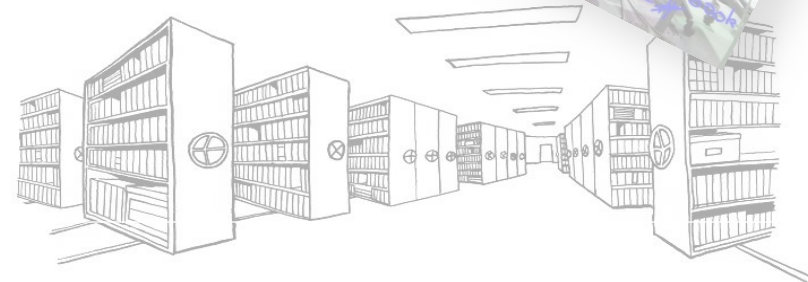
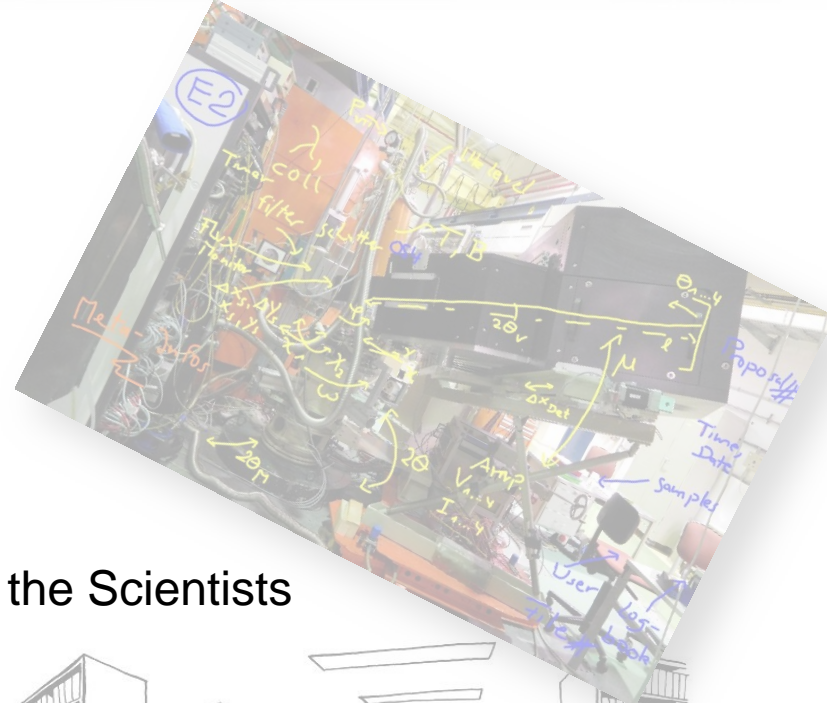


# Data curation at HZB

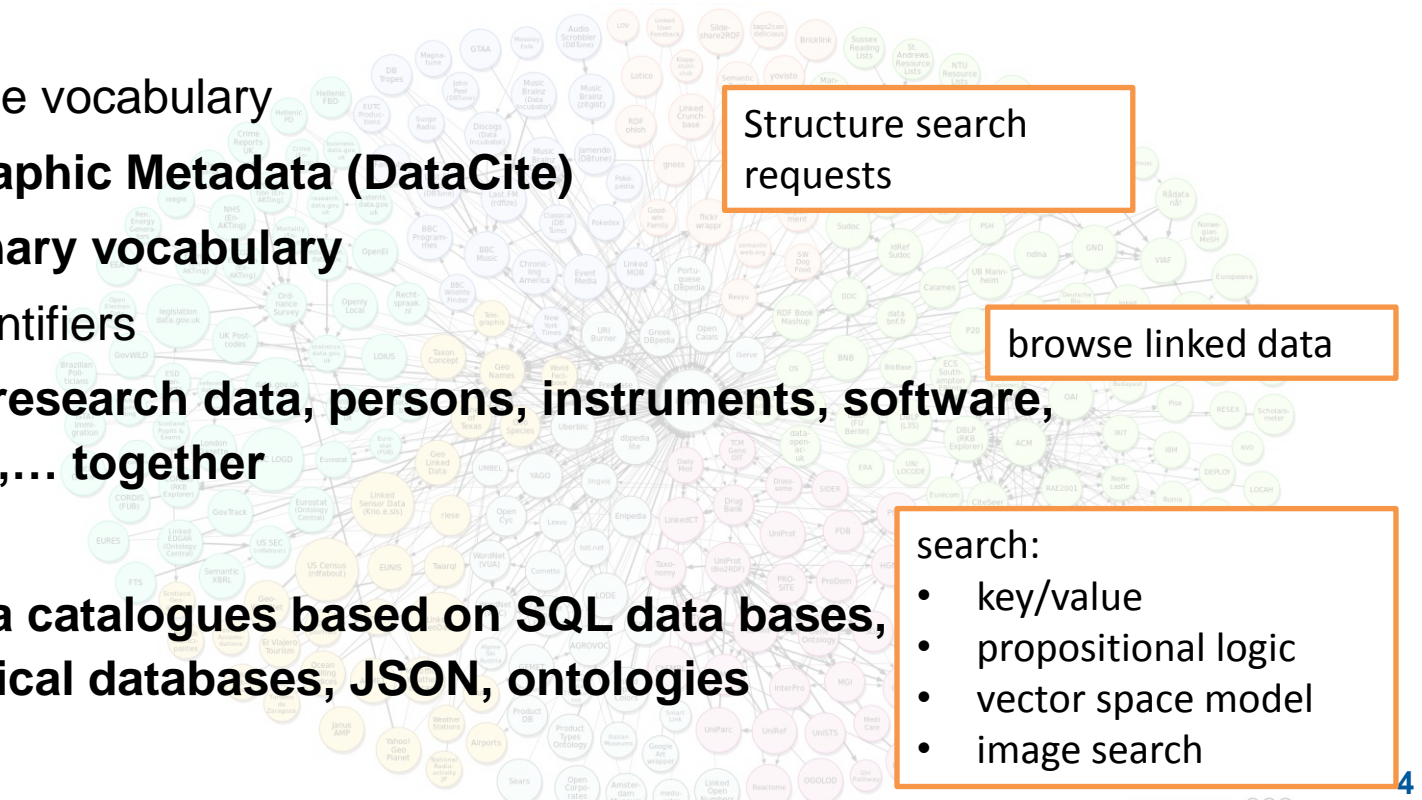
- FAIR Data Principles
  - **What metadata are required**
  - **What metadata standards help us**
- RDM approach at HZB IT
- Implementations
- What IT can offer  $\leftrightarrow$  What we need from the Scientists



## EXCHANGE AND ARCHIVE

- Requirements described in FAIR Data Principles
  - **Findable, Accessible, Interoperable, Re-usable**
- Prerequisite for data exchange:
  - **standardisation and usage on agreed conventions:  
formats, vocabulary, units**
  - **extensive metadata**

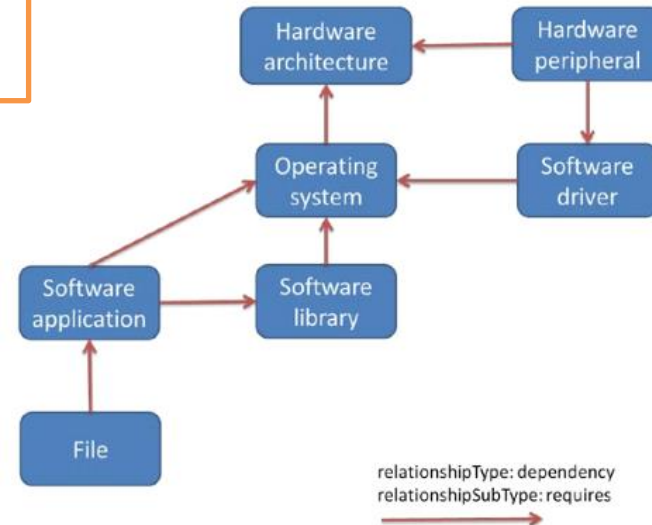
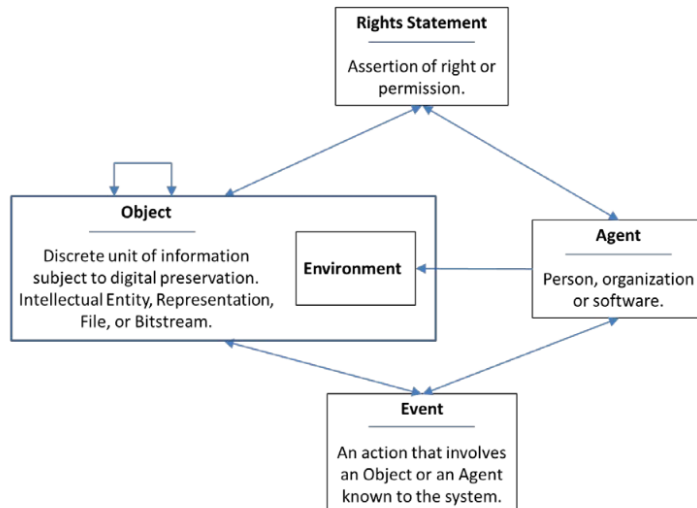
# SEARCH, SORT, FILTER, QUERY, BROWSE

- Data catalogue vocabulary
    - **Bibliographic Metadata (DataCite)**
    - **Disciplinary vocabulary**
  - Persistent identifiers
    - **Linking research data, persons, instruments, software, samples,... together**
  - Technologies
    - **Metadata catalogues based on SQL data bases, geometrical databases, JSON, ontologies**
- Structure search requests
- browse linked data
- search:
- key/value
  - propositional logic
  - vector space model
  - image search
- 

# PROVENANCE, PERMISSIONS/LICENSE, HOW TO OPEN

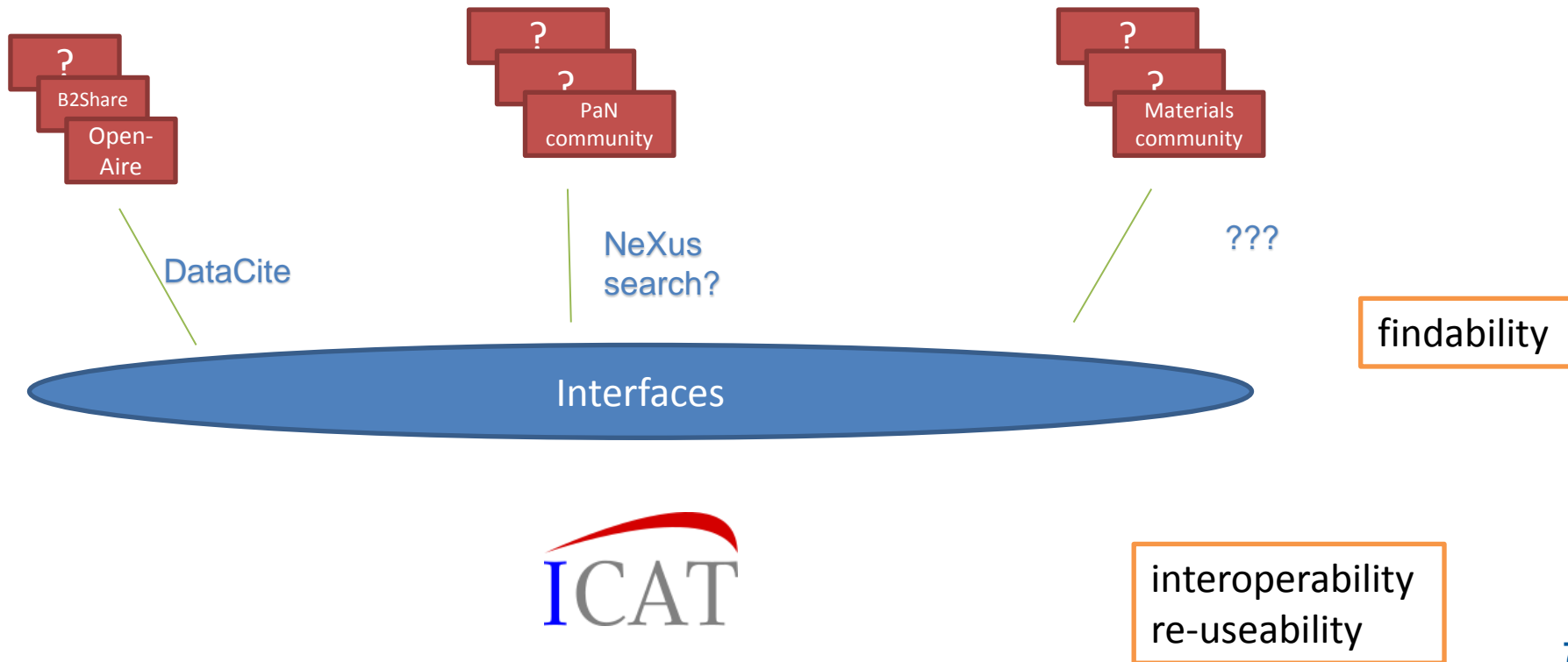
- Not disciplinary
- PREMIS

Trust the data you find  
and know how to use it



## OPEN, VIEW, UNDERSTAND, (RE-)USE, CHANGE, AUTOMATION

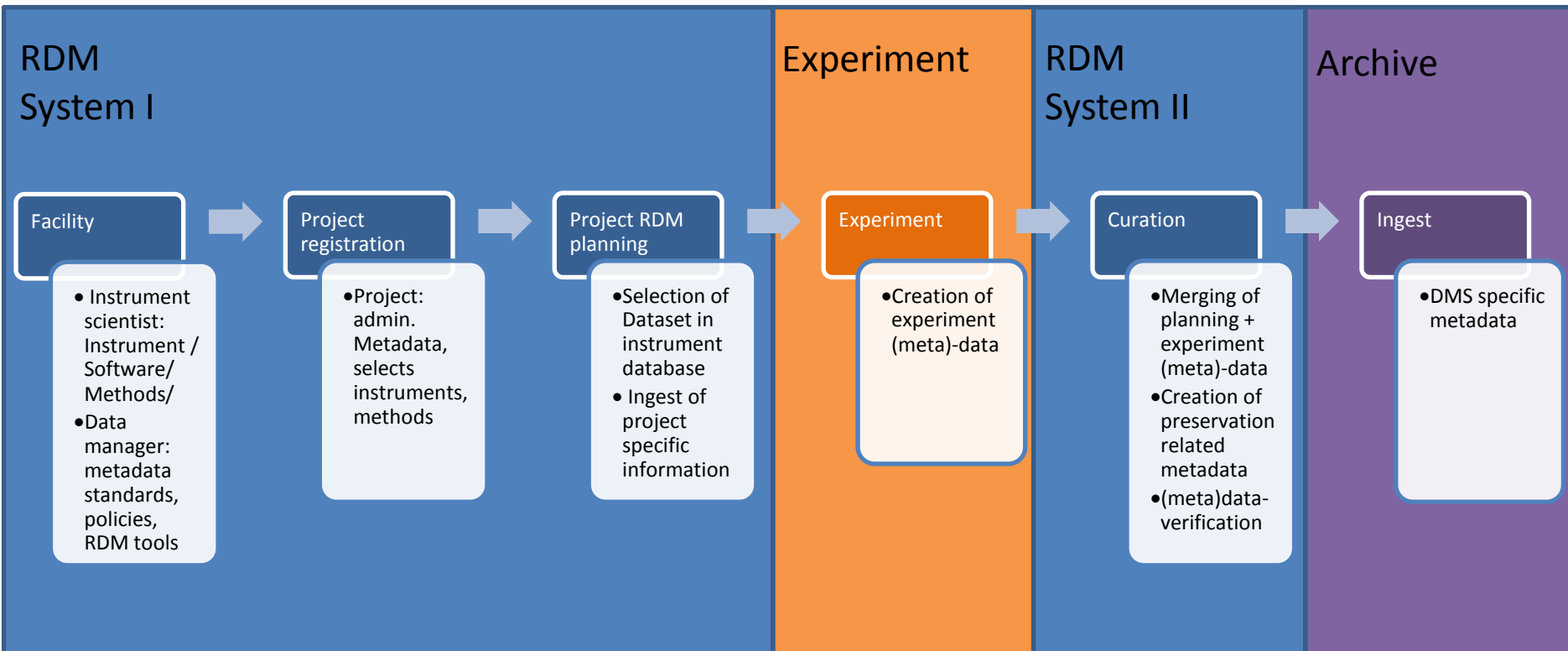
- What is required to use your experiment data for analysis?
- Is this enough information to understand your experiment?
- What else is required to simulate or reproduce the experiment?
  
- For light sources we can use NeXus
- For materials is it NOMADS?
  - **What else is relevant?**



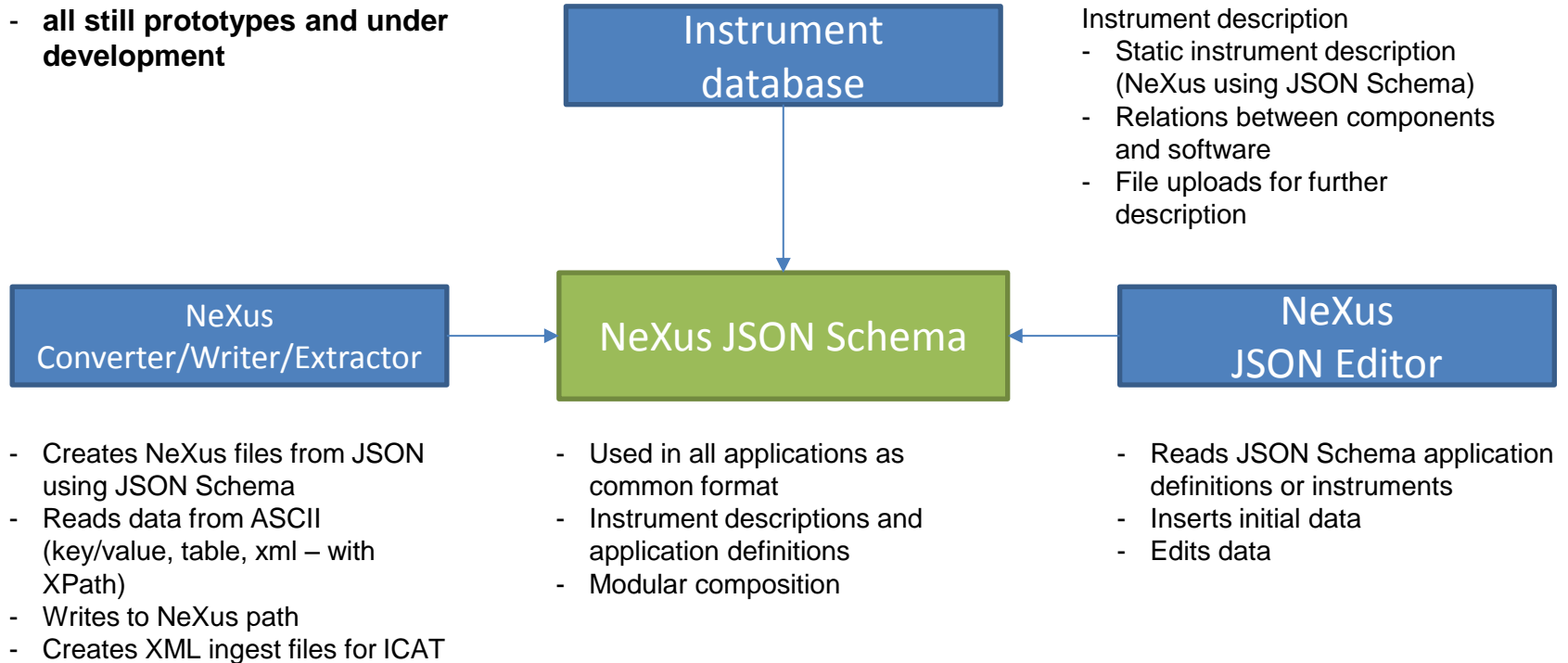
## RE-USE INFORMATION, USE STANDARDS AND AUTOMATE COLLECTION

- Re-use information about data: collect static, not so static and metadata about instruments and software
- Collect and share information about standards
- Automate collection as far as possible





- **all still prototypes and under development**



```
    "insertion_device": {  
      "name": "UE52",  
      "type": "undulator X-ray Source",  
      "@NX_class": "NXinsertion_device",  
      "@short_name": "UE52"  
    },  
    "monochromator": {  
      "local_name": "PGM",  
      "energy": {  
        "min": 85,  
        "max": 1600,  
        "@units": "eV"  
      }  
    },  
  },  
}
```

Required metadata  
standard, validation  
information

Instrum

Metho-  
den

Projekt/Proposal

Admin. data

Premis

Software

Datensätze

Sonstiges (Dateien, verantwortliche Personen)

related information

The screenshot displays the PyDev Package Explorer on the left and a code editor on the right. The Package Explorer shows a project structure for 'nex2icat [metadata-utils master]'. A sub-package 'config' is highlighted with an orange box, containing a 'resources' sub-package. Under 'resources', there are two sub-packages: 'appDefs' and 'UE52\_NanoClusterTrap'. 'appDefs' contains 'NXxas-v1.0.0.json' and 'UE52\_NanoClusterTrap' contains 'instrument.json', 'key\_valueData.json', 'tableData.json', and 'xmlData.json'. The code editor on the right shows the contents of 'NXxas-v1.0.0.json', with lines 1 through 4 highlighted by an orange box:

```
1 [DEFAULT]
2 level=Test
3 experiment=UE52_NanoClusterTrap
4 appDefs=NXxas-v1.0.0,NXreftof-v1.0.0
5
6
7
```

NeXpy v0.10.10

File Edit Data View Magic Window Script Help

NeXus Data

- ▼ Dy01Cp02\_024
  - entry
    - ▶ NXxas
    - ▶ comment
    - ▶ control
    - experiment\_identifier
    - ▼ instrument
      - ▶ detector1
      - ▶ detector2
      - ▼ detector3
        - time\_of\_flight
        - type
        - ▶ insertion
        - ▶ monochromator
        - name
        - ▶ slitwidth
        - ▶ source
        - ▶ intensityIntegrationProcess
        - program\_name
        - ▶ related\_files
        - ▶ sample
        - start\_time
        - ▶ user

detector3:NXdetector  
 @long\_name = 'reflectron time-of-flight mass spectrometer'  
 @short\_name = 'reftof'  
 time\_of\_flight = int16(119x119989)  
 type = 'reflectron time-of-flight mass spectrometer'

index\_1

100  
40  
20  
0

0 20000 40000 60000 80000 100000 120000

time\_of\_flight

signal x y projections options

🏠 ➕ 🔍 ⚙️ 📄 📊 📌 ✅ 🗑️ ➕

```

Jupyter QtConsole 4.3.1
Python 3.6.7 (default, Oct 22 2018, 11:32:17)
Type 'copyright', 'credits' or 'license' for more information
IPython 6.5.0 -- An enhanced Interactive Python. Type '?' for help.
In [1]:
    
```

- Loads fixed values from instrument database (to be implemented)
- Can be edited before measurement

## NeXEdi

Select schema | Select initial values | Download | Restore to Default | Disable/Enable Form | ValidateForm | valid

### UE52 Nanocluster Trap Properties

Basic | Instrument | Sample | Monitor | User

**title**

**start\_time**

**experiment\_identifier**

Unique identifier for the experiment, defined by the facility, possibly linked to the proposals

**experiment\_description**

Brief summary of the experiment, including key objectives.

**definition**

NXxas

**id string**

NXxas

## NeXEdi

Select schema | Select initial values | Download | Restore to Default | Disable/Enable Form | ValidateForm | valid

### NXxas Properties

Basic | Instrument | sample | monitor

**Instrument** Properties

Basic | NXmonochromator | Incoming Beam (NXdetector) | Absorbed Beam (NXdetector)

**NXmonochromator** Properties

Basic | Crystal | Grating

**Local name**

PGM

local name of monochromator.

**Energy** Properties

energy.

**min**

85

**max**

1600

**@units**

eV

**NX\_class**

NXmonochromator

NeXus class. Fixed value

- RDMatDB project with HZDR
  - **NeXus paths as parameter names for ICAT**
- Key words
- Namespaces
  - **nxs/**

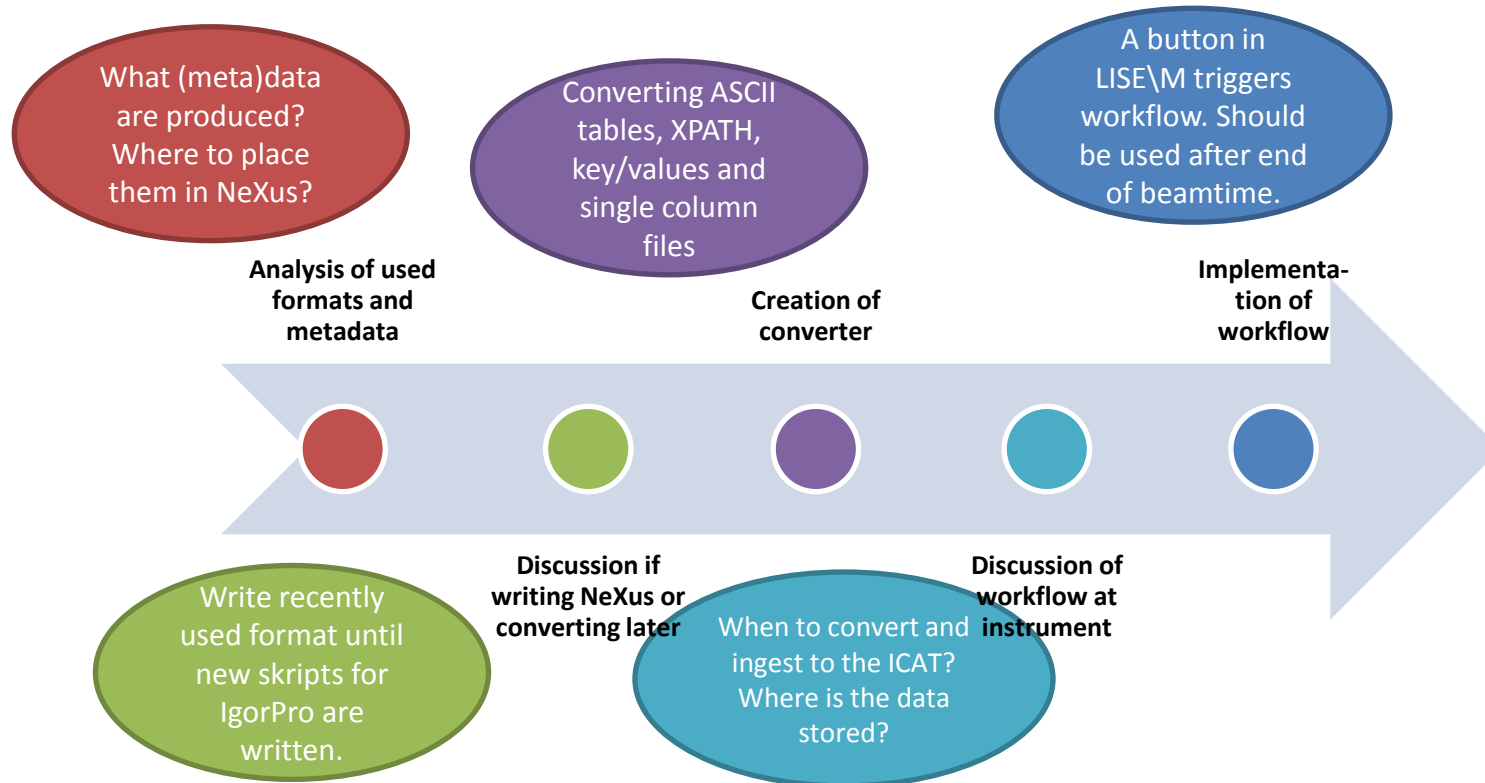
Are there other?  
May be facilitate  
ingest to NOMAD

```

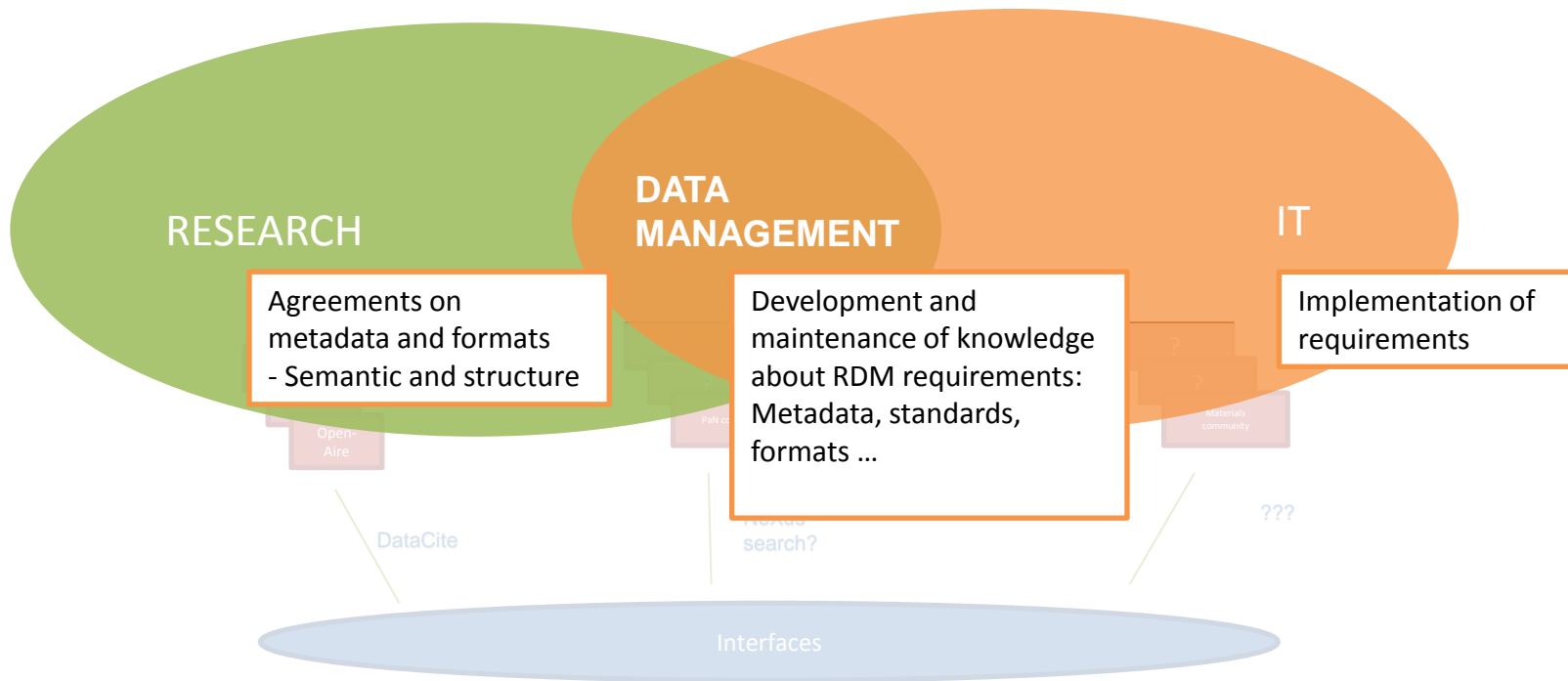
- <datasetParameter>
  <stringValue>incoming_beam</stringValue>
  <dataset ref="Dataset_1"/>
  <type name="nxs/entry/instrument/detector1/@long_name"/>
</datasetParameter>
- <datasetParameter>
  <stringValue>xas1</stringValue>
  <dataset ref="Dataset_1"/>
  <type name="nxs/entry/instrument/detector1/@short_name"/>
</datasetParameter>
- <datasetParameter>
  <numericValue>3.385958474576271e-05</numericValue>
  <rangeBottom>2.7892e-05</rangeBottom>
  <rangeTop>4.1027e-05</rangeTop>
  <dataset ref="Dataset_1"/>
  <type name="nxs/entry/instrument/detector1/data" units="A"/>
</datasetParameter>
- <datasetParameter>
  <stringValue>photo diode</stringValue>
  <dataset ref="Dataset_1"/>
  <type name="nxs/entry/instrument/detector1/type"/>
</datasetParameter>
- <datasetParameter>
  <stringValue>absorbed_beam</stringValue>
  <dataset ref="Dataset_1"/>
  <type name="nxs/entry/instrument/detector2/@long_name"/>
</datasetParameter>

```

# WORKFLOW FOR CREATING CONNECTION







**THANK YOU !**