

Metadata for Experiments in Nanoscience Facilities

Vasily Bunakov

Science and Technology Facilities Council, UK



Workshop on Scientific Data Management for
Photon and Neutron Facilities, 19-20 March 2018

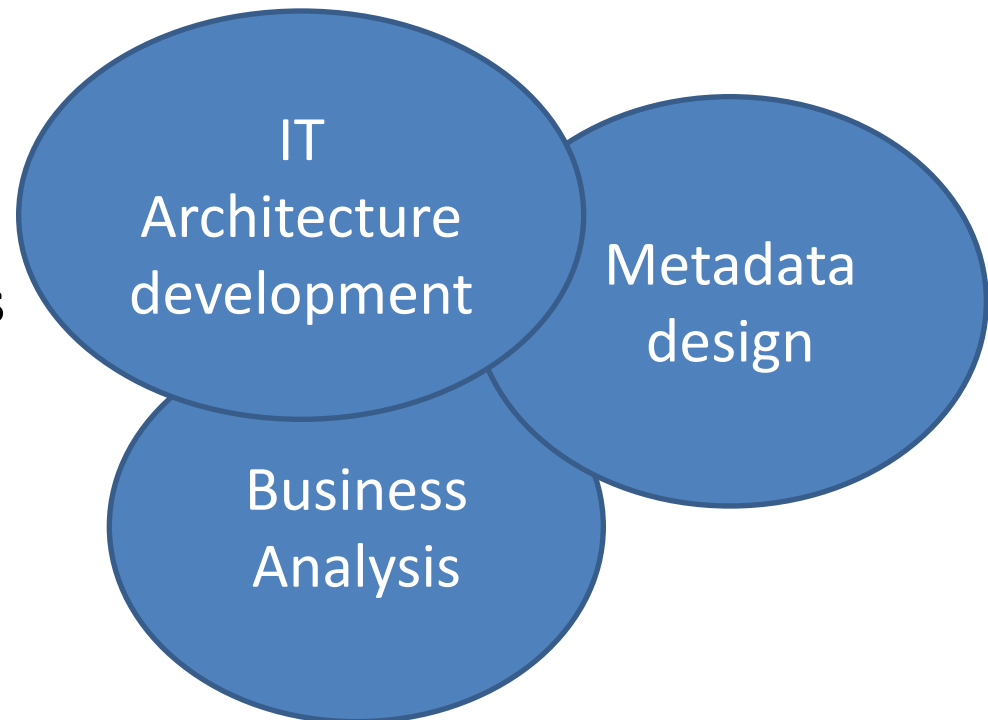
NFFA project



- Horizon 2020 project
- Distributed infrastructure to perform growth, nano-lithography, nano-characterization, theory and simulation and fine-analysis with synchrotron, FEL and neutron radiation sources
- “Virtual research enterprise” with proposals system and data management obligation

Three components of NFFA (virtual) Enterprise Architecture for data management

Enterprise architecture applies architecture principles and practices to guide **organizations** through the business, information, process, and technology changes necessary to execute their **strategies**. †

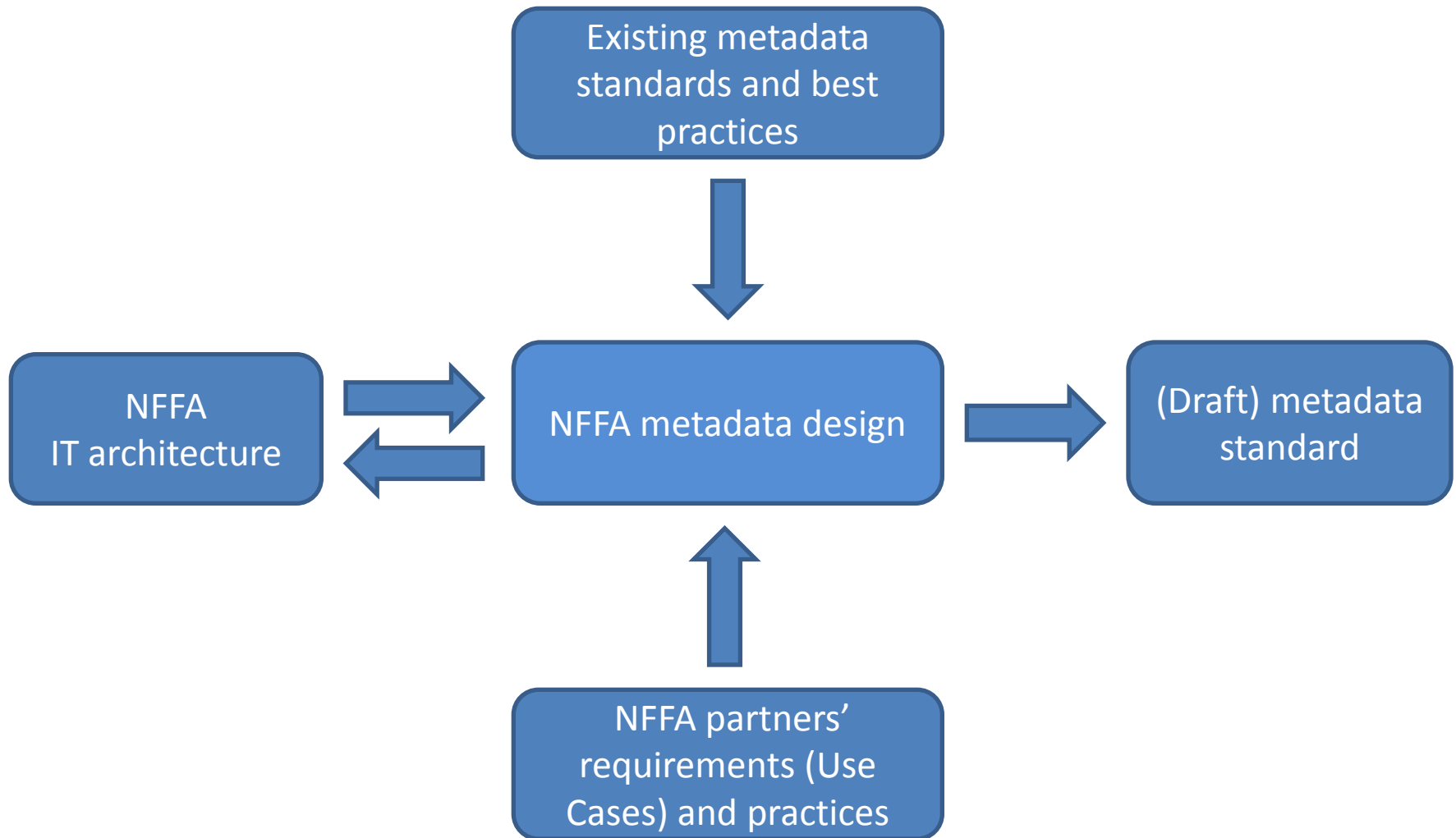


† https://en.wikipedia.org/wiki/Enterprise_architecture

Approach to metadata design

- Overall focus: data lifecycle in nano-facilities
- To satisfy both experimental and computational research
- Top-down route: consider existing data management recommendations and best practices
- Bottom-up route: consider NFFA partners Use Cases and practices
- “Neighbouring” route: consider NFFA IT Architecture
- Synthesis: amalgamate top-down and bottom-up with some “side” (IT architecture) considerations

Sources of metadata design



Metadata artefacts and a potential for their common acceptance

- Common Vocabulary for entities
- Entity-Relationship diagram
- Lists of attributes for entities (with data type and cardinality)
- Particular metadata serialization



Relatively easy to agree upon



Variations inevitable

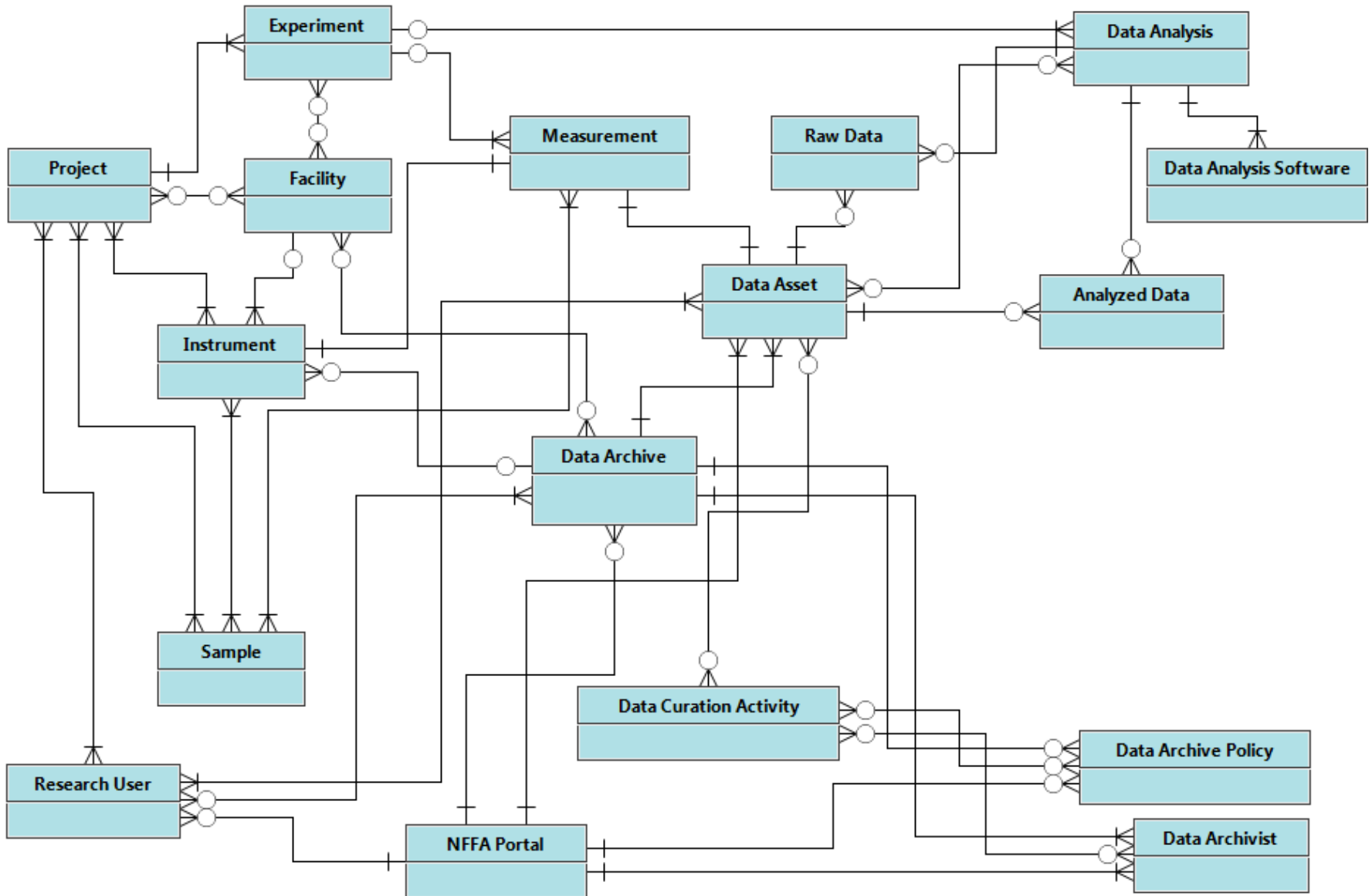


Can be specific to software platform

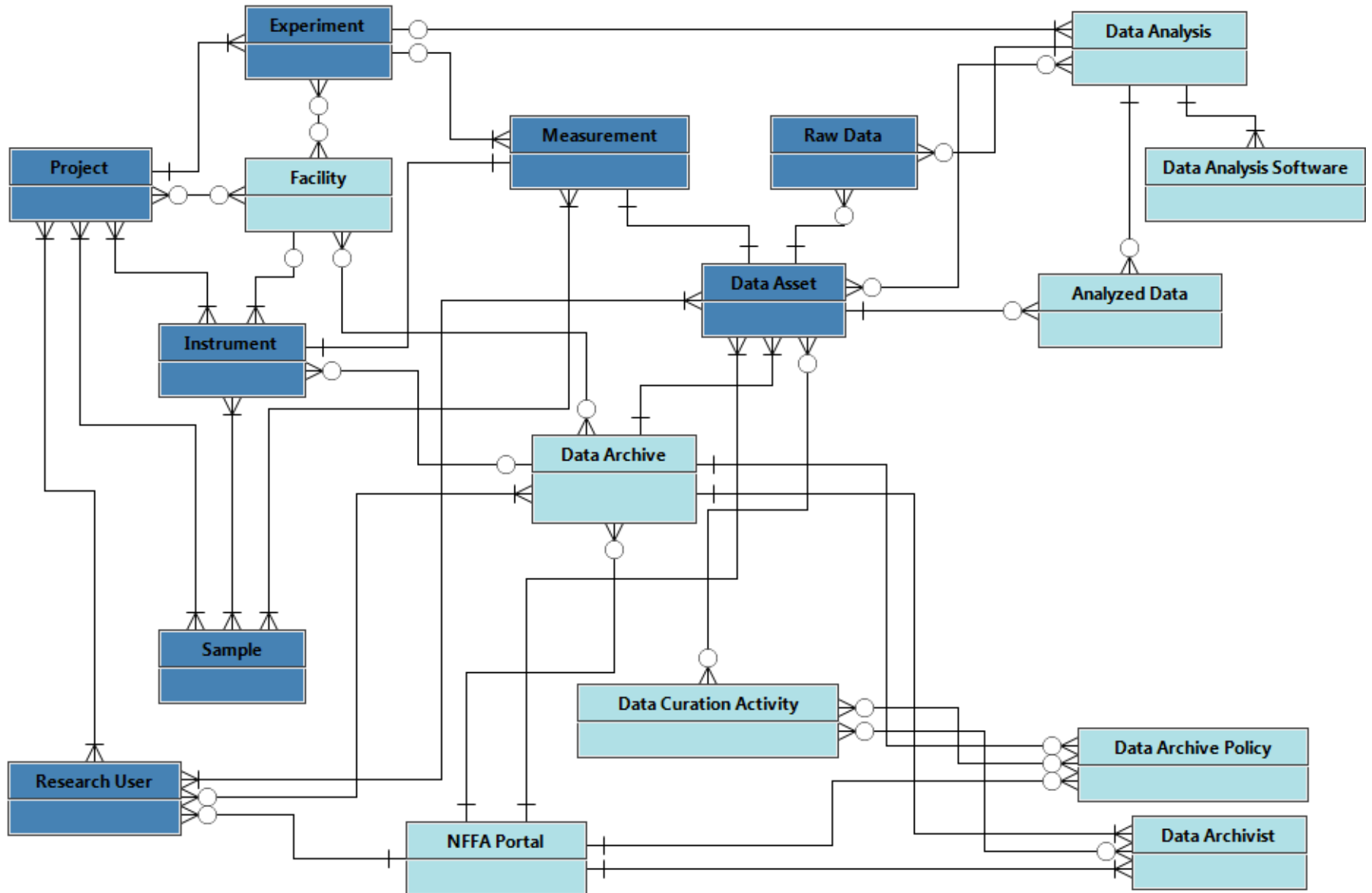
(A fragment of) the Common Vocabulary

- **Research User.** A person, a group of them, or an institution (organization) who conduct Experiment on a nanoscience Facility using a nanoscience Instrument in order to collect and analyze Raw Data, or is interested in data collected or analyzed by other Research Users on the same or other Facilities.
- **Project.** An activity, or a series of activities performed by one or more Research Users on one or more Facilities using one or more Instruments for taking one or more Measurements of one or more Samples during one or more Experiments. Facility, Instrument, Measurement and Sample can refer to computer simulation environment.
- **Facility.** An institution (organization), or a division of it that operates one or more nanoscience Instruments for Research Users. For computer simulation, Facility can be a software platform that allows to order and manage computational experiments (so that the software platform serves the purpose of managing software modules that can be considered virtual Instruments).
- **Instrument.** Identifiable equipment (such as a device or a stand or a line) that allows conducting an independent nanoscience research, perhaps without involvement of other Instruments. Instrument is hosted by Facility and used by Research User. Instrument produces Raw Data in the course of Experiment. Instrument can be in fact a software for computer simulation (a software module or/and a particular configuration of it).

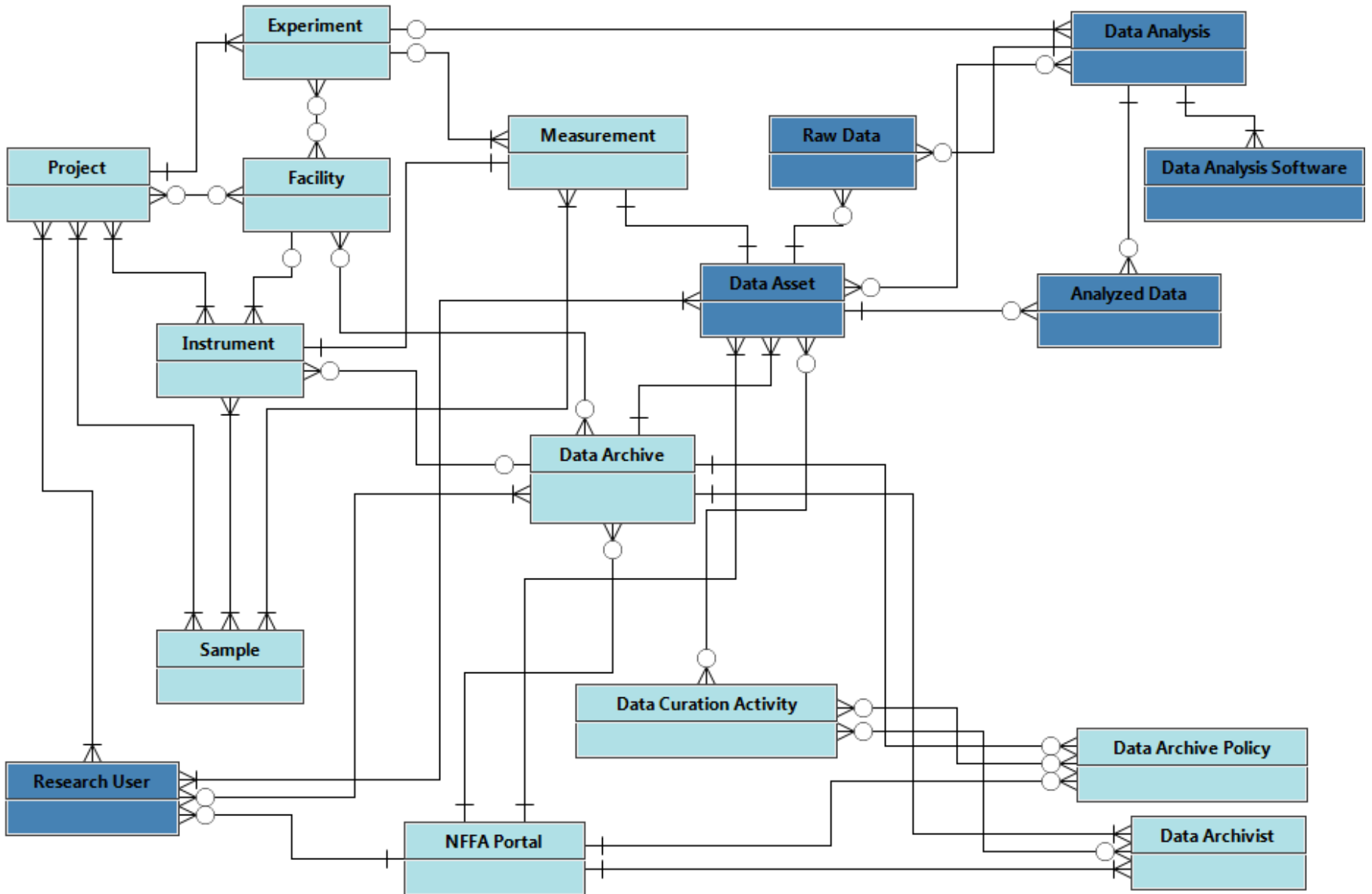
ER diagram



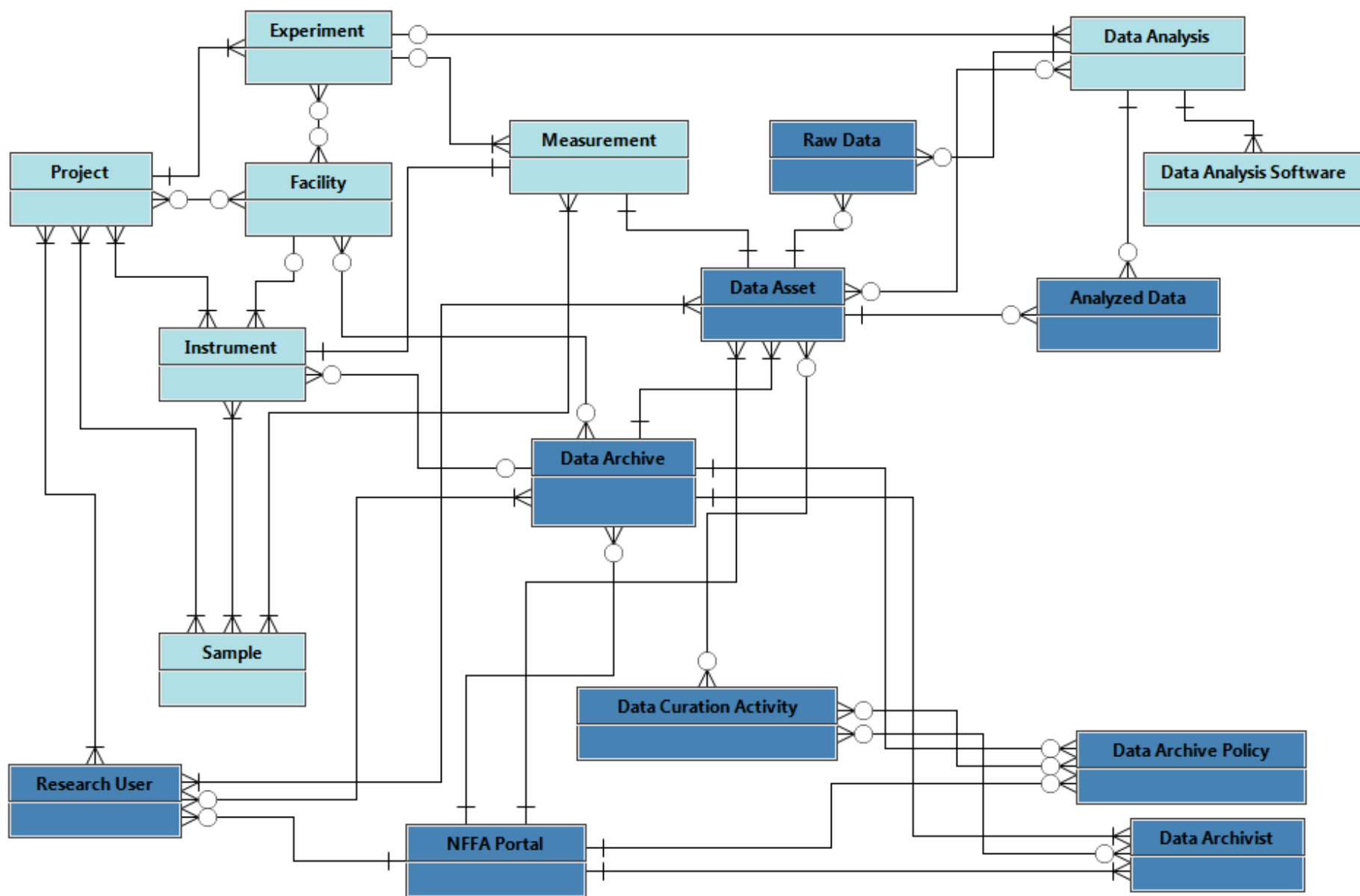
“Experiment” perspective



“Data analysis” perspective



“Data Archive” perspective



Other metadata models for nanoscience and nanotechnology

- CODATA-VAMAS: a focus on modelling the computation for nanoscience with *a posteriori* approach to the extraction of metadata from data files
- Cross-walks between NFFA and CODATA-VAMAS models are possible using three NFFA model entities: Sample (an input file), Experiment (a set of topical keys) and Measurement (a program run).

CODATA: International Council for Science: Committee on Data for Science and Technology

VAMAS: Versailles Project on Advanced Materials and Standards (VAMAS)

Other metadata models for nanoscience and nanotechnology (continued)

- NOMAD: a focus on a nano-object with the metadata categories (sections) for the description of the object shape, size, physical structure, chemical composition, crystallographic structure and surface description.
- Cross-walks between NFFA and NOMAD models are possible using three NFFA model entities: Sample, Experiment and Measurement.

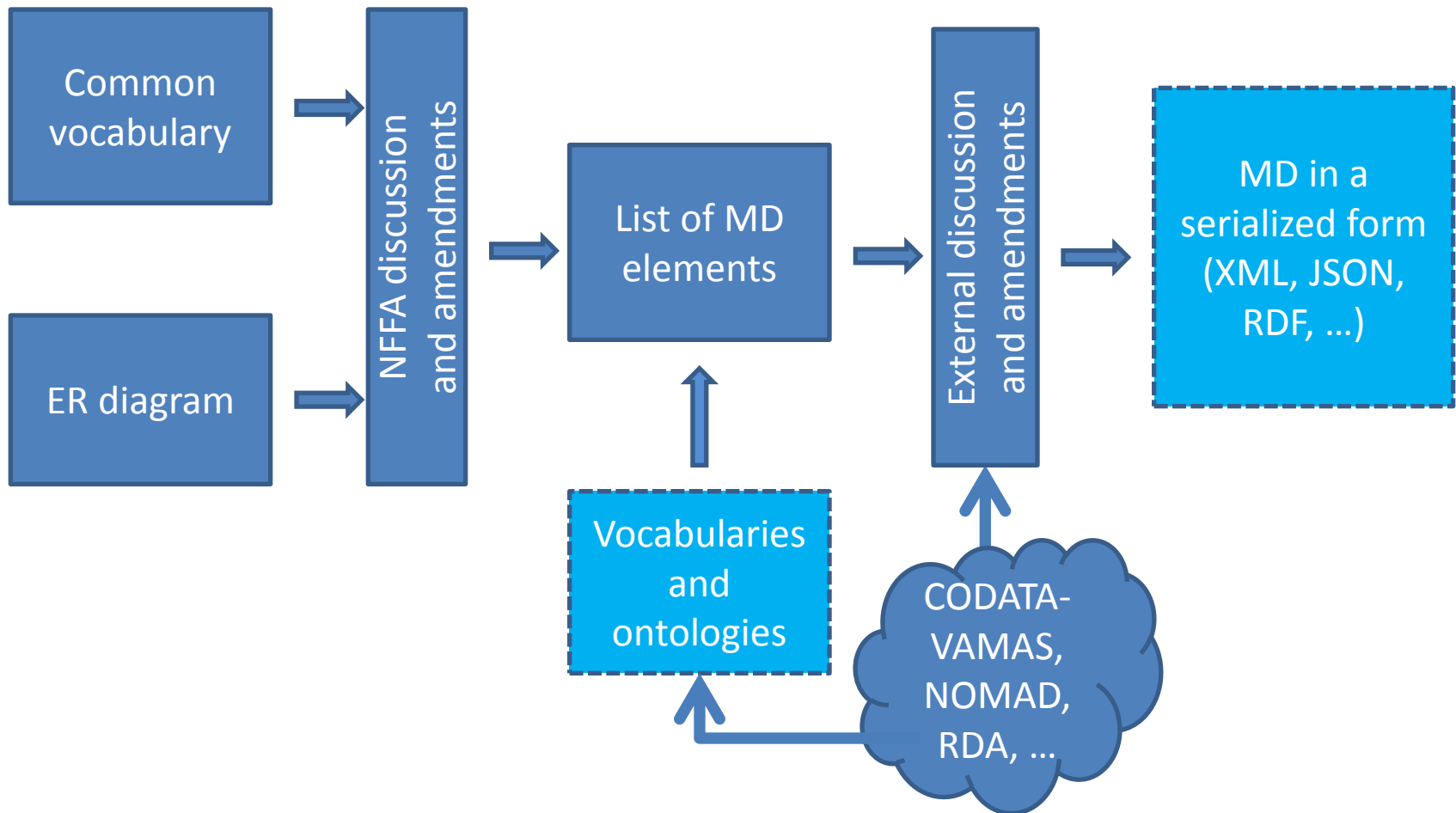
NOMAD: Novel Materials Discovery Centre of Excellence

NOMAD data repository: <https://repository.nomad-coe.eu/>

Cross-walks across NFFA, CODATA-VAMAS and NOMAD metadata models

NFFA concept	CODATA-VAMAS concept	NOMAD concept
Experiment	Nano-object production steps	Series of s/w runs
Measurement	Nano-object testing steps	S/w run
Sample	Nano-object or collection of objects	Input data
Data Asset		Output data

Stages of metadata design ("the process")



Cases for NFFA metadata use and collaboration with other projects and organizations

- Can be applied in any facility research (not necessarily nanoscience) that is interested in having a structured description of data lifecycle
- Can be applied to physical or / and computational experiments, and include data analysis phase Ω
- Can facilitate data exchange and interoperability of software platforms: KIT Data Manager ¹⁾, EUDAT services ²⁾, ICAT ³⁾, ...
- Can be used with controlled vocabularies and ontologies

1) <http://datamanager.kit.edu/>

2) <https://eudat.eu/services>

3) <https://icatproject.org/>

Ω The most advanced (and most promising) case

Further reading & Contacts

NFFA project deliverables (available upon request)

Conference papers:

"Metadata for Experiments in Nanoscience Foundries"

<https://zenodo.org/record/1175958>

"Metadata for nanotechnology: interoperability aspects"

<https://zenodo.org/record/1175964>

NFFA project: <http://nffa.eu>

STCF Scientific Computing: <http://www.stfc.ac.uk/SCD>

Vasily Bunakov vasily.bunakov@stfc.ac.uk

Thank you!

