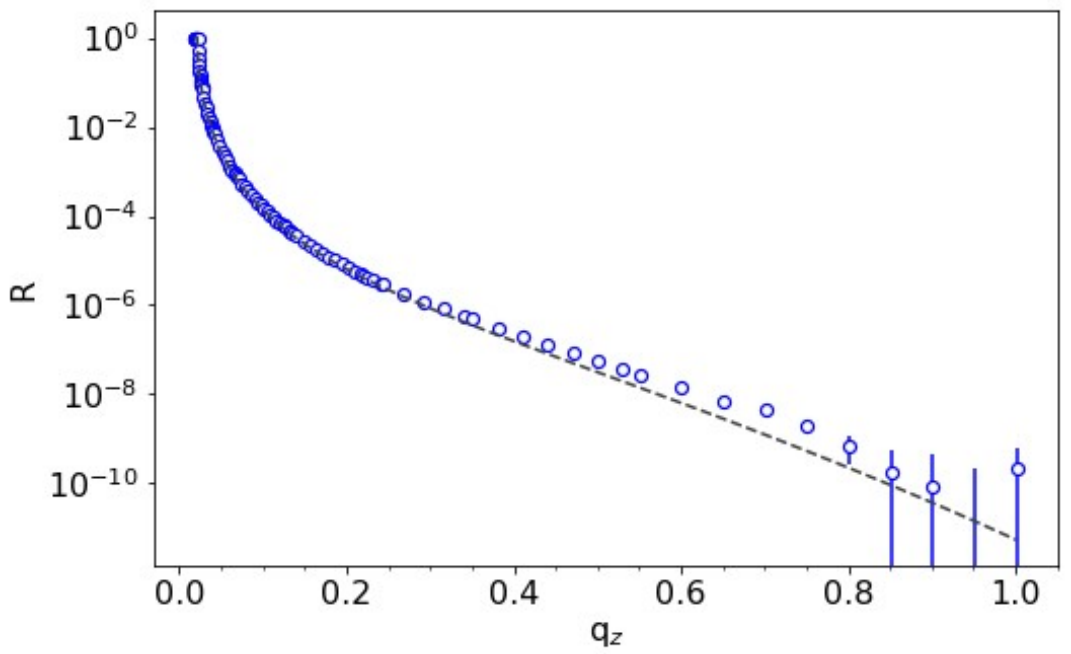
A path towards NeXus

Questions in view of a use case of XRR



Christoph Rosemann, DESY FS-SC
February 13, 2024 NeXus show-and-tell

HELMHOLTZ

DESY.

Outline of presentation

Central idea: clarification on uncertainties, ask questions

The idea and task

- Intermediate data format for specific instrument

Use case evolution and status

- Specific beamline at PETRA III
- Special instrument

Challenges and open questions

- How to create new application definition(s)
- How to address undefined things

My background

~15 years in hep

- Worked in large experiment collaborations
- Software, Computing, Analysis
- Standard frameworks, data formats across languages
- ROOT, CMSSW, LCIO, ...

~8 years in photon science

- Worked on more basic things
- Several designs and implementations on
  - Small processing tools, algorithms, prototype frameworks
  - Data formats for 1D, 2D, 3D and 4D data
  - Loosely based on NeXus/hdf5

The basic idea

A piece of the puzzle that is data processing

Data format in intermediate processing between DAQ (control system) and final analysis

- Abstraction from control system layer, allowing portable use
- Single data format for transient usage and persistency
- Standardisation allows re-use of effort (plus being FAIR)
- Being on the path to having a single file format

Obvious (?) choice NeXus (hdf5)

- Practically there is no choice (due to the lack of acceptable alternatives)
- Still rather low acceptance in scientific communities
- Implementation details are the topic of this presentation

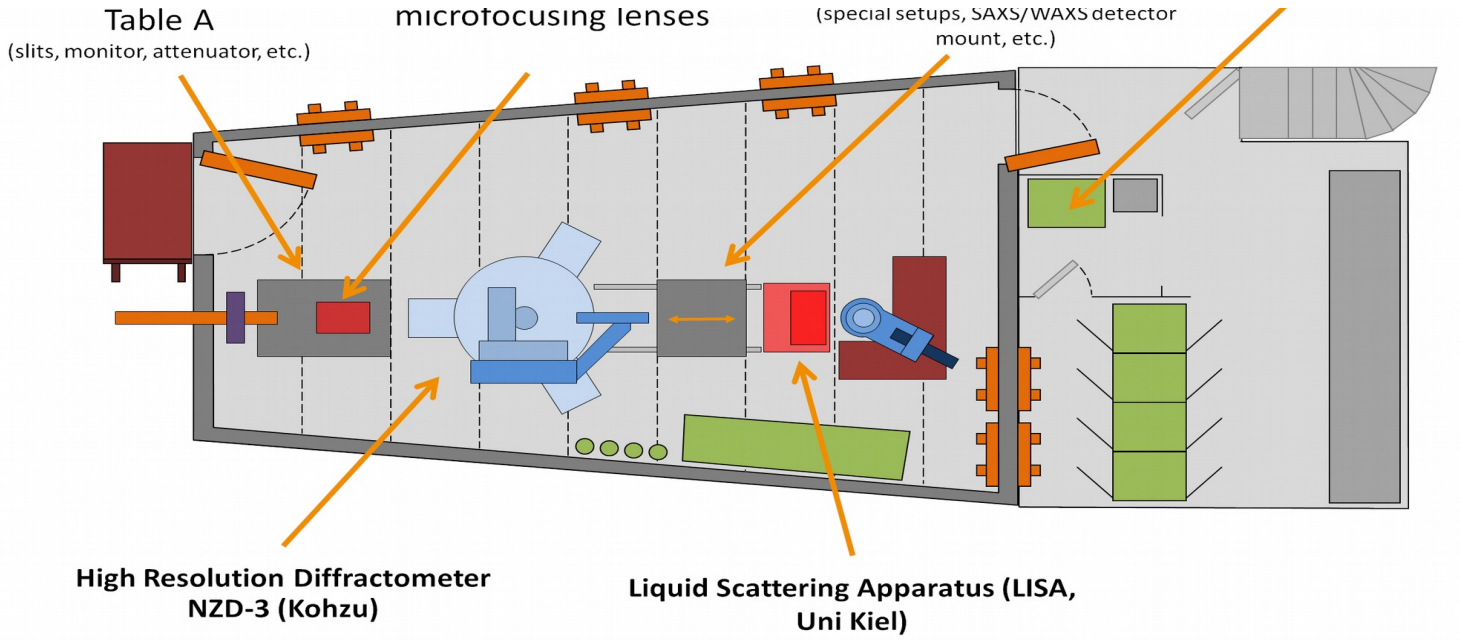
The use case:
LISA at P08 at PETRA III
X-ray reflectometry of liquids

LISA at P08

Liquid Interface Scattering Apparatus

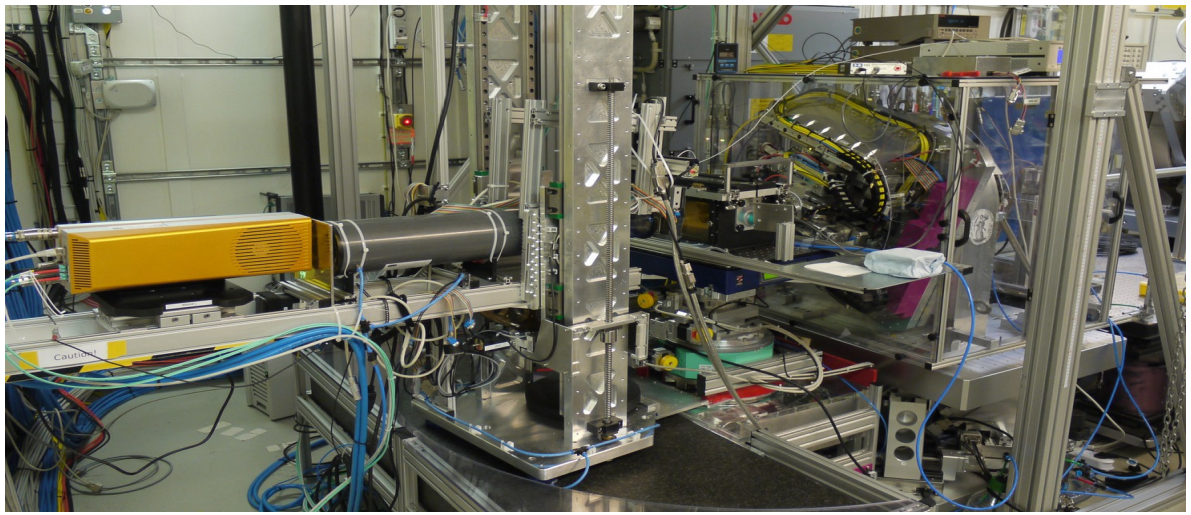
P08 – High resolution diffraction

- Features a 6-circle diffractometer
- Second measurement station – LISA
- XRF is optional
- In addition powerful laser available



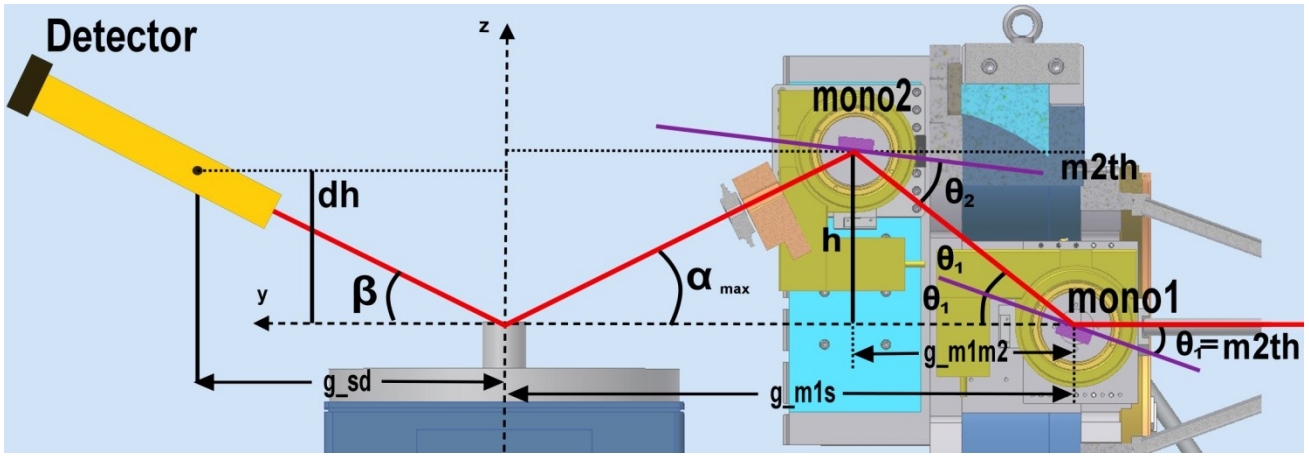
References

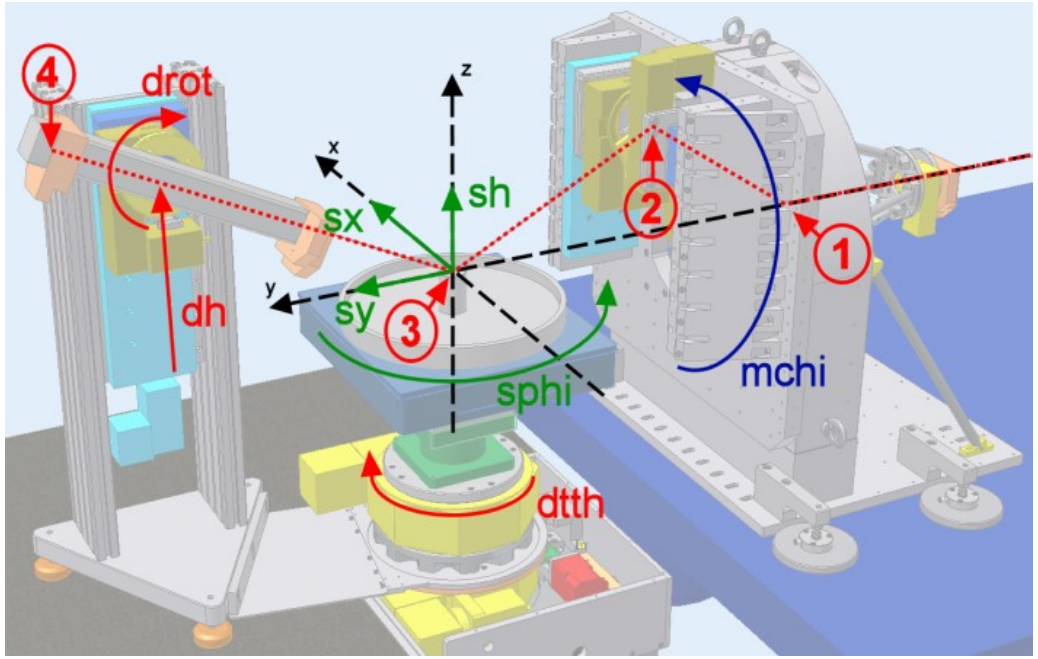
- [1] O.H. Seeck et al., Journal of Synchrotron Radiation 19, 30-38, (2012). [10.1107/S0909049511047236]
- [2] B.M. Murphy, et al, Journal of synchrotron radiation 21 (1), 45-56 (2014) [10.1107/S1600577513026192]

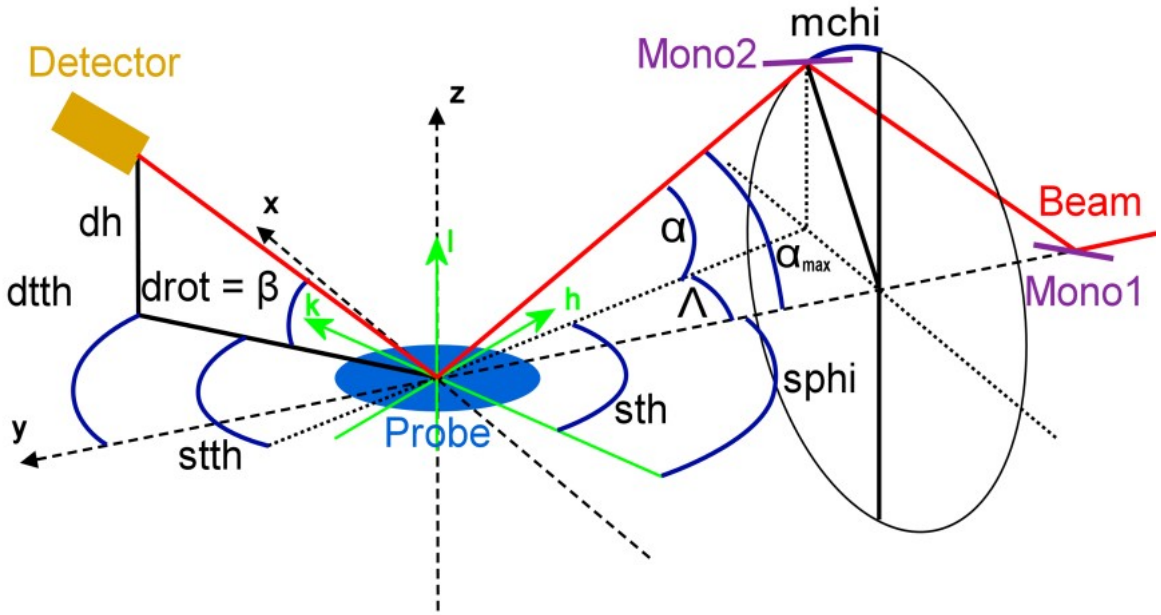


Geometry setup and motor names in control system

Just some more detail







Data acquisition

TANGO/SARDANA driven

Setup

- Separate data streams for imaging detector and everything else (motors, counters,...)
- File format can be chosen: hdf5/NeXus; as well as ASCII based
- All naming conventions stem from the control system
- Ongoing development to type and check values (Jan Kotanski, FS-EC)

Challenges

- Brittle setup, probably inherent to the task
  - Any change of a name, cable or address will alter the output without warning
- No import or export of this data, requires local expert knowledge
- Neither user nor analysis friendly

Division into single tasks

The data flow

Initial data from control system

- Discover data
- Parse given format
- Use pre-defined list of names
- Validate data in terms of type and range

Transient formulation

- Find coherent representation of data
- XRR data has an inherent hierarchy
- Guiding thought: how is the data used, how is it accessed
- Define abstract structure
- Do implementation including IO

Persistency

- NeXus has two layers of classes
- “translate” the formal definition into NeXus vocabulary
- This is not straightforward

Progression

Input for the task from control system

Initial data from control system

- FIO (“header and column”)
- Hdf5 files from 2D detector

```
!
! Comments
!
%c
qscan 0.0 0.0 0.0 0.0 0.018 0.027 16 1.0
user p08user Acquisition started at Thu Apr 21 07:24:52 2022
!
! Parameter
!
%p
rois lambda  = [27, 70, 50, 9
abs = 5.0001
atten = 1.97893
bcl = 0
```

```
%d
 Col 1 mchi DOUBLE
 Col 2 beta DOUBLE
 Col 3 dtth DOUBLE
 Col 4 sphi DOUBLE
 Col 5 exp_t01 DOUBLE
 Col 6 lambda_max DOUBLE
 Col 7 lambda_roi1 DOUBLE
 Col 8 lambda_roi2 DOUBLE
 Col 9 lambda_roi3 DOUBLE
 Col 10 lambda_roi4 DOUBLE
 Col 11 lambda_roi5 DOUBLE
 Col 12 lambda_roi6 DOUBLE
 Col 13 lambda_index DOUBLE
 Col 14 lom_foil DOUBLE
 Col 15 lom_foil_vfc DOUBLE
 Col 16 petra_beamcurrent DOUBLE
 Col 17 ion_bl DOUBLE
 Col 18 apd DOUBLE
 Col 19 apd2 DOUBLE
 Col 20 ion1 DOUBLE
 Col 21 ion2 DOUBLE
 Col 22 alpha_pos DOUBLE
 Col 23 beta_pos DOUBLE
 Col 24 sth_pos DOUBLE
 Col 25 stth_pos DOUBLE
 Col 26 qx DOUBLE
 Col 27 qy DOUBLE
 Col 28 qz DOUBLE
 Col 29 qpar DOUBLE
 Col 30 q DOUBLE
 Col 31 epoch DOUBLE
 Col 32 timestamp DOUBLE
 89.60180232558429 0.055964759644922424 -7.725231048508495 7.720100848112113
00002 6035.0 0.056451415463587645 0.05641476002627854 0.005046550597289645 -9
0518700.5069897 6.708765506744385
```

Name	Description	Type	Shape
test_00367_00000.nxs		NXroot	
entry		NXentry	
instrument		NXinstrument	
detector	"haspp0...	NXdetector	
acquisition_mode	"Twenty...	string	scalar
bit_depth_readout	[24]	uint32	1
calibration_date	""	string	scalar
collection		NXcollection	
count_time	[1000]	float64	1
countrate_correction_applied	[1]	int8	1
data	Compres...	int32	17 × 516 × 1
description	"Lambda"	string	scalar
detector_readout_time	[0]	float64	1
flatfield	Compres...	float32	0 × 516 × 15
flatfield_applied	[0]	int8	1
geometry		NXgeometry	

The data class description

Hierarchy

- A dataset is a set of scans, typically the data over a whole range of qz values
- Alternative name “measurement”
- Constant for a dataset is the detector and the sample
- A “scan” is a concept from the control system

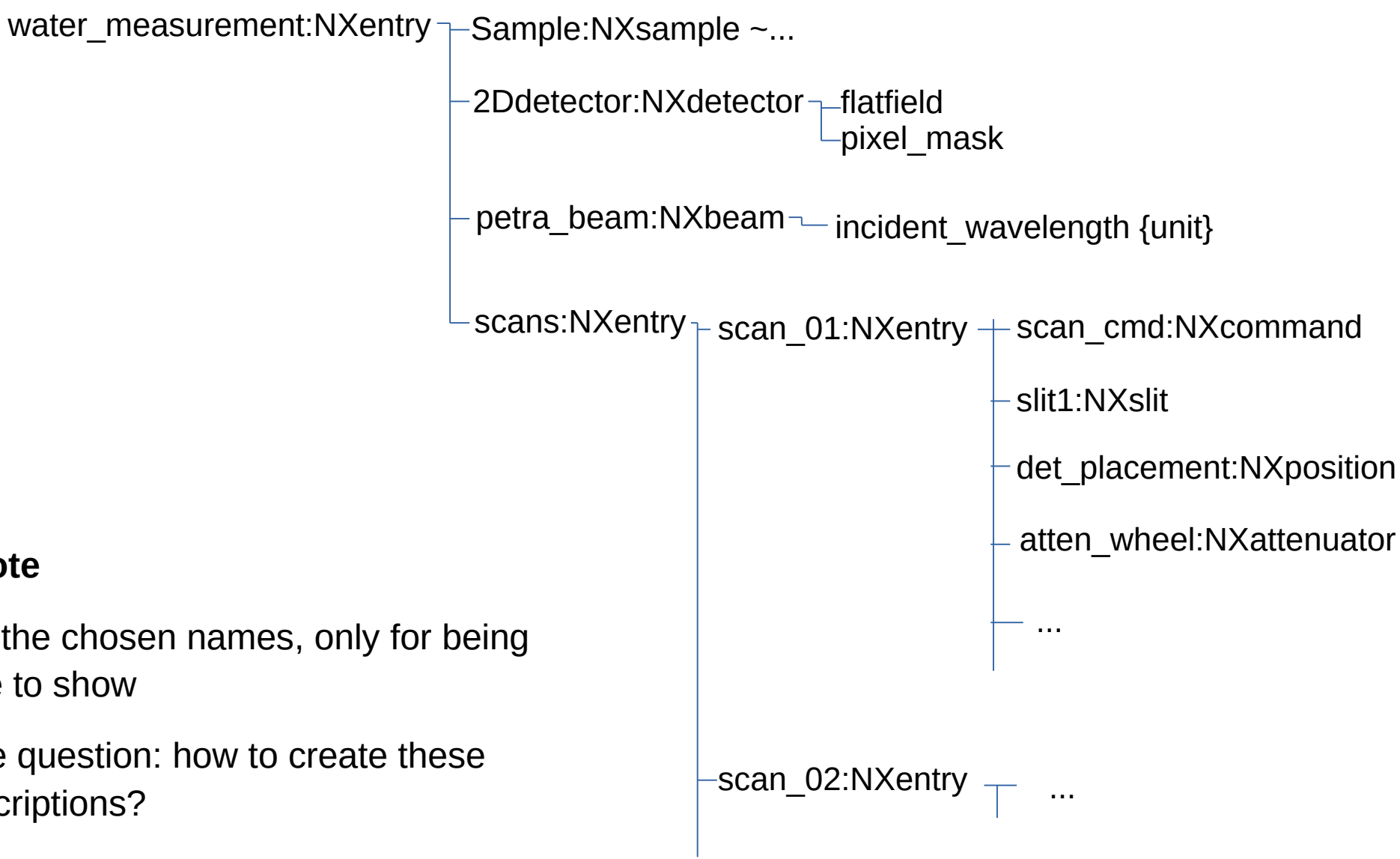
A scan

- consists of scan points
- is defined by the parameters of:
  - qz start and end
  - Exposure time
  - Absorber position

A scanpoint

- is defined by the values of:
  - detector counts
  - angles (motor positions)
  - apd, ion chamber, beam current

NeXus definition



Side note

- Not the chosen names, only for being able to show
- Side question: how to create these descriptions?

NXDL definition

```
    # LISA - a draft description of data taken with the LISA instrument at P08 at PETRA III

-->

<definition name="LISAdata" extends="NXobject" type="group"
    category="application"
    xmlns="http://definition.nexusformat.org/nxdl/3.1"
    xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
    xsi:schemaLocation="http://definition.nexusformat.org/nxdl/3.1 ../nxdl.xsd"
    >

    <doc>This is a draft definition for data taken with the LISA instrument at beamline P08 at PETRA III at DESY.</doc>


        <group type="NXentry" name="entry">
                <field name="title"  minOccurs="0" maxOccurs="1"/>

                <group type="NXsample" name="sample">
                                <field name="name">
                                <doc>Descriptive name of sample</doc>
                                </field>
                </group>

        <group type="NXdetector" name="detector">
                        <field name="flat_field" type="NX_INT" signal="1">
                                <dimensions rank="2">
                                <dim index="1" value="xSize" />
                                <dim index="2" value="ySize" />
                                </dimensions>
                        </field>
                        <field name="pixel_mask" type="NX_INT" signal="1">
```

Main questions

Big (stupid?) question

- Is there a transient class instantiation? (e.g. class constructors)

Mapping of classes, mapping of hierarchy

- Is there a better way to represent the hierarchy of data in NeXus?

Developing a NXDL definition

- Right now entirely manual process (using a XML editor)
- Is there a possibility to automate this procedure?

How to extend NeXus with functionality

- eg. develop a NeXus compliant framework
- Would you recommend libraries (vs. applications)?

How to represent geometry in NeXus

- Is there a tutorial/more extensive resources to geometry description?

Thank you

Contact

Deutsches Elektronen-
Synchrotron DESY

www.desy.de

Christoph Rosemann
FS-SC
christoph.rosemann@desy.de