Project information

Project full title	LEAPS pilot to foster open innovation for accelerator-based light sources in Europe
Project acronym	LEAPS-INNOV
Grant agreement no.	101004728
Instrument	Research and Innovation Action (RIA)
Duration	01/04/2021 – 31/03/2025
Website	https://www.leaps-innov.eu/

Deliverable information

Deliverable no.	3.2
Deliverable title	Universal data format specification for surface topography description
Deliverable responsible	ESRF
Related Work-Package/Task	WP3/ Task 3.4 Development of new metrology methods and measurement protocols suitable for production environments
Type (e.g. Report; other)	Report
Author(s)	F. Polack, R. Barrett
Dissemination level	Public
Document Version	1.0
Date	29/09/23
Download page	https://www.leaps-superflat.eu

Document information

Version no.	**Date**	**Author(s)**	**Comment**
1.0	29/09/23	F. Polack, R. Barrett	

This project has received funding from the European Union´s Horizon 2020 research and innovation programme under grant agreement No. 101004728

Abstract

Superflat is Workpackage (WP) 3 of the LEAPS-INNOV project. WP3 draws together ten LEAPS facilities[1] with common needs for the development of high-performance reflective X-ray optics. Cross-validation of the surface topography metrology data of these optics is a critical step in improving their quality. Here a specification for an Interchange Metrology data format has been developed in order to facilitate exchange of data between the principal stakeholders which span end-user facilities,metrology laboratories and optical manufacturers. The specification defines the minimum data and metadata content required for unambiguous exploitation. It proposes a data structure which is compatible with the NeXus format which is already widely adopted by LEAPS facilities and their users.

[1] Further information regarding the Superflat consortium members is available at leaps-superflat.eu

This project has received funding from the European Union´s Horizon 2020 research and innovation programme under grant agreement No. 101004728

Contents

Glossary

Acronym	Definition
LEAPS	League of Accelerator based photon sources
PCP	Pre-commercial Procurement

This project has received funding from the European Union´s Horizon 2020 research and innovation programme under grant agreement No. 101004728

1 Introduction

One of the principal objectives of the WP3 Superflat is the development of new metrology methods suitable for implementation in industrial environments. One of the key activities to ensuring the validity of these methods is the exchange of optics between stake holders in cross-validation exercises and subsequent comparison of surface metrology data acquired with different instruments and by different laboratories. With improved surface topographies it is increasingly important to ensure that the different datasets are fully documented with regard to the data acquisition and processing steps. To date critical information (which may often reside in so-called 'metadata') is usually not stored directly or in any universal format with the surface metrology data. This may introduce errors in the comparison and exploitation of data.

To overcome these shortfalls, we are attempting to facilitate the exchange of data between surface metrology laboratories and optics manufacturers via the development of a universal data format for describing the measured surface topography of X-ray reflective optics. The core of this report presents the principal requirements and features of the proposed format. The full format description is given in appendix.

2 Principal considerations for the universal data format

- The data format is not intended to replace the native data format of the measuring instrument is only required to define the **minimum of data and metadata** needed for reliable interpretation without external information and the organisation of these data and metadata in the file. For the purpose of data exchange, only the final consolidated data are relevant. Nevertheless, the data format should be readily extendable such that, if required, it can incorporate intermediate measurement data or even the raw data.
- To encourage adoption of the format it is important that open-source software tools be available to facilitate reading and writing of the files. Whilst this is outside the scope of the deliverable, preliminary tools have been developed and more sophisticated implementations in data analysis software are currently being considered. These constitute some of the ongoing activities of the WP task 3.4.
- The data format should facilitate the application of FAIR data management

3 Key features of the universal data format

- The file format distinguishes four broad categories of information which may be stored in the file:
    1. Surface measurement results
    2. Sample description
    3. Instrument description
    4. Environment data
- In view of the required information to be stored, it has been chosen to adopt the hdf5 container format whilst structuring the data according to the NeXus conventions. This structure is already widely used in LEAPS and other RIs (and their user communities). Its

This project has received funding from the European Union´s Horizon 2020 research and innovation programme under grant agreement No. 101004728

adoption has the advantage of allowing the use of many existing open-source software tools to navigate, visualise and process the data sets. This should allow the development efforts within the X-ray reflective surface metrology community to be concentrated upon the specific needs of the application.

4 Conclusions and perspectives

- A file specification for surface topography metrology data for X-ray reflective optics has been developed.
- The format will be tested initially within the Superflat consortium in order to verify the suitability of the specifications in real-use situations. If necessary the specification can be extended or adapted after these field trials. The use of the NeXus and hdf5 conventions should facilitate any such modifications to the specification.
- Once the specification has proven to be well adapted to the needs as determined within the project consortium, the file description will be more widely disseminated within the optics manufacturing and the international metrology community. This will exploit various channels including personal contacts with companies involved in the Superflat PCP (Task 3.2), international workshops and the leaps-superflat.eu website.

5 Appendices

5.1 Specification of a format for exchanging optical surface metrology data

This project has received funding from the European Union´s Horizon 2020 research and innovation programme under grant agreement No. 101004728

Project information

Project name	LEAPS-Superflat
Project scope	LEAPS-Superflat is workpackage 3 of the LEAPS-INNOV project. This workpackage aims to improve European capabilities for the production of very high-quality X-ray mirror and grating substrates which are key to achieving the ultimate performance of our user facilities.
Grant agreement no.	101004728
Instrument	Research and Innovation Action (RIA)
Duration	01/04/2021 – 31/03/2025
Website	leaps-superflat.eu

Specification of a format for exchanging optical surface metrology data

This document describes the functional specifications which should be fulfilled by an interchange Metrology data format usable by the members of the LEAPS-INNOV Superflat consortium and their Optics providers, in order to facilitate interoperation and cross-validation of the data. The intent is to define the minimum data and metadata content required for unambiguous exploitation and propose a data structure compatible in a widely available file format, while preserving a possibility of extension to match individual requirements.,

Superflat is part of the LEAPS-Innovation project, which has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 101004728

Contents

Superflat is part of the LEAPS-Innovation project, which has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 101004728

1 Purpose and scope

As the required accuracy of the surface topography of manufactured optical surfaces increases, it approaches the accuracy limits of the metrology which are primarily limited by the calibration of the measuring instruments, compensation of systematic errors and correction of other measurement bias. Manufacturers, and most of the institutions gathered in the LEAPS-Superflat consortium have their own, often internally developed, metrology instruments and procedures. The ever more stringent accuracy requirements tend to reveal significant differences in the surface metrology determined by different instruments. It is therefore increasingly important to compare the measured data with those produced in another facility.

In addition to the variety of existing instruments and measurement procedures, there is also a variety of output data formats, where all the information required for accurate cross-comparison is not always included and needs to be provided externally. The need for a common exchange format of surface metrology data is therefore widely recognized. This common format is not intended to replace the native, instrument specific recording format, but should only define the **minimum of data and metadata** needed for reliable interpretation without external information and the organisation of these data and metadata in the file. For the purpose of data exchange, only the final consolidated data are relevant. It is however important that the chosen format remains **extendable** within the same organisation framework, so that it can incorporate intermediate measurement data or be adapted to further evolution of optical surface metrology and the needs of specific instruments. Finally, the format should allow long term archiving and therefore be **compatible with broadly available reader and writer applications. To promote straightforward data exchange these tools should preferably be open source.**

2 Review of the mandatory data

Depending on the sample and instrument, the collected data which should be stored in the file may differ. These can be broadly classified into four categories:

1. Surface measurement results
2. Sample description
3. Instrument description
4. Environment data

The Sample and Instrument sections, include data which are constant during one measurement set, but which can vary from one measurement set to another if several measurement sets are present in the file, e.g. measurement of an actively modified surface with different values of the actuators. These variations of configuration should be tracked in the Sample or Instrument data and properly referred to in each measurement set. Finally, the Environment data will group all parameters which can have an influence on the results but cannot be attributed to the sample or the instrument.

LEAPS INNOVATION
Co-funded by the European Union
Superflat is part of the LEAPS-Innovation project, which has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 101004728

In this section the review of the data which must be recorded will be made independently of their format and physical distribution in the file, mainly focusing on the parameter definitions and the interconnections between the different data groups.

2.1 Optical surface measurement data

Depending on the measuring instrument, the shape of the optical surface can be defined by the **height** and/or the **slope** gradient versus the position in the 'surface plane'. This requires the definition of a spatial reference frame in which position and slope are defined. This **Surface reference frame** may or may not be identical to the **Sample reference frame**, in which case the space transformation between the two frames should be specified.[1]

The Surface reference frame should be oriented in such way that the height is represented by the Z coordinate, with the Z axis pointing outside of the physical surface (drawing). The XY plane is considered as the '*surface plane*'. The slope gradient, if needed, is given by its two components Sx=dZ/dX and Sy=dZ/dY. This decomposition depends on the chosen frame orientation. The definition is unambiguous provided that the Surface to Sample transformation is given. A concave surface in the XZ plane will have a positive curvature d^2Z/dX^2 = dSx/dX, and a convex surface a negative curvature.

It is recommended, where possible, to align the X axis of the surface reference plane parallel to the projected trace of the X-ray beam propagation on the XY surface plane with the positive direction in the same sense as the propagation.

A measurement point is in principle expressed by a tuplet {Z, dZ/dX, dZ/dY, X, Y}, but some elements may be useless and not given, according to the measurement method (e.g. slopes with interferometry).

Surface metrology data are very often processed to remove a known surface shape, such as a sphere, a toroid, or an elliptical cylinder, from the raw measurement values. While processed data are useful for visualising the final metrology result, keeping track of the raw data and of the applied processing is highly desirable. If data result from some kind of processing, metadata describing it should be added and attached to the dataset; the source dataset from which they are derived should be present in the file and referred to by the processing metadata.

2.1.1 Discussion

In previous discussions there were claims to define the X and Y axis in directions related to the destination use of the object. From a strict logical point of view, cross-comparison should not be

[1] In discussions regarding the data format some users suggested that the Surface reference frame should have its origin at the physical centre of the surface. With the definition of a surface reference frame this is possible, but since a surface centre can be difficult to define accurately, the space transformation between the two reference frames should be provided anyway.

LEAPS INNOVATION

Co-funded by the European Union

Superflat is part of the LEAPS-Innovation project, which has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 101004728

related to a particular direction of use. Measurement orientation information must be included in the data format, but since perfect orientation is difficult to achieve it should not be defined by the format. However, it is convenient that the object owner specifies his preferred orientation of the axes of coordinate. This orientation should be part of the Metrology request specification, in particular when profiler instruments (NOM, LTP) are used. In such case the line profile direction should be specified by reference to some Sample reference marks.

2.2 Sample data

The sample data should give all information required for traceability and for precisely locating the Optical surface measurements with respect to the physical object. It should include:

- A physical description of the object, allowing the reproduction of the measurement configuration. This includes the shape and overall dimensions of the sample.
- The clear aperture of the Optical surface on which the data analysis is meaningful and should be performed. If several clear subapertures are defined on the surface they would normally be described by separate measurement data records.
- Reference mark positions. In all cases, measurement should be located with respect to reference marks, which can be either engraved or printed on the object, or any physical features of the object, edge, corner, which can be unambiguously located.
- Optionally, but highly desirable, a drawing, describing the geometry and the location of these elements.
- Configuration parameter which may influence the measurement. Position in space, for gravity influence; Actuators, in case of active optics; etc …
- Any information which can be useful for object traceability: Name, owner, origin, identification numbers.

The descriptive parameters will be referred to a Sample reference frame, in which all numerical coordinates will be expressed. As the measurement data are related to the Optical surface, the Sample reference frame will be preferably related to the '*Surface plane'*, as above defined, but this is not mandatory.

2.3 Instrument data

This section should provide all parameters needed for exploiting the measurements which are fixed and common to all measurement sets present in the file. It should also give all information enabling the measurement to be reproduced. If the file solely contains consolidated final data, this second part is only informative, and the level of detail can be left to the data producer as appropriate.

The minimum requirement is the name of the instrument and a version number. If several configurations are possible the used one should be identified.

LEAPS INNOVATION
Co-funded by the European Union
Superflat is part of the LEAPS-Innovation project, which has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 101004728

Instruments usually display two almost independent parts: a sample stage to position the sample for the measurement and a detector which can be static or more usually positioned with a detector stage. Case of LTPs or NOMs where the detector scheme has both a static and a moving part, is a little trickier but could be forced into the model.

Description of the instrument movements should be referred to the laboratory frame. The only requirement is that the Z direction is aligned on the gravity and pointing upwards. The XY plane can be chosen for convenience, e.g. aligned on the main translation axis.

2.3.1 Detector

The detector section should primarily describe the parameters of the detectors needed to interpret the results. It will depend on its dimensionality. It must include:

- A geometric physical description of detector pixel shape, pixel size (x, y), pixel pitch (x,y). The description might not be obvious for some scanning beam instruments. In this case the description should pertain to the measurement probe characteristics. For instance a LTP could be described as shape="gaussian" , size= "gaussian σ".
- The mapping of this geometry to the sample reference frame. Origin and direction of storage scans; magnifications (x,y).

2.3.2 Sample/detector stage

The sample and /or detector stages sections are only required if the file contains a sequence of measurements of the same object. description should include:

- A kinematic description of all actuators: direction of action, place in the kinematic chain.
- The records of all positions used in the measurements. If measurements are done on the fly, it should give the parameters allowing to reconstruct the acquisition position, speed, starting position and time, sampling interval etc …

When one of these sections exist, the mapping of the detector or sample reference frame to the instrument frame should be also available in the file.

2.4 Environment data

This section should collect all data which are not directly related to the measurement method but may have an influence on the result. It will be primarily the temperature at various positions in the measurement enclosure, sample detector, etc … Accidental events occurring during the measurement should also be recorded here.

3 File organisation

Despite the variety of instruments, the data and metadata should be organised in the file according to the principles of FAIR data management, that is they should be Findable, Accessible, Interoperable

LEAPS INNOVATION

Co-funded by the European Union

Superflat is part of the LEAPS-Innovation project, which has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 101004728

and Reusable. This approach is also implemented by other scientific programs inside the LEAPS scientific community (e.g. NEP data management published by NFFA project: https://www.nffa.eu/about/data-management/data-management-plan/)

- **F**indable, requires the data and metadata are well defined with a common vocabulary
- **A**ccessible, that all data and metadata are consolidated in a single and file. They should be also self-documented.
- **I**nteroperable, means a file format which can be written and read by open source, multi-platform software., widely spread and available inside our scientific community. Standard data schema should be associated to the format.
- **R**eusable, the format should be flexible to accommodate data resulting from various instruments of different types. It should be also easily extensible to adapt to new experimental schemes and methods.

3.1 The NeXus format

To satisfy those requirements, we propose to adopt the NeXus data format (http://www.nexusformat.org/), as has been done already by a number of user communities inside LEAPS facilities. This was also the choice of the NFFA project (cited above), with which we share the similar objectives of creation of an advanced system for data and metadata management and the implementation of Metadata Schemes for the Metrology Techniques for which a commonly accepted Metadata Standard is not yet available.

NeXus is an international standard for the storage and exchange of neutron, x-ray, and muon Experiment data, but can in principle be extended to other techniques. The structure of NeXus files is extremely flexible, allowing the storage of both simple data sets, such as a single data array and its axes, and highly complex data and their associated Metadata, such as Measurements on a multi-component Instrument or numerical simulations. NeXus is built on top of the container format HDF5 (Hierarchical Data Format 5: https://www.hdfgroup.org/solutions/hdf5/) and adds domain-specific rules for organising data within HDF5 files in addition to a dictionary of well-defined domain-specific field names. The documentation of the NeXus format can be found at: https://manual.nexusformat.org/ref_doc.html.

NeXus provides an already well-developed framework for the creation of the nomenclature, Vocabulary and Metadata we require, and it can suitably enriched with new entries consistent with it, whenever necessary

3.2 NeXus features

3.2.1 Hierarchical structure and random data access.

Build on top of HDF5, NeXus provides a hierarchical structure of nested groups and datasets (or fields). Groups and fields may be also given attributes. It is therefore easy to match the organization sketched above in section 2. Data and relevant metadata grouped in one group can be distributed

LEAPS INNOVATION
Co-funded by the European Union
Superflat is part of the LEAPS-Innovation project, which has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 101004728

into semantically distinct subgroups. When the entry skeleton is in defined, specific data can be directly accessed (read/write) by their paths inside the hierarchy.

3.2.2 Reliable high-level open-access software

HDF5 is a de-facto standard file architecture in the LEAPS facilities. It is platform independent, and benefits from well-developed and maintained open-access software tools, which take charge of all implementation burden, and expose simple high level read/write functions. In particular, the interface with Python is extremely simple and intuitive. A very large part of the acquisition and analysis software used in our facilities can read and write HDF5 files.

3.2.3 Standardized

Inside the HDF5 framework, Nexus provides a supplementary organization layer, in the form of classes, the structure of which is standardized. Each group of a NeXus file must have a NXclass attribute indicating the class name. A class defines a minimal structure of entries and attributes which must be present and several optional entries. Compliance with these data schemes ensure interoperability. The standard is open; new classes can be added to offer new structures or extend existing ones. Such extensions are intended to be eventually included in the standard if supported by a user community.

3.2.4 Non-redundancy and multi-access

It appears in the inventory of section 2, that some parts of the data could be logically attached to several entries. In order to avoid data duplication, the HDF5 architecture allows to write the data only once and to preserve access from any of these entries, by using a link mechanism. The link is a logical pointer to the location of the physical data. Data are written only once and the link which point to them, is the only duplicated object. Two different data paths can therefore lead to the same storage location. The NeXus overlayer however adds a location tag by adding an explicit "*target*" attribute pointing to the entry which physically holds the data.

3.2.5 Extensible and self-documented

Custom additions to the standard Nexus definitions are not limited by the format. New entries and attributes can always be added to the standard ones when needed. A reader program will simply ignore undocumented entries, but attributes may be used to internally document these additions. Browser software is able to navigate into the file, despite its unknown structure. A definition file can be given to the browser to document the data and metadata storage schema

4 Propositions for a NeXus based data management schema

In NeXus terminology, the data management scheme applicable to a particular class of scientific application is called "**application definition**". Many application definitions are already registered by NeXus organisation and can be found on NeXus website, but none is directly applicable to the whole

Superflat is part of the LEAPS-Innovation project, which has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 101004728

set of data and metadata listed in section 2. A special **application definition** suitable for optical surface metrology must be therefore agreed upon.

Another issue is that the NeXus base classes are not immediately suitable to store complete surface metrology data and metadata, especially for recording the raw data and the applied processing, and the inclusion of some sample data, such as fiducial marks, may require significant extension to the standard NXsample class.

However, a large part of the data enumerated in section 2, can perfectly fit into the standard base-classes listed on NeXus website provided special fields are added to them. These additions must be described in the **application definition** schema.

4.1 Principles and strategy

Fixing a complete **application definition** schema including all the kind of measuring instruments and measuring techniques, is a desirable target but it carries high risks that its complexity be dissuasive to adopt the file standard. It is more reasonable to define a progressive strategy, starting from simple cases and progressively extending it to more complex cases such as the inclusion of raw stitching data. A progressive strategy will allow validation time between steps and provide better opportunities to tailor the standard to the needs.

A minimal **application definition** should allow to access unambiguously the final result of the metrology. In a first step, the scheme will only apply to final consolidated data, in view of facilitating the exchange between LEAPS partners and with their suppliers.

- Starting from use cases, interferometric height map or LTP or NOM profiles, define the strictly necessary elements and their organisation.
- Keep provision for extending the same basic structure to more complex cases (SHarper, stitching) or to private uses such as raw data storage.

4.1.1 Respect the recommendations of the NeXus group

- Strictly adhere to the NeXus terminology and organization conventions. In particular, the data:NXdata group should be present in the default path.
- Use the standard NeXus base classes whenever they can be applied. When extra fields are needed in a class, they are defined in the **application definition** scheme.
- Use standard attributes.
- Use explicit chaining paths. No name parsing should be required to find a data group
- Use links for cross-referencing data without duplication.

4.1.2 Specific rules and classes

- Follow the logical structure described in section 2.
- Define specific new classes only when required (e.g.: fiducial marks), and keep them as general as possible.

LEAPS INNOVATION
Co-funded by the European Union
Superflat is part of the LEAPS-Innovation project, which has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 101004728

- Make a large use of comment attributes to document the data, especially instrument specific data.
- Use links to provide an access to all data which can be logically connected to the current open group.

4.2 Propositions for a data management schema

The division of data and metadata into 4 categories, measurement data, sample, instrument, and environment corresponds to the standard architecture of a NXentry group inside a NeXus data file. NeXus allows several entries to be present in the file, but only one should be addressed as default data, others should be considered as complimentary entries. Each entry should have a field named "definition" identifying the application definition scheme to which it complies.

The **application definition** scheme organizes the internal structure of the different groups inside a NXentry group. It specifies the hierarchy of classes of the subgroups and the names and types of the data fields and attributes inside each groups.

4.2.1 Minimal structure for metrology data

A minimal scheme is shown in figure 1. Each NXentry complying to the proposed NXmetrology application definition should have, at least, the following four direct child groups:

1. A NXdata group named **data**

This group contains the final metrology results in the numeric array fields **height**, **slope_x** and **slope_y**, containing the optical surface height and height gradients with respect to the sampling grid direction, either in 1D or 2D arrays. At least one of these fields is required; the others are optional. **x** and **y** are 1D numeric arrays containing the abscissa and ordinate values of the sampling grid[2]. The dependency of the data fields on **x** and **y** is indicated by group attribute *axes*, as defined by the NeXus NXdata standard. One of these two field is required at least. Note that NeXus standard expects **x** and **y** tables, rather than a starting value and a stepsize.

If the surface reference frame is not coincident with the sample reference frame, three optional fields, **origin_x** , **origin_y** and **rotation** define the transform of the measurement frame with respect to the sample frame. If these values are not given the two frames are assumed to be identical.

All numeric data fields should have their *unit* attributes set to a proper value.

All fields can be links to another storage location inside (or outside[3])the file. The whole **data** group itself can be a link to another NXdata group of the file, which contains the required data fields.

2. A NXsample group named **sample:**

[2] The use of the two 1D arrays means that 2D data are implicitly assumed to be distributed over a rectangular grid in the sample reference frame XY (surface) plane.

[3] Nevertheless, at least for consolidated surface data, for simplicity, it is suggested to save all information in a single file.

LEAPS INNOVATION
Superflat is part of the LEAPS-Innovation project, which has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 101004728
Co-funded by the European Union

The sample group should primarily describe the optical surface being measured. The optical surface is assumed to be essentially lying in the x, y coordinate plane, the z direction being the direction of the height measurements. By convention, the z direction should be pointing outside of the optics.

The sample group defines the reference frame in which the measurement is done. If a **fiducials** group is present, the origin and orientation of this frame is defined by the x, y coordinates of the fiducial points. The corners of the optical surface can be used as fiducial points, if it has a rectangular shape. If no fiducial group is present, the reference frame is assumed to be oriented with the x direction aligned with the long dimension; the origin is assumed to be at the bottom left corner when viewing normally toward the surface and the x axis will be aligned with the bottom edge. The x, y, z trihedra must be direct.

The standard NXsample base class described in the NeXus documentation is intended to be very versatile, but is oriented toward the description of an object submitted to X-ray or neutron beams. Despite the large number of proposed optional fields, only the **name** and **description** fields are pertinent for an optical surface. In case the object has an identification number or reference, it is suggested to add an **ID** alphanumeric field.

The **orientation** field, documents how the surface is oriented during the measurement, with respect to the vertical, for eventual gravity sag correction. It can take the following enumerated string values:
- "top" means surface is facing upwards and viewed from top;
- "side" means the surface is facing laterally and viewed from the side;
- "bottom" means the surface is facing downward and viewed from below;
- "transform" means the surface orientation is defined by an NXtransformation group defining the general coordinate transform from instrument frame to sample frame (see NeXus documentation).

The **clear_aperture** field is an array numeric parameter defining a clear aperture, usually a rectangle. In this case, it is proposed the parameter be x_min, y_min, x_size, y_size. If the meaning is different this should be specified by a *format* attribute. The field should have a *unit* attribute.

LEAPS INNOVATION
Co-funded by the European Union
Superflat is part of the LEAPS-Innovation project, which has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 101004728

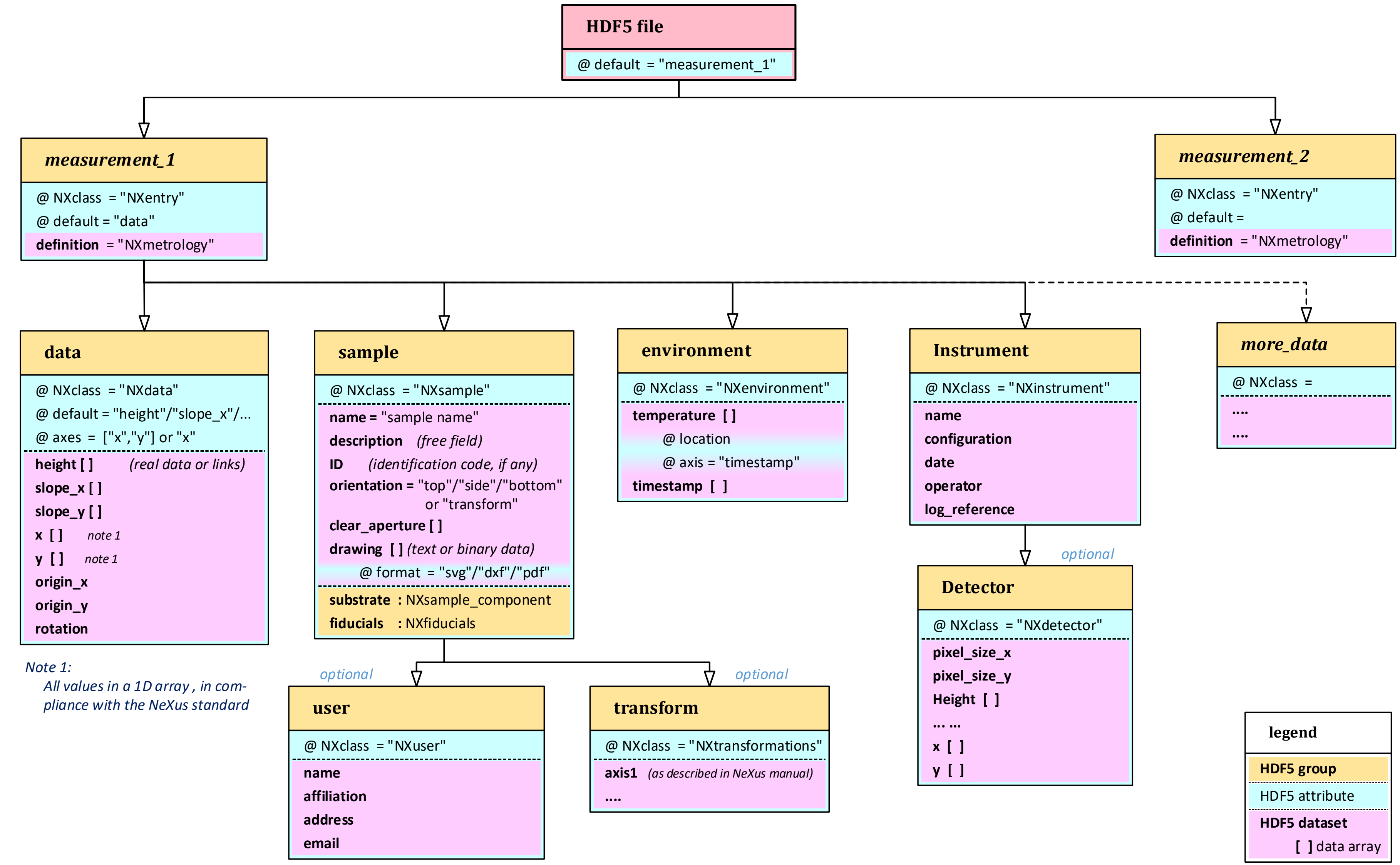

Figure 1 The data structure corresponding to NXmetrology application definition.

Superflat is part of the LEAPS-Innovation project, which has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 101004728

Co-funded by the European Union

The HDF5 architecture allows ascii or binary character array fields. Such fields can store the content of small data files, especially text files. It is suggested to use this feature to store a simple drawing of the sample surface, directly in the file, through a character array field named **drawing**. This field will have a *format* attribute, indicating the file extension of the original file. It is recommended to avoid binary file formats and rather use SVG or DXF formats which are only using plain ascii. PDF format can be also accepted, though it contains some non ascii binary blocks.PDF files are usually bigger than SVG or DXF files, which should be preferred instead.

The above fields describe the optical surface, but it is also useful to have information on the substrate, shape, material, and also crystalline orientation, as this is regularly needed for evaluating a gravity sag or a thermal expansion. It is suggested this information is included in a **substrate** group of class NXsample_component. A proposal for the **substrate** and the **fiducials** groups is given below in a separate paragraph.

Finally, a standard NXuser group can optionally add information about the owner of the artefact affiliation, contact information, manufacturer etc. ...

3. A NXinstrument group named **instrument:**

The **instrument** group is the standard location to place the information about the instrument and the particular configuration, if applicable, used for the measurement. It is also a convenient place to store information on the measurement session, date, operator, logbook reference. These fields are informational and can be used liberally.

To the **instrument** group is usually attached a **detector** group, class NXdetector. It will keep information on the detection geometry such as magnification, aperture, and pixel size. The detector group can be a location to store measurement raw data, which will be linked to from other locations in the file.

4. A NXenvironment group named **environment:**

The **environment** group is also a standard location in a NeXus file. This is a location to place parameters which can have an influence on the measurement. As an example, the actuator positions of a deformable mirror for the considered measurement. For data recorded in stitching mode, the same group can record the collected values of temperatures during the whole measurement period with the associated timestamp value of each temperature point. A *location* attribute may refer the location where these temperatures were collected.

4.2.2 The **substrate** and **fiducials** children of the **sample** group

As said before, the main purpose of the sample group is to define the optical surface which is the measurement subject. It is also useful to define the artefact bearing this surface especially if gravity sag correction is intended. It is proposed here to put the corresponding information in a child group of **sample**, named **substrate** and having the class NXsample_component. As shown on figure 2, this group will define the substrate material, with an optional attribute to specify the crystalline orientation of the optical surface, the shape of the blank and appropriate parameters for defining this shape. Note that thickness always refers to the direction normal to the optical surface. If needed a drawing can be added in the same way as defined for the **sample** group.

LEAPS INNOVATION
Co-funded by the European Union
Superflat is part of the LEAPS-Innovation project, which has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 101004728

We propose to group information related to fiducial marks which may exist on the surface in a separate child group of **sample**. The group will be essentially a collection of fields giving the coordinates, in principle measured values, of the fiducial points in the sample frame. The name of these fields is free; their location on the surface could be described by a *location* attribute and a *marker* attribute could describe their aspect.

If the sample is rectangular, the corners and borders might be considered as fiducials. A simple naming convention for these points seems adequate. We suggest names like **corner_sw** which are shorter though as explicit, as corner_bottom_left. If one coordinate is unknown, or undefined as for borders, it should be written as a NaN.

Finally a drawing could be added to the fiducials group, to explicit the location of the listed points and give their theoretical positions.

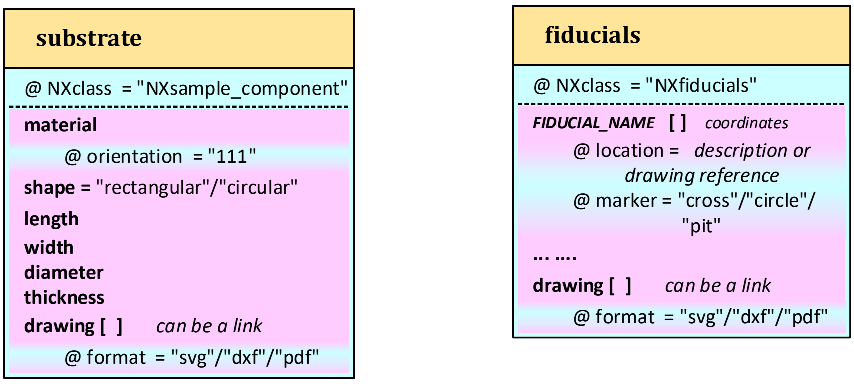

Figure 2. Diagrams of the **substrate** and **fiducials** child classes of the **sample** group.

This **fiducials** group could be given a special class name, such as NXfiducial, though the general purpose NXcollection class, defined in NeXus base classes, could be convenient as well.

4.2.3 Keeping track of data processing

The final metrology result is often a height or a slope error, which is the difference between the absolute height or slope of the surface and some model. As different model can be applied, to the same absolute data it is useful to keep track not only of the final result but also of the original data.

The NeXus framework provides a mechanism to keep track of the data processing chain illustrated in figure 3. A NXdata group resulting from some processing can be given a NXprocess child group documenting the kind of processing applied by means of its **program** field and the corresponding processing parameters. The source data are given as a child group of the NXprocess group. If processing was only applied in a subarea of the source data grid, this will be documented with a NXregion group.

LEAPS INNOVATION
Co-funded by the European Union
Superflat is part of the LEAPS-Innovation project, which has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No. 101004728

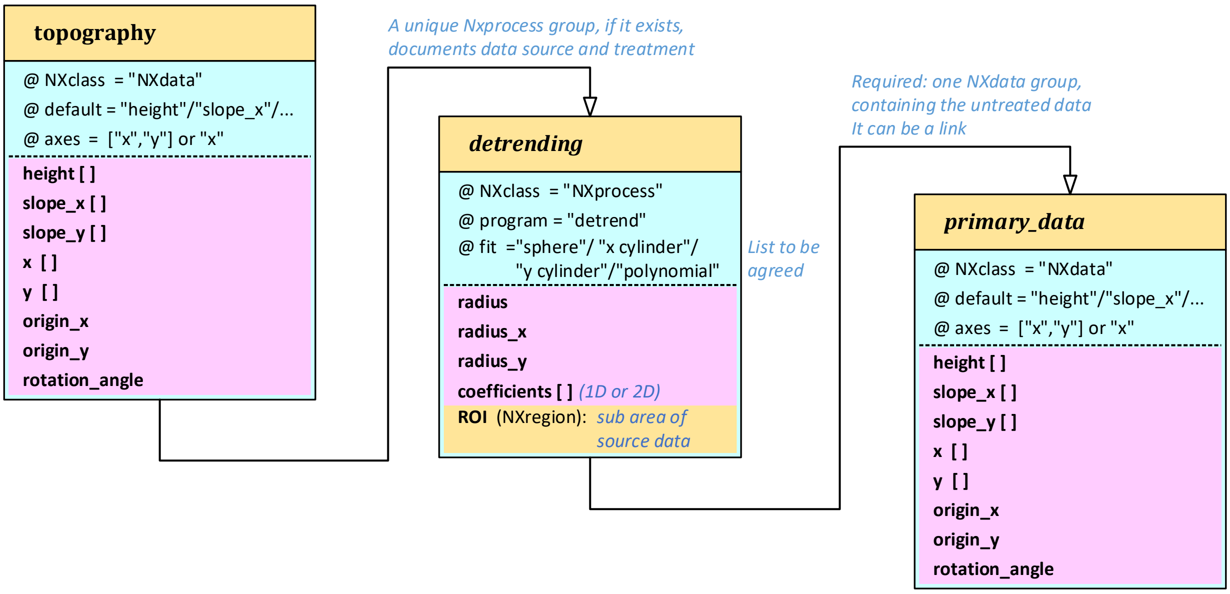

Figure 3. Example of a NXdata and NXprocess chain for keeping track of some data processing

Several usual processes can be easily implemented as a start, such as detrending, spatial filtering and gravity sag correction, as they all rely on a unique source data group. The set of parameters required by each "*program*" needs to be agreed on later. Extension to processes depending on more than one data group, e.g. stitching, could be implemented in a second stage.

The tracking of successive processes applied to the data could be maintain by chaining process and, optionally, intermediate data group. Note that intermediate data groups are not requested in the chain for complete tracing of the data processing.

5 Conclusion and outlook

Starting from the observation that a common and universal file format for storing optical surface metrology data would be most helpful for exchanging reference data and develop common processing software, we inventoried the data and metadata which should be made available in such files and evaluated their requirement level. It led us to propose a progressive strategy to develop the format starting from the simpler case of storing final consolidated data, while remaining able to extend later to unprocessed data generated by specific instruments and various metrology methods.

An agreement was easily achieved on the use of HDF5 file format as the physical storage architecture. It is also proposed to comply to the NeXus format standard for the internal organisation and data management scheme of the files. These choices fulfil a fundamental requirement of data management to be FAIR (i.e. making the stored data Findable, Accessible, Interoperable and Reusable).

From this baseline a first draft of data management specification has been constructed, initially limited to final metrology results but able to store height and slope data within the same organisation scheme. Provision is also made to keep track of simple data processing of the raw data, such as spatial filtering or detrending.

LEAPS INNOVATION
Co-funded by the European Union
Superflat is part of the LEAPS-Innovation project, which has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 101004728

The next steps will involve a concrete implementation of the format for storing real metrology data. It could be first applied to metrology results produced at the end of phase 2 of Superflat PCP, where each of the three participant companies has produced and measured a “metrology sample”, which was also measured by four metrology laboratories of the consortium using different instruments. These measurement results are available under different formats, but have been processed, for cross comparison’s sake, by a common python script. Extending this script to read and write in the newly defined format will be a first test of applicability. In longer term, the NXmetrology storage format could be implemented in the PyLOSt software, developed at ESRF, which is becoming a standard for processing stitching interferometric measurements. It is expected that experimenting with the format will raise some issues and necessitate certain modifications of the present scheme. We are confident that the flexibility of NeXus and HDF5 will favour further improvements and extensions.

LEAPS INNOVATION
Co-funded by the European Union
Superflat is part of the LEAPS-Innovation project, which has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No. 101004728