

Elettra Sincrotrone Trieste

LEAPS-IT

NeXus Data Format 14.05.2019

10 years ago: PaNdata Europe

- Objective inc. "Definition of standards for common scientific data formats"
- The focus was on : **NeXus**
- Most facilities were on-board
- We realised that is was very challenging to deploy a common format with common metadata
 - It remains a challenge 10 years after





Challenges in 2009 (Elettra's perspective)

- Limited knowledge of the concept METADATA
 - Both from IT and Scientists
- Not really supported from ANALYSIS software
 - e.g. no HDF5 in PyMca, Matlab (still in beta), h5py in version 1.
- HDF5 was not mature enough (NeXus is based on it)
- No services like cataloguing (e.g. ICAT) were available
- The dominant formats were RAW, TEXT, TIFF and some custom ones



In the meantime: Today is Elettra

- HDF5 is everywhere
 - Analysis software and development systems support it
- Metadata and their importance is well accepted and understood
 - \circ But still we do not have **common** ones
- No NeXus for now (see next slides on *Challenges* and *Next*)
- All Beamlines with modern TANGO-based Experimental control systems are using HDF5
 - FERMI FEL (LDM, DIPROI, TIMER, TIMEX, MAGNEDYN, TERAFERMI)
 & Few beamlines of Elettra (XRF, TWINMIC, SYRMEP, ESCAMICR)
 - The upgrade Elettra 2.0 will increase its use
 - They have custom structures (on HDF5) developed by Elettra



Challenges

- Even with custom HDF5 there is still need for converters
 - HDF5 to whatever (suboptimal) format the user requires (incl. TEXT)
 - Easier to develop a converter than rewrite the analysis software
- Agreeing on metadata is difficult
 - Imposing metadata of an application (i.e. Computed Tomography) from a different beamline to a local one is not always feasible
- The instruments are custom and often they are in continuous development
 - Matching them with the existing [NeXus Application Definitions] is difficult
- New acquisition/processing models
 - Event-based acquisitions (FERMI), Stream-like processing (Kafka)
- Competition
 - Other than in-house/custom solutions, there are alternatives like the [Scientific Data Exchange] (USA) De Carlo et al. 10.1107/S160057751401604X



Future: towards harmonised solutions

- A harmonised common solution is necessary
 - In the context of the past PaNdata work but also in LEAPS-IT, CalipsoPlus, PaNOSC, ExPaNDS
- For Elettra this is summarised as:
 - CalipsoPlus + PaNOSC + ExPaNDS + Elettra 2.0 = Data Policy + FAIR + EOSC
 - Solutions will be based on the VUO (vuo.elettra.eu) as a virtual laboratory platform and web portal to access the available services and other solutions emerging in the community as result of the above mentioned projects
 - Multi tier storage for datasets (scratch, online, offline, remote)
 - Storage capacity and bandwidth availability is going to increase dramatically
- For NeXus:
 - We will keep examining the technology
 - \circ $\,$ We will increase the use of HDF5 $\,$
 - We will develop HDF5-to-NeXus converters where necessary





www.elettra.eu