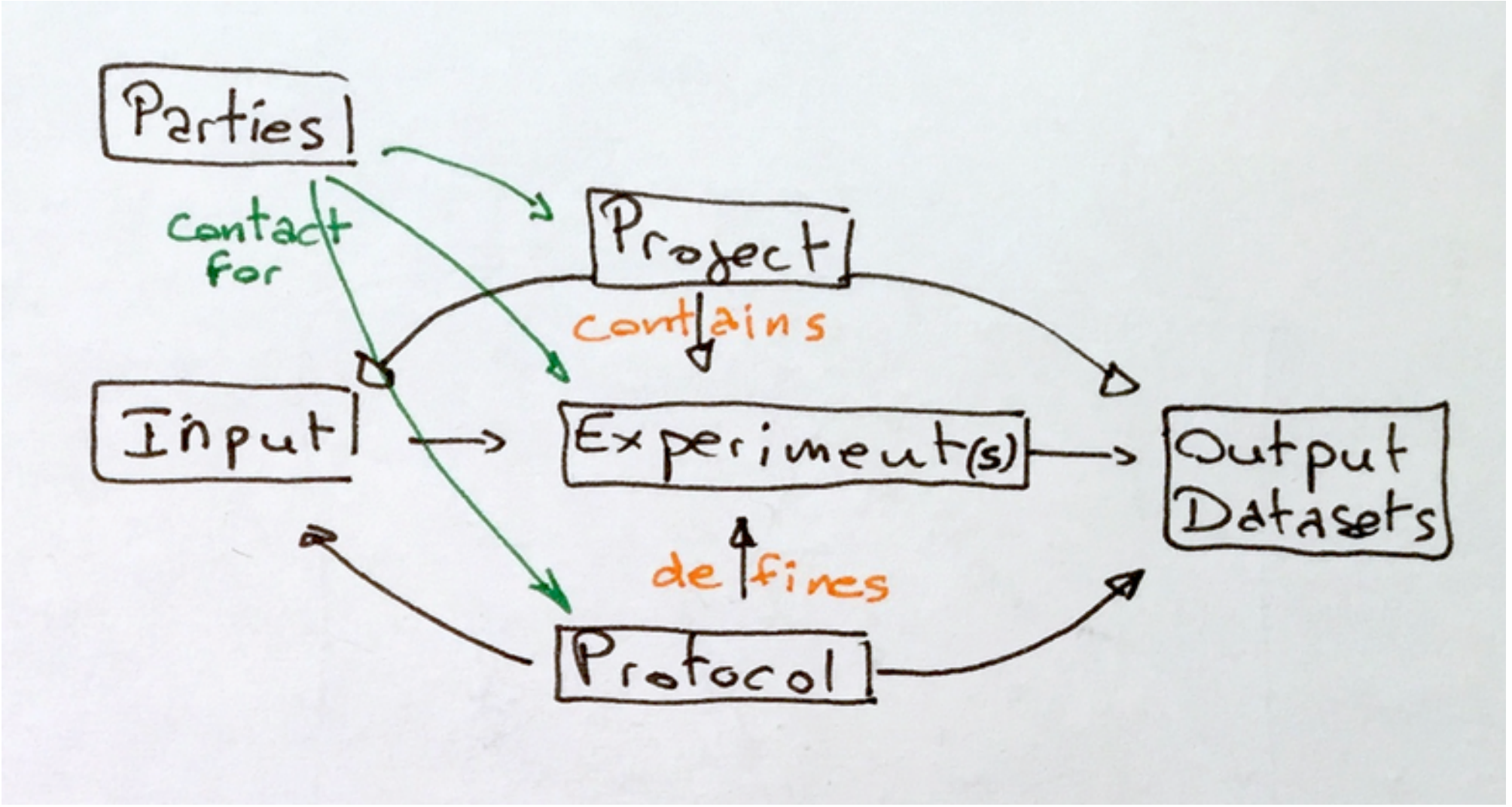
How does SimDM handle data provenance ?

Provenance Day

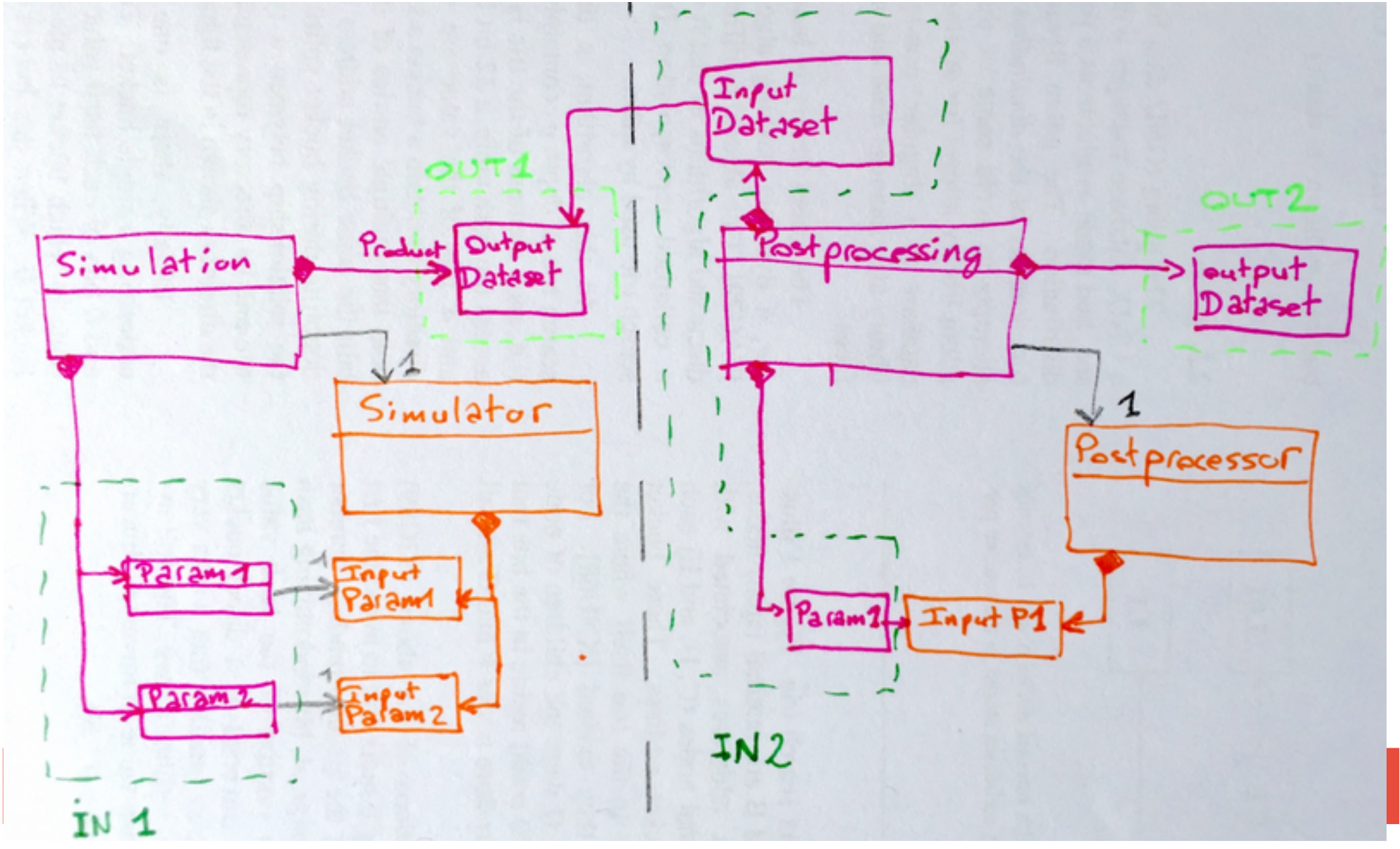
Strasbourg

2016.12.13

Basic simulation DM



Simulations chaining



The whole information is an XML set

- For now, only an XML serialization of SImDM is standard
- A project is a set of XML serializations of the SimDM packages
- The reference between XML files can be ether
  - Local to project namespace
  - Global through ivoid
- The API is any XML access method
  - Xpath query
  - Custom relational index from XML files with text search
  - Etc...

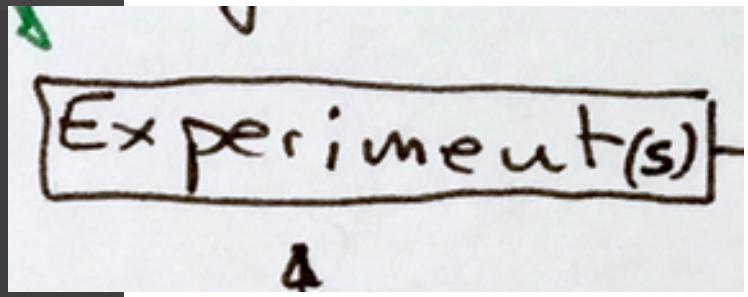
Resource metadata

- Experiment, Protocol and Project are Resources
- A Resource has a set of common metadata

```
<?xml version="1.0" ?>
<ns2:aSimulation xmlns:ns2="http://www.ivoa.net/xml/SimDM/v1.0">

    <identity publisherDID="DM51NoPAH_3_20"/>
    <name>DM51NoPAH_3_20</name>
    <description/>
    <referenceURL/>
    <created>2016-01-20T14:33:17.285163</created>

    <contact>
        <identity publisherDID="n/a"/>
        <role>owner</role>
        <party publisherDID="n/a"/>
    </contact>
```

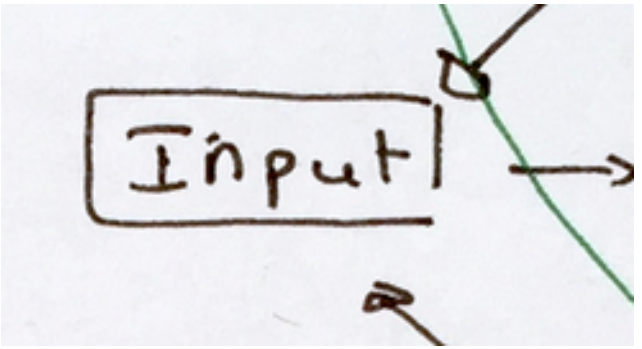


Experiment metadata

- Basically, an Experiment has metadata about
  - Its input
  - Its produced data (output datasets)

```
<parameter>
    <numericValue>
        <value>2.000000e+01</value>
        <unit>mag</unit>
    </numericValue>
    <inputParameter publisherDID="avmax"/>
</parameter>
```

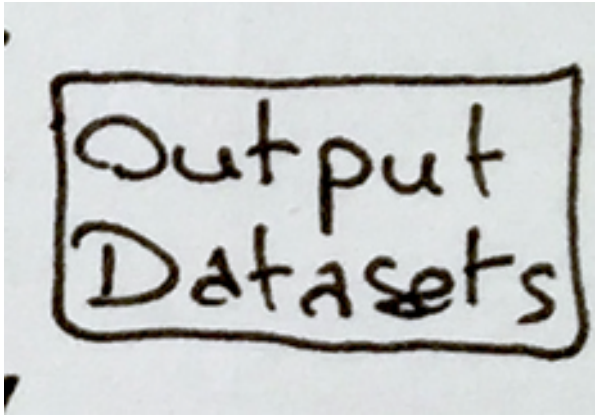
ID to link Protocol XML file



Experiment metadata

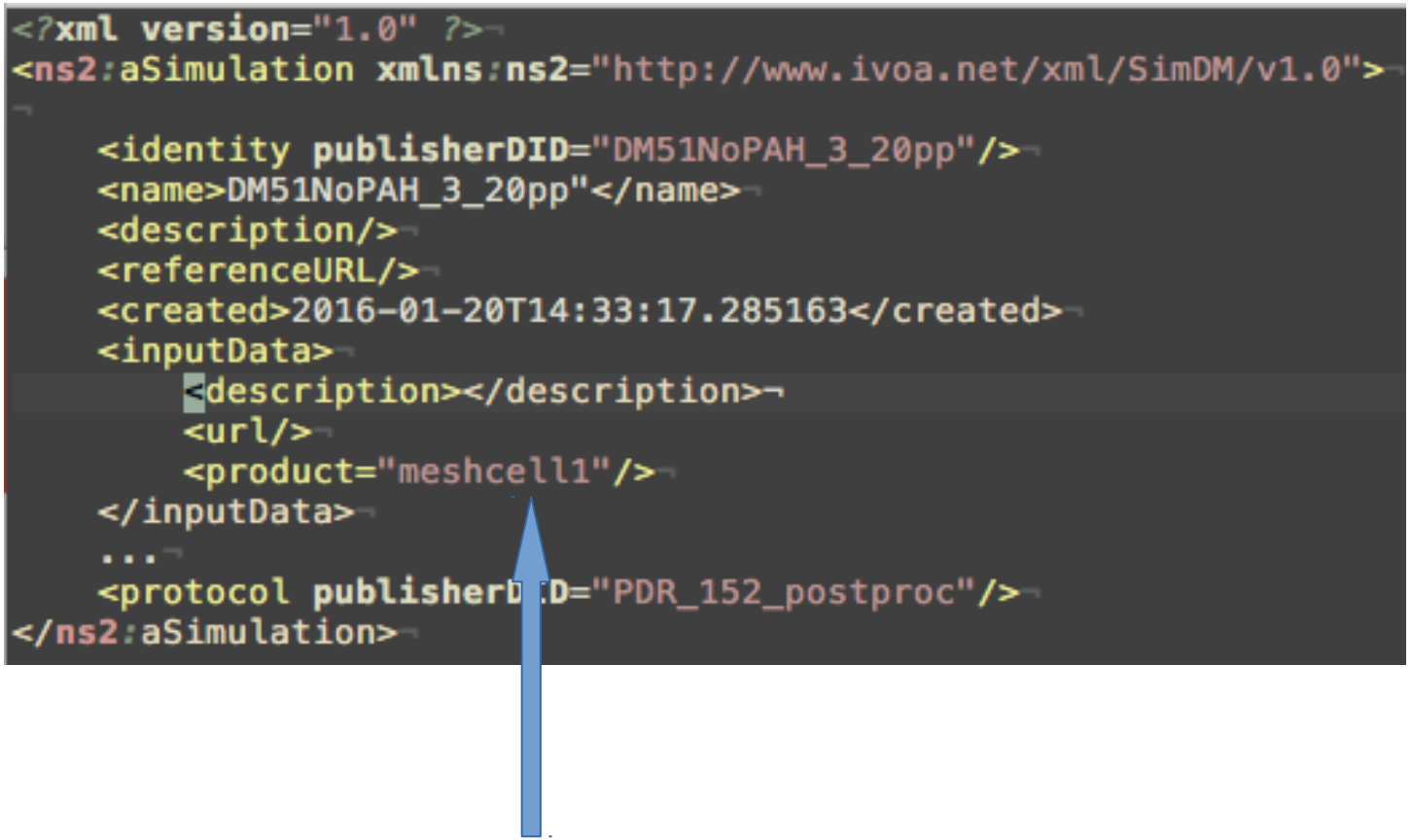
```
<outputData>
    <identity publisherDID="meshcell1"/>
    <numberOfObjects>871</numberOfObjects>
    <accessURL/>
    <characterisation>
        <statistic>min</statistic>
        <numericValue>
            <value>2.064447e+05</value>
            <unit>cm-3</unit>
        </numericValue>
        <axis publisherDID="protdens"/>
    </characterisation>
    <objectType publisherDID="meshcell"/>

</outputData>
<protocol publisherDID="PDR_1[illegible]2_1[illegible]28_mathis"/>
```



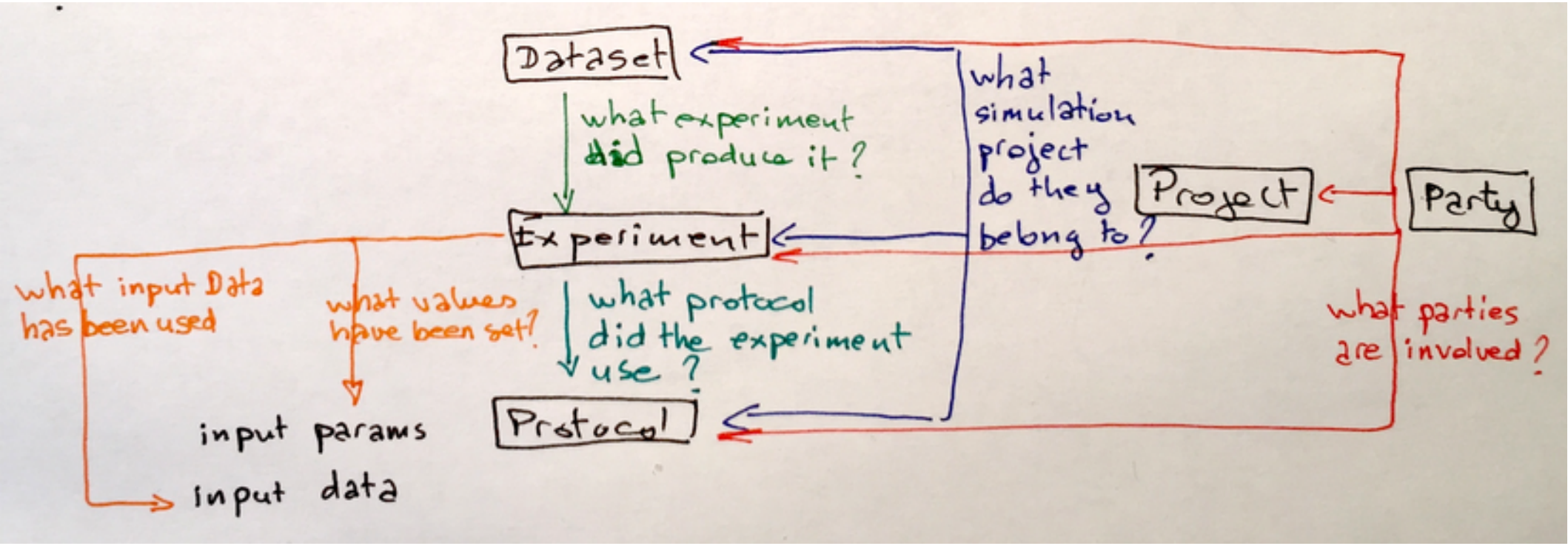
ID to link Protocol XML file

Experiment metadata (chaining)



ID to link output in another Experiment XML file

Where does this dataset come from ?



Take away

- **You can pick only what you want from SimDM**
  - And get very simple data model
- **Basic provenance metadata included in SimDM**
  - Included simulation chaining
- **The only one API to interact with SimDM is XML query**
  - The only one standard serialization is XML
  - Room to develop more user-friendly/use-case-specific UI
  - Not simple to do generic tool because of high abstraction of parts of SimDM. In particular EAV pattern (need attached vocabulary/semantic)