Metadata for material science at the Lightsource BESSY II

T. Birke[1],H. Görzig[1], V. Laux[1], T. Mertens[1],R. Muller[1], M. Ries[1], A. Schaelicke[1], P. Schnizer[1], T. Unold[1], L. Vera Ramirez[1], J. Viefhaus[1]

[1]Helmholtz-Zentrum Berlin (HZB), Germany

11/07/2019

Table of Contents

Table of Contents

BESSY II – STORAGE RING

HELMHOLTZ
RESEARCH FOR GRAND CHALLENGES

HZB Helmholtz Zentrum Berlin

- in user operation since 1998
- “compact”, high ratio of straight sections, ~ 38 (27) beamlines
- diverse user community
- offering short pulses

parameters	
Energy	1.7 GeV
Circumference	240 m
Horizontal emittance	5 nm rad
Beam current	300 mA
RF frequency	500 MHz
max. RF voltage	2 MV
Bunch length low-α	15 ps 2 ps
Mom. Comp. factor low-α	7.5×10^{-4} 3.5×10^{-5}

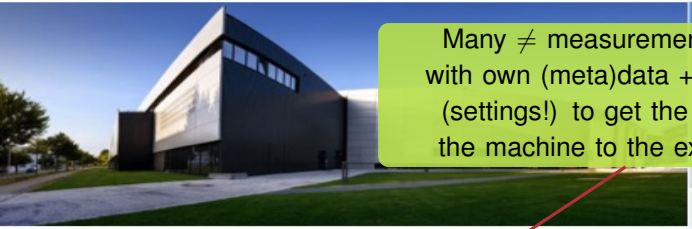
BESSY II: a 3rd generation light source optimized for soft x-ray range

courtesy of M. Ries

BESSY II – STORAGE RING

HELMHOLTZ
RESEARCH FOR GRAND CHALLENGES

HZB Helmholtz Zentrum Berlin



Many ≠ measurement devices with own (meta)data + beamlines (settings!) to get the light from the machine to the experiment.

- in user operation since 1998
- "compact", high ratio of straight sections, ~ 38 (27) beamlines
- diverse user community
- offering short pulses

parameters	
	1.7 GeV
ce	240 m
emittance	5 nm rad
Beam current	300 mA
RF frequency	500 MHz
max. RF voltage	2 MV
Bunch length low-α	15 ps 2 ps
Mom. Comp. factor low-α	7.5 ×10↑−4 3.5 ×10↑−5

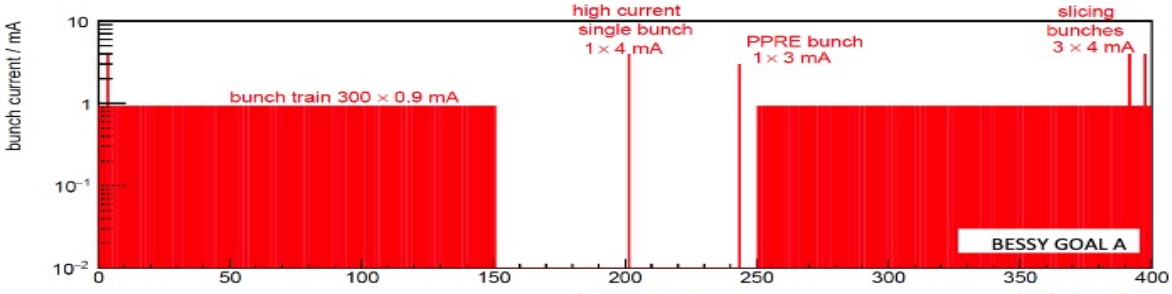
BESSY II: a 3rd generation light source optimized for soft x-ray range

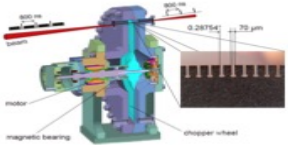
courtesy of M. Ries

BESSY II – MULTI USER MACHINE

HELMHOLTZ
RESEARCH FOR GRAND CHALLENGES

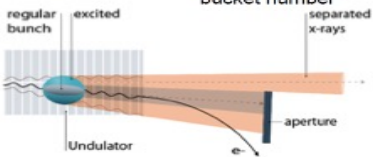
HZB Helmholtz Zentrum Berlin





Talk P. Goslawski

Tour T. Mertens



simultaneously serving many users applying many different techniques

courtesy of M. Ries

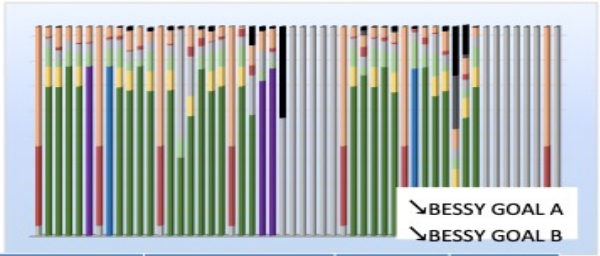
BESSY II – OPERATION

HELMHOLTZ RESEARCH FOR GRAND CHALLENGES

HZB Helmholtz Zentrum Berlin

- 24/7 operation
- on average:
  4600 h / year with 96 % ± 2% availability
- positive trend trough systematic identification of root causes
- common metrics application

Tour A. Schälicke



Year	Scheduled	Downtime	# Outages	Availability	MTBF	MTTR
2013	4505 h	159.3 h	105	96.5%	42.9 h	1.52 h
2014	5408 h	384.4 h	136	92.9%	39.8 h	2.83 h
2015	3896 h	92.5 h	90	97.6%	43.3 h	1.03 h
2016	4855 h	62.9 h	69	98.7%	70.4 h	0.91 h
2017	4290 h	241.5 h	62	94.4%	69.2 h	3.90 h

BESSY II features high availability and reliability

courtesy of M. Ries

ROLLING UPGRADES – EMIL

HELMHOLTZ RESEARCH FOR GRAND CHALLENGES

HZB Helmholtz Zentrum Berlin

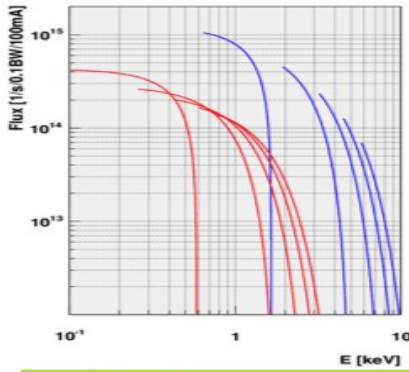
integrate a canted undulators scheme following user request

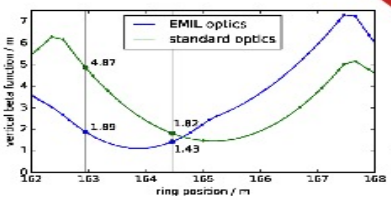
↘BESSY GOAL A
↘BESSY GOAL B

ENGERY

Tour T. Mertens

Tour C. Kuhn





Monochromator Section

PINK

SISSY-1

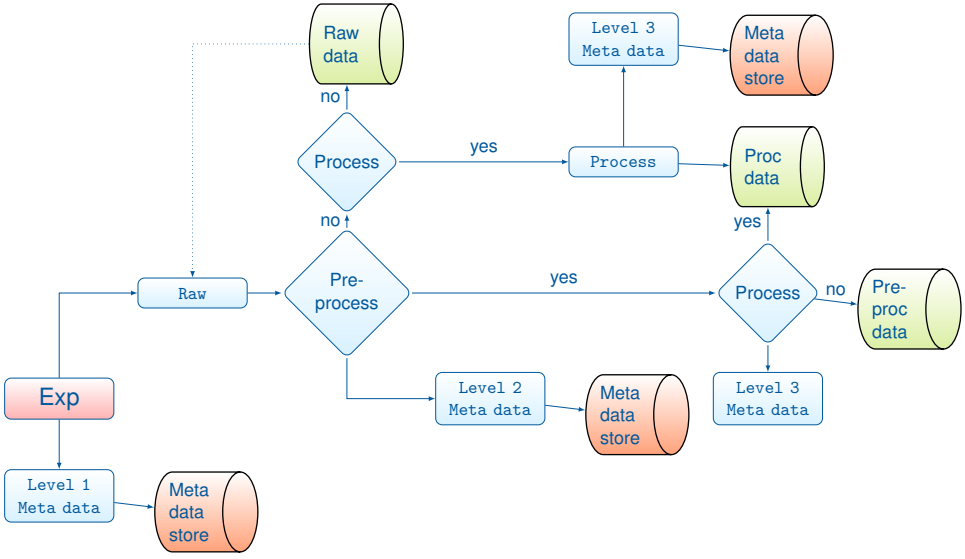
PEEM

flexibility: constantly evolving following the needs of the user community

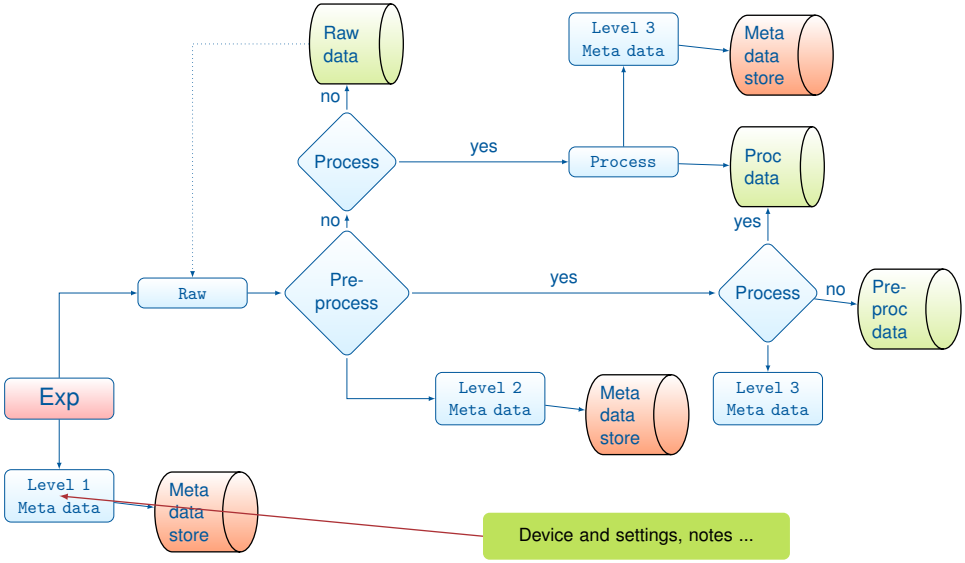
courtesy of M. Ries

Table of Contents

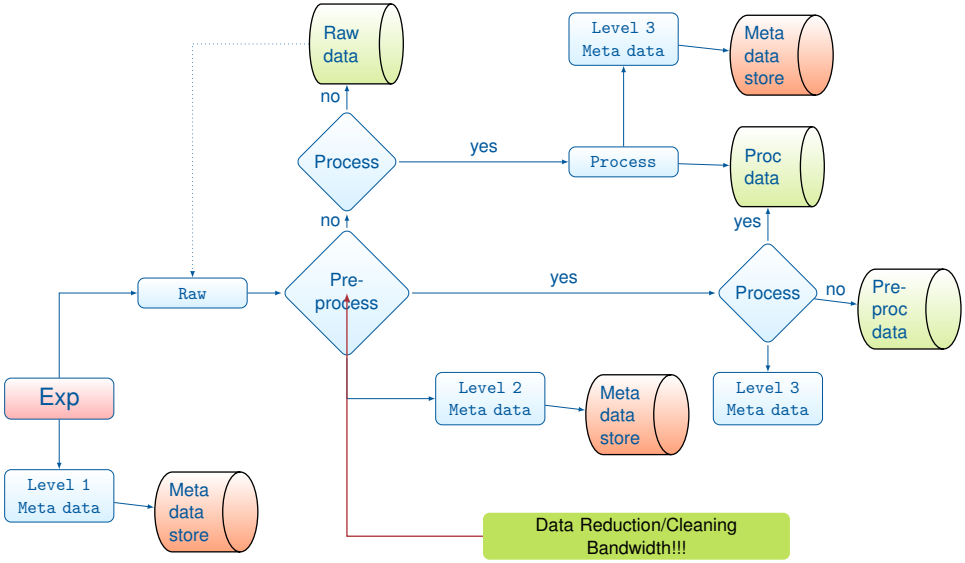
Data flow schematic



Data flow schematic



Data flow schematic



Data flow schematic

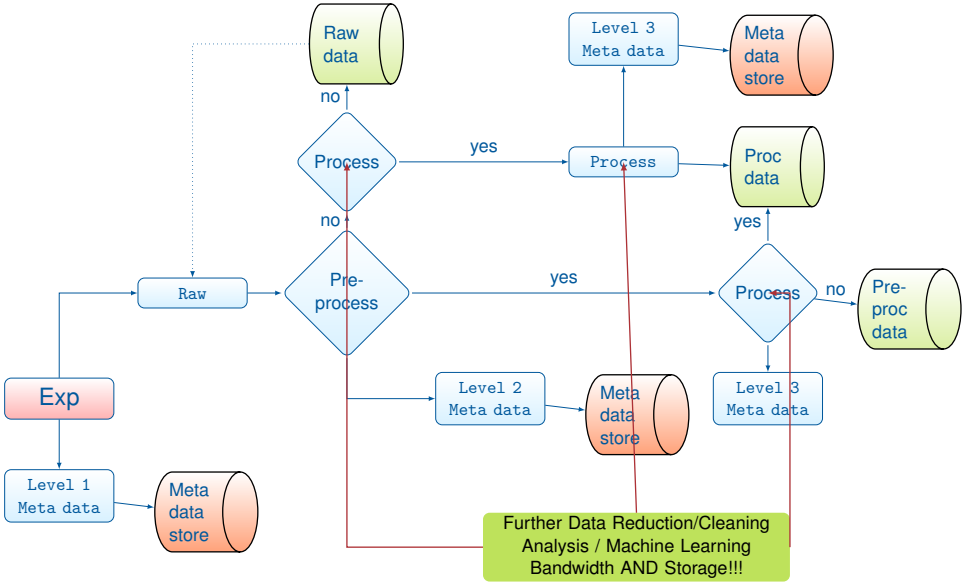


Table of Contents

Some (meta)data considerations

- What do we consider as metadata?
- Auto-generated? → Most
- Manual? → Some
  - Digital (elog entries)
  - Analog (handwritten)
- Data lives! → Persistent IDentification of data (unique, versioned,...)
- Logbook interface → Sometimes
- Storage: SQL/noSQL, file formats → NEXUS + mongodb + ElasticSearch
- Scalability: can we handle the future? → Elastic Stack + NEXUS?
- **Data Model?**
- Findability vs amount of metadata

Some side notes (issues acc phys):

- Accelerator Physics is experience based (an Art)

  (https://www.youtube.com/watch?v=A1L2xODZSI4)
- Trains, cell phones, cars, ...
- Running software on different systems
- Version control
- File formats
- Maintainability of packages/tools (cookiecutter, code certificates,)
- Containerization : **Singularity** with a SCIentific File System (SCIF)

Table of Contents

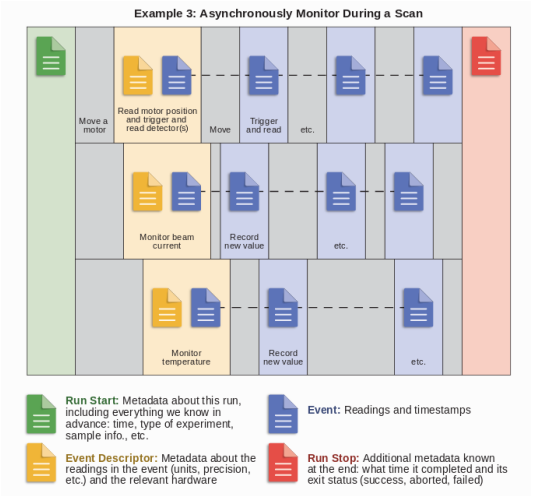
BESSY II - the machine

- Control System BESSY: Epics 3.14 (3.15 under way)
- Epics variables with unique naming convention (location, device type, etc...) **metadata**
- all channels logged in archiver
- accelerator metadata = data
- Question: which data is metadata for some exp
- **Meaning meta(data)** depends on settings (changed manually and can **NOT** be inferred from data)

- We want: Near Real Time Simulation / Analysis
- Why? Performance optimization, fast recovery, machine protection and maintenance
- Experiments: beam commissioning

Example: Bluesky and Ophyd

- Python
- Ophyd for device abstraction (epics, labview, but also extendable)
- Bluesky for experiment control and planning
- Nice data model (see right)
- Databroker available (base: sqlite and MongoDb)
- Suitcase for elasticsearch developed in house (ohters can be easily produced)
- possibility to store data in external files but keep links to data in database (adaptors can be written in straightforward way to load/save the data)
- Generates unique ID for each experiment
- Can talk to Olog
- **METADATA** and data hints
- live plotting and fitting



Example bluesky metadata


```
[{'_id': 'OPDFjmoBqZm8A59lj93a',
  '_index': 'start_run',
  '_score': 3.9233167,
  '_source': {'detectors': ['bpm'],
              'hints': {'dimensions': [[['motor2'], 'primary'],
                                       [['master_clock_frequency_readback'],
                                        'primary']]},
              'motors': ['motor2', 'master_clock_frequency'],
              'num_intervals': 49,
              'num_points': 50,
              'plan_args': {'cycler': "(cycler(MasterClockFrequency(prefix='', "
                                      "name='master_clock_frequency', "
                                      "parent='master_clock', settle_time=0.0, "
                                      "timeout=2.0, read_attrs=['setpoint', "
                                      "'readback', 'offset'], "
                                      'configuration_attrs=[], limits=(499626, '
                                      "499634), egu='kHz'), [499623.43033, "
                                      '499623.65255222225, 499623.87477444444, '
                                      '499624.0969966667, 499624.3192188889, '
                                      '499624.5414411111, 499624.7636633333, '
                                      '499624.98588555556, 499625.20810777775, '
                                      '499625.43033]) * '
                                      "cycler(SynAxis(prefix='', "
                                      "name='motor2', read_attrs=['readback', "
                                      "'setpoint'], "
                                      "configuration_attrs=['velocity', "
                                      "'acceleration']), [0, 1, 2, 3, 4]))",
                            'detectors': ["BPMStorageRing(prefix='', "
                                          "name='bpm', read_attrs=['stat', "
                                          "'stat.mean_x', 'stat.mean_y', "
                                          "'stat.rms_x', 'stat.rms_y', "
                                          "'waveform', 'waveform.packed_data', "
                                          "'waveform.counter', "
                                          "'waveform.ready', 'waveform.pos_x', "
                                          "'waveform.pos_y', "
                                          "'waveform.intensity_z', "
                                          "'waveform.intensity_s', "
                                          "'waveform.status', 'waveform.gain', "
                                          "'waveform.rms_x', "
                                          "'waveform.rms_y'], "
                                          "configuration_attrs=['stat', "
                                          "'waveform'])"],
                            'per_step': 'None'},
              'plan_name': 'scan_nd',
              'plan_type': 'generator',
              'scan_id': 1,
```

Table of Contents

Beamlines - Let's complain a bit.

- Different Experiment Control Softwares with different ways to control Sample Environment
- Complex software interface protocols
- Time consuming integration of new equipment with little mobility.
- No metadata standards for Sample Environments (SECoP with NEXUS?)
- Physicists programming drivers.
- Human Readable variable names?
- No generic work-/dataflow - user/beamline specific, what about authentication?
- New "workflows" (related to (meta)data) should not change users habits or create extra overhead
- Metadata database for static metadata (see further)
- Experiment controls: EPICS, Tango, Sardana $\rightarrow$ Python?
- Common Data Model Architecture vs NoMaD $\rightarrow$ **Data Model?**

Table of Contents

HZB Helmholtz Zentrum Berlin

RE-USE INFORMATION, USE STANDARDS AND AUTOMATE COLLECTION

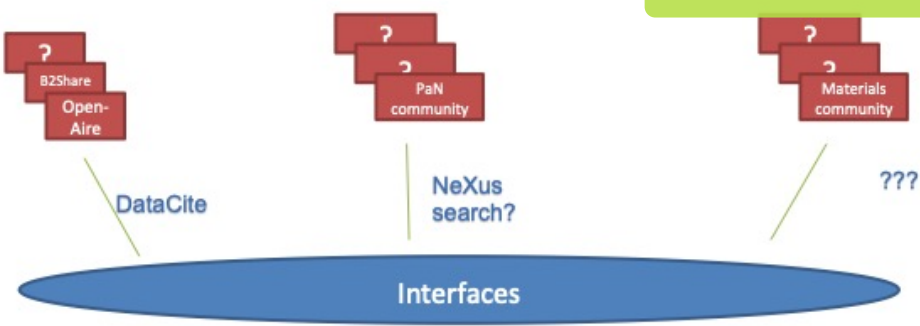
- Re-use information about data: collect static, not so static and metadata about instruments and software
- Collect and share information about standards
- Automate collection as far as possible

courtesy of H. Görzig

HOW DO WE GET THERE?

HZB Helmholtz Zentrum Berlin

ICAT : metadata catalogue



findability

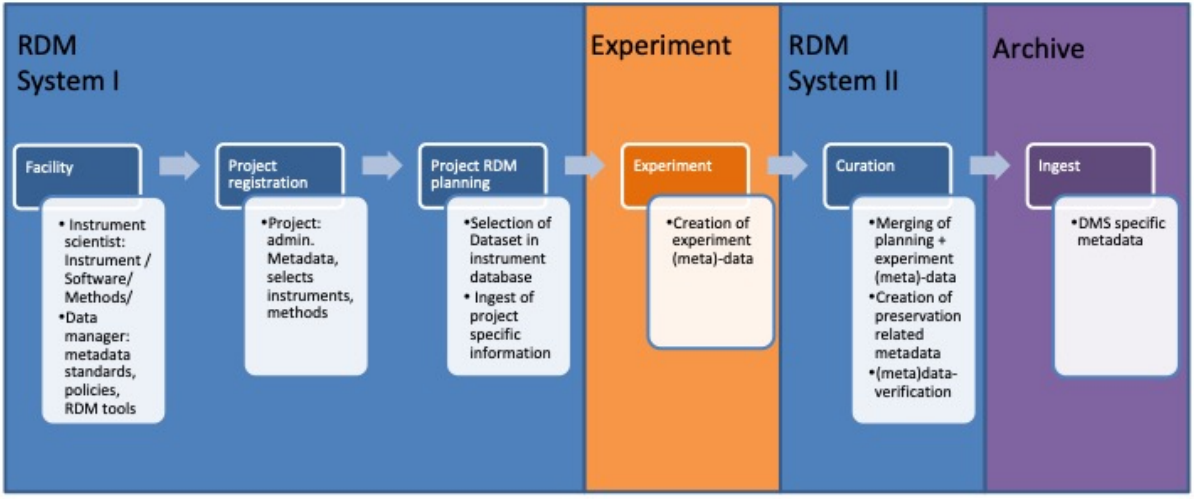
ICAT

interoperability
re-useability

courtesy of H. Görzig

WORKFLOW

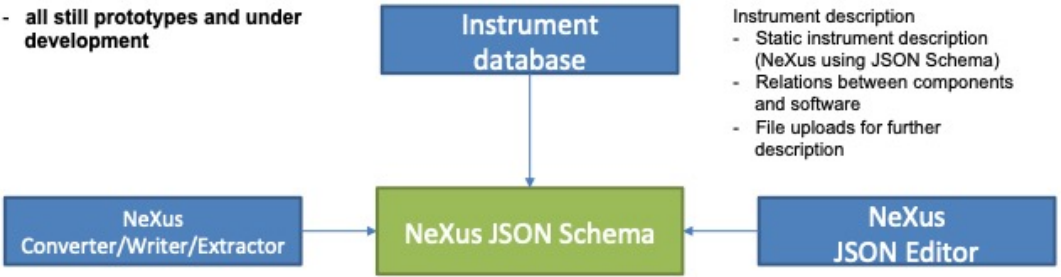
HZB Helmholtz Zentrum Berlin



courtesy of H. Görzig

HZB Helmholtz Zentrum Berlin

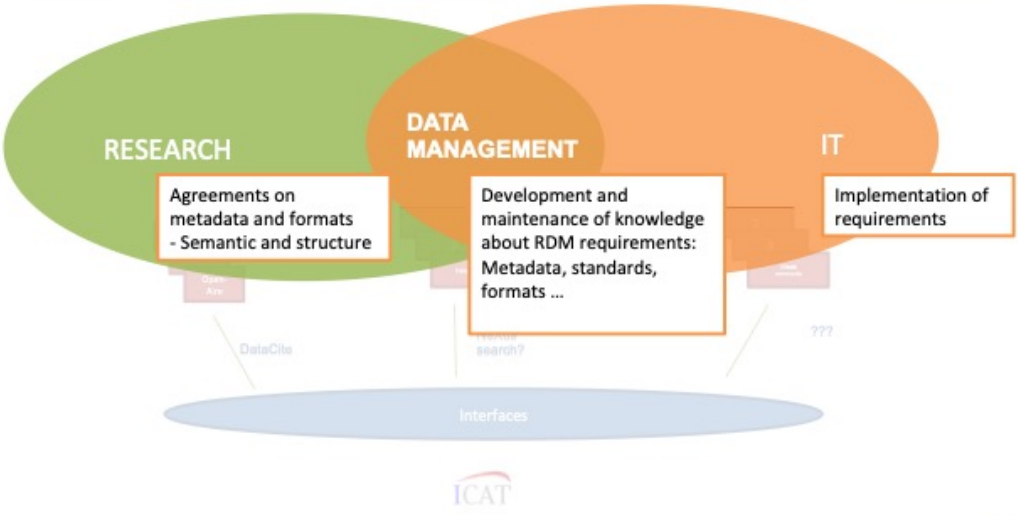
- **all still prototypes and under development**



courtesy of H. Görzig

HOW DO WE GET THERE?

HZB Helmholtz Zentrum Berlin



courtesy of H. Görzig

Table of Contents

Summary

- Many open questions.
- Initial work started both from machine side as from experiments side.
- EPICS with Bluesky (Python) + Archiver for the machine.
- ICAT database for static metadata under development.
- NEXUS interfaces for data/metadata under development.
- Some limitations can not be solved easily (bandwidth, standardization, storage, access rights,...)
- Working towards FAIR data.

Links

- Singularity :https://sylabs.io/docs/
- Bluesky, Ophyd :https://github.com/bluesky
- Scientific Filesytem :https://sci-f.github.io/