inst.dlr: a semantic instrument database for scientific large-scale facilities

Witold Arndt[1] // Christian Langenbach[1]

[1] DLR German Aerospace Center – Research Data Management

Motivation for inst.dlr

High-quality metadata is the key to sustainable use and an essential prerequisite for the efficient and reproducible re-use of research data - whether by third parties after publication, by colleagues in the same department or, not least, by the researchers themselves a few years later.

The most efficient solution is to link the creation of data directly with the description and semantically precise linking of machine-readable and schema-compliant metadata.

Key points

- **Duration**: 2023 – 2025 (36 mo)
- **Budget**: ~1 MEUR
- **Personnel**: 3 FTE
- **Scope**: Internal metadata reuse and 3rd party reuse, e.g. technology transfer, marketing, accounting, potentially other research orgs
- **Implementation**: 3 pilot institutes, central IT services
- **Funding**: 100% DLR internal digitalisation programme (PK-D)

DLR locations and mission

DLR is the Federal Republic of Germany's research centre for aeronautics and space.

We conduct research and development activities in the fields of **aeronautics**, **space**, **energy**, **transport**, **security** and **digitalisation**. The German Space Agency at DLR plans and implements the **national space programme** on behalf of the federal government. Two DLR **project management agencies** oversee funding programmes and support knowledge transfer.

DLR operates **150+ large-scale research facilities**, national and international **missions and projects at 50+ sites and offices**. This is operated and established by 11k employees (5.8k scientists) on a yearly budget of ~1350 MEUR.

Main Technical Challenges

Compared to commercial systems:

- Almost always individual items / configurations
- High complexity through modular design
- Iterative changes in set-up and operation
- Long-term usage of components and different life cycles for sub-assemblies
- Components and instruments built by different suppliers and again not off-the-shelf
- Very heterogeneous documentation of components, their maintenance and interactions

Main Organisational Challenges

- Unclear and heterogenous processes for data retention and data exchange
- Concerns regarding data protection and secrecy, NDAs, embargoes
- Limited resources for data annotation, scientists' training

System Structure
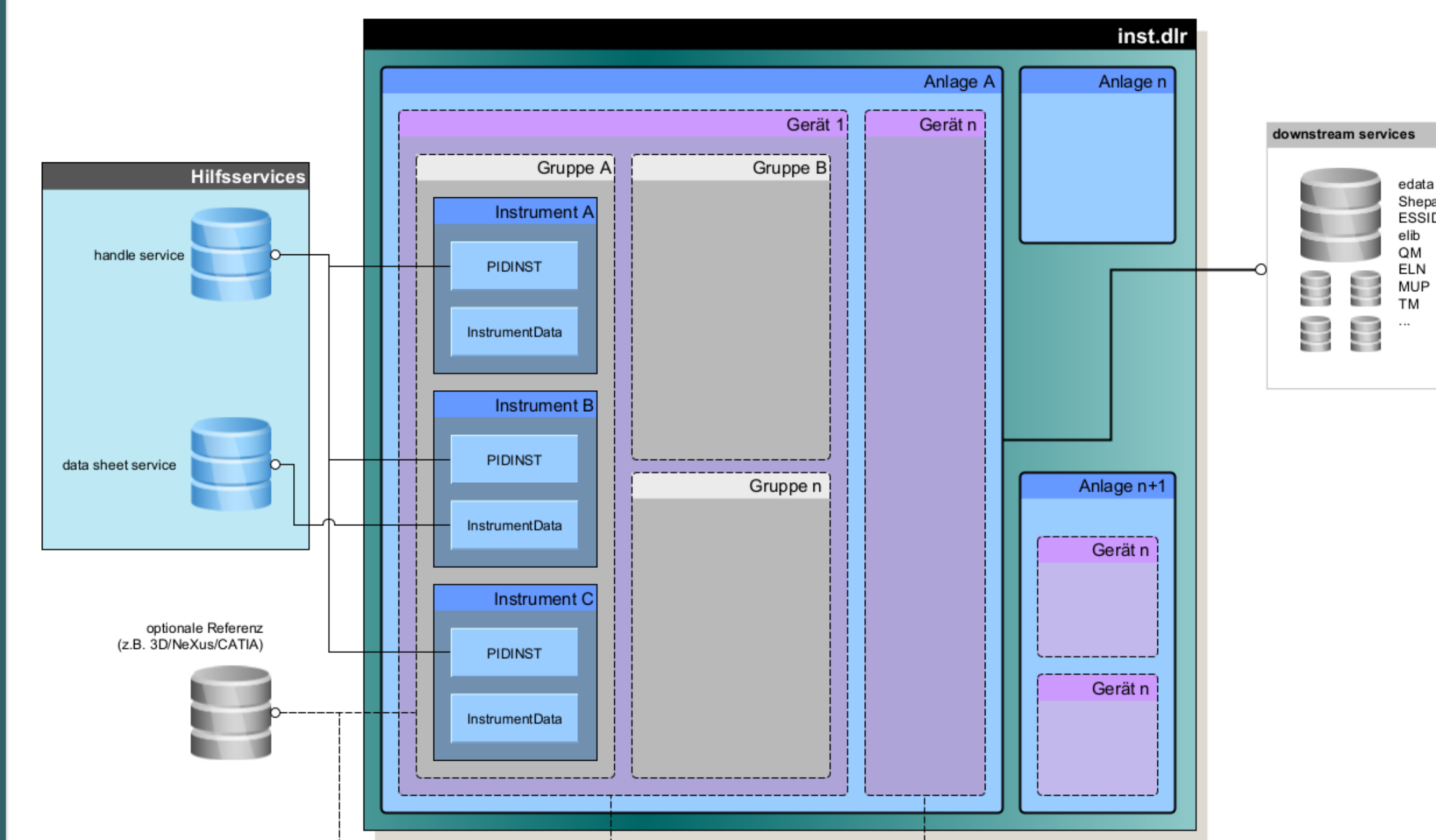
The instrument as the central element in the data model is referenced via PIDINST and provides further metadata as InstrumentData. The two help services for handle and data sheet are connected.

Via an API, metadata of the system, devices, groups or instruments can be passed on to internal or external services with controlled granularity.

Main Organisational Challenges

- Unclear and heterogenous processes for data retention and data exchange
- Concerns regarding data protection and secrecy, NDAs, embargoes
- Limited resources for data annotation, scientists' training

DLR, inst.dlr and HMC

inst.dlr is a contribution to develop and deploy tools, processes and training in metadata generation, data publication and data **re-use for large-scale facilities** with a link between **industry** and **research**.
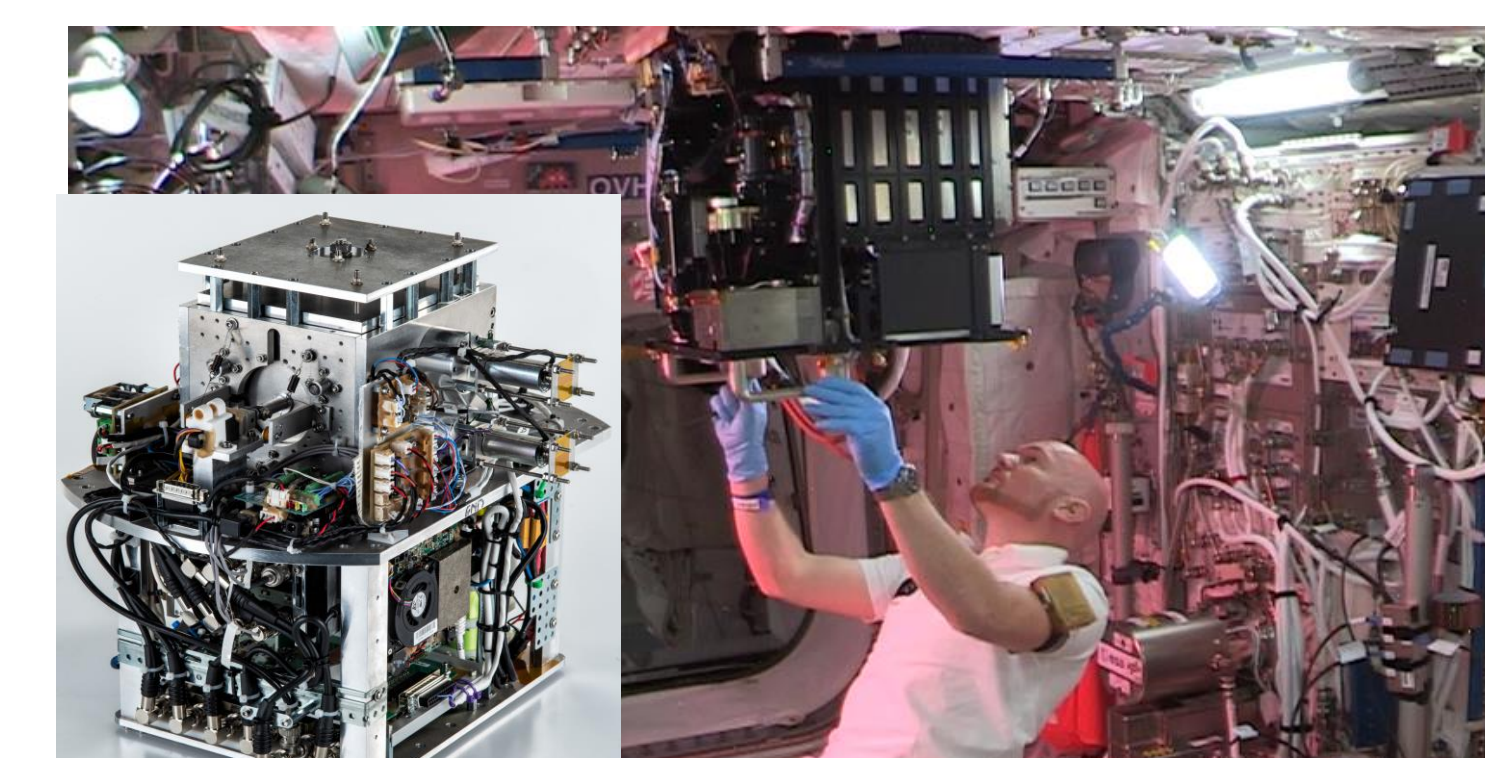
HMC / DLR Contact:
Witold Arndt <witold.arndt@dlr.de>

Foundation

inst.dlr builds upon preliminary work:

- RDO PIDNINST standard
- HZB/HMC RDMinfoPool
- NFDI4Ing OntoHuman

NFDI4ing

PID   RDA RESEARCH DATA ALLIANCE   HZB Helmholtz Zentrum Berlin

Pilot Infrastructures

Increasing complexity, # of interfaces

MEGraMa (Magnetically Excited Granular Matter)

Test Facility for Thermal Energy Storage in Molten Salt (TESIS:com)

DLR Flight Experiments