

# An HDF-5/NeXus based file format for STXM

Benjamin Watts, Joerg Raabe

Swiss Light Source, Paul Scherrer Institute, 5232 Villigen, Switzerland

## **Introduction**

Scanning Transmission X-ray Microscopy (STXM) is a relatively recent and rapidly growing X-ray imaging technique with niche capabilities that are applicable across a wide set of scientific disciplines. The basic principle of STXM involves focusing an energy-tunable, monochromatic X-ray beam onto a raster-scanned sample while measuring the transmitted X-rays as a function of sample position. The position of the X-ray spot (typically smaller than 50 nm) on the sample is usually measured via interferometer to very high precision ( $>1$  nm) in order to avoid image distortions. The chief advantage of STXM lies in the use of soft X-ray spectroscopy to obtain strong, natural contrast in high-resolution imaging. Commonly utilised contrast mechanisms include spectroscopic effects based on elemental composition, oxidation state[A,B], molecular structure,[G,] molecular orientation (via linear dichroism)[K,L,M] and magnetisation (via X-ray magnetic circular dichroism). Furthermore, trace elements can be mapped via fluorescence detection,[C,D] electron detection has been used to map surfaces [E,F], dynamics can be imaged in pump-probe experiments,[G,H] and area detectors allow for differential-phase contrast imaging[I,J] and ptychography. The sample mounting in STXMs also tend to be quite accessible and numerous examples exist of environmental cells and in-situ operation of technological devices. [N,O,P,Q,R,S,T] The many combinations of contrast mechanisms, detection schemes, scanning modes/patterns and sample configurations demonstrate the enormous range of STXM experiments.

Despite the wide variety of STXM experiments, there remain some elements that are ubiquitous; the aim of STXM experiments is always the acquisition of images and spectra and so most STXM data tend to be measured via the following scan types:

1. Point spectrum: a scan through photon energies at a single spatial point on the sample.
2. Line spectrum: a set of spectra, measured in parallel, at a set of spatial points forming a line across the sample.
3. Image: a 2-dimensional scan of the sample (or OSA or detector) at a single photon energy.
4. Image stack: a set of images of the same sample area over a set of photon energies.
5. Focus: repeated scans across a feature on the sample (or OSA) at a set of zoneplate positions.

The above scan types represent the typical operation of a STXM for both instrument alignment and scientific data acquisition. Currently, the file format in broadest use for STXM data is the *self describing format* (SDF) developed by Peter Hitchcock that combines a JSON-like metadata file coupled with separate files containing ascii tables of the measured values.

Over the past few years, a collaboration between the Paul Scherrer Institute, the Max Planck Institute for Intelligent Systems, and Semafor Informatik & Energie AG has developed a new STXM control software called *Pixelator*. [in preparation] *Pixelator* is designed to have a clean, modular interface in order to accommodate and fully integrate future extensions to STXM instruments such as novel scanning modes and detector types. Alongside the new instrument control capabilities of *Pixelator* comes the requirement to store all the types of data that may be produced by *Pixelator*-controlled systems so that the data file format doesn't limit the capabilities of the STXM.

Our aim in defining the NXstxm file format has three aims:

1. have sufficient flexibility to accommodate any scan type
2. the set of common, standard scans should be easy to read and plot

▼	Sample_Stack_Example.hdf5	
▼	entry1	NXentry
	definition	
	start_time	
	end_time	
	title	
▼	counter0	NXdata
	count_time	
	data	
	photon_energy	
	sample_x	
	sample_y	
	scan_type	
▶	counter1	NXdata
▶	analog_input0	NXdata
▶	ring_current	NXmonitor
	data	
	photon_energy	
	sample_x	
	sample_y	
▼	instrument	NXinstrument
▶	entrance_slit	NXaperture
▶	exit_slit	NXaperture
▶	source	NXbendmagnet
▶	counter0	NXdetector
▶	mirror	NXmirror
▶	sample_x	NXdetector
▶	sample_y	NXdetector
▶	facility	NXsource
▶	zone_plate	NXzoneplate
▼	sample	NXsample
	rotation_angle	
▼	collection	NXcollection
▶	beam_shutter	NXpositioner
▶	coarse_x	NXpositioner
▶	coarse_y	NXpositioner
▶	detector_x	NXpositioner
▶	detector_y	NXpositioner
▶	detector_z	NXpositioner
▶	entrance_slit	NXpositioner
▶	exit_slit_h	NXpositioner
▶	exit_slit_v	NXpositioner
▶	fine_x	NXpositioner
▶	fine_y	NXpositioner
▶	osa_x	NXpositioner
▶	osa_y	NXpositioner
▶	photon_energy	NXpositioner
▶	polarisation	NXpositioner
▶	ring_current	NXpositioner
▶	ring_y_asym	NXpositioner
▶	scan_request	
▶	...	

- encourage saving all beamline and instrument parameters

These aims align well with those of NeXus,[U,V] however the variety of scan types produced by STXM instruments makes for some conflict between aims #1 and #2. Existing NeXus application definitions have side-stepped this challenge by tending to define separate application definitions for different scan types. Following this strategy for STXM would either be very limiting in the types of defined scans, or produce an unreasonably large number of application definitions and quite likely allowing a lot of STXM data to fall in between the formally defined scan types. Therefore, we have opted to make NXstxm very flexible in order to incorporate all data produced by STXM, while introducing a few extra rules that provide simplicity in the case where it can be reasonably achieved. The aim of this article is to give a description of the NXstxm format, while discussing the reasoning and intention behind the strategic choices.

## File Hierarchy

The HDF5 container format[W,X] used by NeXus contains a hierarchy of groups and fields that can be considered analogous to the directories and files of a computer file system. HDF5 places no restrictions on how these group and field objects are arranged within the file, but NeXus provides standardising guidelines designed to balance flexibility and accessibility. Figure 1 illustrates a schematic of an NXstxm compliant HDF5 file hierarchy, with the various groups/classes discussed in the following subsections.

Figure 1. Structure of an example NXstxm file written by *Pixelator*. The NeXus class of each HDF5 group is given to the right.

### NXentry

NeXus provides NXentry as a way of allowing multiple measurements in the one file, if there is some reason for grouping them together. A set of measurements that are made in parallel or form a consecutive set, for example, could be written into separate NXentry's within the same HDF5 file. *Pixelator*, and other stxm control programs, allow a user to scan hyperspectral image “stacks” of multiple spatial regions together such that the instrument will execute scans of each regions in sequence using the first photon energy before repeating the set of spatial regions using the next photon energy and so on for each requested

photon energy. In this case, a conforming NXstxm file will contain one NXentry group for each spatial region such that the contents of each NXentry group is identical to the case where that spatial region alone had been scanned.

### **NXinstrument**

The NXinstrument group is where physically meaningful descriptions of the experimental apparatus and a full accounting of the measured data can be found. Following the standard NeXus guidelines, implementations of NXstxm are encouraged to provide a full description of the experiment. Of particular importance to NXstxm are the NXdetector groups, which represent each of the detector devices being operated (including the phosphor/photomultiplier tube point-detector that is most commonly used) as well as the interferometer axes that measure the sample position relative to that of the zone plate. Therefore the detector names:

- sample\_x
- sample\_y
- sample\_z

are reserved for the detectors representing measurements of the sample scan axes. All other detectors can be freely named. Data stored within these NXdetector groups will be written as a list of values, or arrays in the case of detectors returning more than one value per pixel. In other words, detector data is stored as an array with the first dimension corresponding to the time-ordered pixel number and subsequent dimensions corresponding to the output of the detector. Thus a point-detector scanned through a 10 by 10 pixel image will have an NXdetector data array of 100 items in a single dimension, while a 5 by 5 pixel area-detector will have an NXdetector data array of size (100, 5, 5). At the same time, the NXdetector groups corresponding to the interferometer axes will store 100 position values. In this way, the data stored within the NXinstrument group doesn't rely on any assumptions about how the scan is conducted and the data can always be assembled into an image by iterating through the first axis of every NXdetector group to read the position and detector values for each point. The NXinstrument section of the NXstxm application definition therefore follows the general NeXus guidelines very closely.

### **NXdata**

The NXdata groups in NeXus are intended to allow easy access to plottable data by grouping the relevant information together. An NXstxm file should contain one NXdata group per (identically named) NXdetector group present within the NXinstrument group (except for those corresponding to the interferometer axes). The cartesian raster patterns and simple point-detector setups that form the vast majority of STXM measurements today makes for a straightforward application of NeXus principles to constitute the NXdata groups. Every STXM measurement that is of scientific interest will involve X-ray spectroscopy and imaging of the sample, thus NXstxm requires the following data fields always be present in every NXdata group:

- photon\_energy
- sample\_x
- sample\_y

In scans that vary these parameters, then the usual 'axis=n' attribute, where 'n' refers to the dimension of the data array corresponding to the parameter, must be applied. In scans where the parameter is not scanned, then the field must still be present, but containing only a single value and without the 'axis' attribute.

Since STXM encompasses a wide range of possible experiments, with 1D and 2D detectors and/or arbitrary scanning patterns of any number of instrument axes, it is prudent to consider the likely edge cases. In the interests of keeping the data found in NXdata simple to plot, the NXstxm application definition describes NXdata as a simplified view of the measurements. For setups involving detectors that return multiple values, such as a 2D area detector, NXstxm requires that the array of values returned by the detector be summarised into a single value (e.g. the mean) for use in

the NXdata group. (Note that the full array will still be available in the corresponding NXdetector group that resides within the NXinstrument group.) In the case of scan patterns that do not fit into a simple rectangular array easily, or scanning modes that do not prioritise positioning accuracy, it is accepted in NXstxm that the plottable data found within an NXdata group may be distorted or present only an approximate view of the experimental data. Thus the set of values provided in the fields that label the data array dimensions (e.g. sample\_x and sample\_y for an image of the sample) may not be accurate and also the arrangement of values with the data array is not guaranteed to be accurate relative to either the axis labels or to the relative positions of the other array values.

A 'scan\_type' field is also provided in the NXdata groups, containing a string describing the type of scan that has been recorded from the following choices:

- “sample point spectrum”
- “sample line spectrum”
- “sample image”
- “sample image stack”
- “sample focus”
- “osa image”
- “osa focus”
- “detector image”
- “generic scan”

These strings describe the set of typical STXM scans as discussed in the introduction, except for the “generic scan” string, which is to be used whenever none of the others are applicable. Although the same information could be discerned by examining the fields having 'axis' attributes, this 'scan type' field is provided as a convenience for human readers.

### **NXmonitor**

NXmonitor groups in NeXus are nearly identical to NXdata groups, but differ in that they are intended to be used as normalisation information for other data sets, rather than as an independent measurement. In order to account for time-variations in photon flux, STXM measurements are typically normalised against the current of the synchrotron storage-ring, for which NXstxm expects a value per pixel measurement that is to be included as an NXdetector and NXmonitor group with the reserved name “ring\_current”.

### **Coordinate System**

In a STXM instrument, the X-ray beam is typically fixed in position while the sample is raster-scanned to build up the set of pixels that form the measured image. However, the image data is presented in the reference frame of the scanned object, which mirrors the axes. The guiding principles for coordinate system choice in STXM are:

1. Provide image data corresponding to a view of the sample as would be seen from the zone plate.
2. Make positioning along the optical axis easy to understand in order to avoid crashes.

These principles correspond to the coordinate system where:

1. Within the reference frame of the sample/scanned object, the positive Y-axis points upwards and the positive X-axis is perpendicularly to the right when viewed from the zone plate position.
2. The positive Z-axis is parallel to the optical axis, oriented in the direction of X-ray propagation and with the origin at the OSA position.
3. Images are written with a right-handed coordinate system in which +Y points up, +Z points towards the X-ray source and +X is oriented horizontally to the right when viewed from the X-ray source.

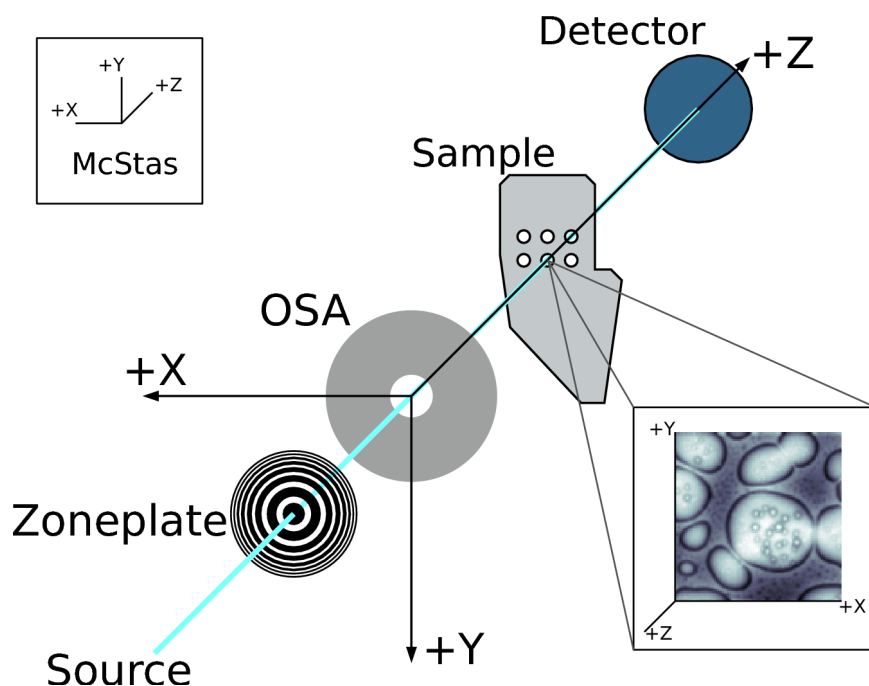


Figure 2. A schematic of a STXM instrument, demonstrating a left-handed coordinate system with +Z oriented in the X-ray propagation direction, +Y pointing down and +X oriented horizontally, to the left when viewed from the X-ray source. However, due to the scanning nature of the instrument, images are written in the reference frame of the scanned object and so the axes are mirrored to produce a right-handed coordinate system (lower right inset). The McStas coordinate system typically used in NeXus is shown in the upper left inset.

Figure 2 illustrates the coordinate system of a typical STXM instrument and the reversed image coordinate system (lower right inset), alongside the McStas coordinate system typically used in NeXus. Note that the origin of the instrument coordinate system is placed at the center of the OSA. Also, the sample mounting plate will typically have positions for up to six separate samples, with the center of the lower middle position being set to (0,0) on the sample X-Y scanning stage. Sample positions are typically separated by 5 mm (centre to centre) with holes of diameter 2.7 mm for easy mounting of both TEM grids (diameter 3 mm) and silicon nitride membranes (with 5 x 5 mm substrates).

## Conclusions

The NXstxm application definition described here has been designed for a balance of convenience and flexibility. While most current STXM analysis software will only wish to read the simple NXdata version of the data, the ability to read a more detailed version of the measurements from the NXinstrument group provides opportunities for more sophisticated use of the data that may become more commonly utilised in future. A disadvantage of this approach is that the data is essentially written twice, although STXM data rates are usually low enough that the larger files are not a significant problem and the increase can also be offset by utilising the compression features that are built into HDF5.

The NXstxm application definition is currently at its starting point and it will inevitably evolve through future versions. The authors encourage all STXM users to provide constructive input and be a part of the NeXus and NXstxm community. Extensions to possibly be included in future versions of the NXstxm application definition include further metadata for detectors in order to automatically indicate the appropriate normalisation procedures.

## References

- [A] J. Miot, K. Benzerara, G. Morin, A. Kappler, S. Bernard, M. Obst, C. Ferard, F. Skouri-Panet, J. M. Guigner, N. Posth, M. Galvez, G. E. Brown, F. Guyot, *Geochim. Cosmochim. Acta* 2009, 73, 696.
- [B] S. Takahama, S. Gilardoni, L. M. Russell, *J. Geophys. Res.-Atmos.* 2008, 113, D14205.
- [K] S. Kowarik, K. Broch, A. Hinderhofer, A. Schwartzberg, J. O. Osso, D. Kilcoyne, F. Schreiber, S. R. Leone, *J. Phys. Chem. C* 2010, 114, 13061.
- [L] T. Schuettfort, B. Watts, L. Thomsen, M. Lee, H. Sirringhaus, C. R. McNeill, *ACS Nano* 2012, 6, 1849.
- [M] B. Watts, T. Schuettfort, C. R. McNeill, *Adv. Funct. Mater.* 2011, 21, 1122.
- [C] A. P. Hitchcock, M. Obst, J. Wang, Y. S. Lu, T. Tyliczszak, *Environ. Sci. Technol.* 2012, 46, 2821.
- [D] B. Kaulich, A. Gianoncelli, A. Beran, D. Eichert, I. Kreft, P. Pongrac, M. Regvar, K. Vogel-Mikus, M. Kiskinova, *J. R. Soc. Interface* 2009, 6, S641.
- [E] B. Watts, C. R. McNeill, *Macromol. Rapid Comm.* 2010, 31, 1706.
- [F] C. Hub, S. Wenzel, J. Raabe, H. Ade, R. H. Fink, *Rev. Sci. Instru.* 2010, 81, 033704.
- [G] H. Ade, H. Stoll, *Nat. Mater.* 2009, 8, 281.
- [H] A. Fernandez-Pacheco, L. E. Serrano-Ramon, T. Tyliczszak, K. W. Chou, R. Cordoba, A. Szkudlarek, L. O. Brien, C. Kapusta, M. R. Ibarra, J. M. De Teresa, *Nanotechnol.* 2012, 23, 105703.
- [I] A. Menzel, C. M. Kewish, P. Kraft, B. Henrich, K. Jefimovs, J. Vila-Comamala, C. David, M. Dierolf, P. Thibault, F. Pfeiffer, O. Bunk, *Ultramicroscopy* 2010, 110, 1143.
- [J] J. Raabe, B. Watts, G. Tzvetkov, R. H. Fink, C. Quitmann, *J. Phys. Conf. Ser.* 2009, 186, 012012.
- [N] T. Huthwelker, V. Zelenay, M. Birrer, A. Krepelova, J. Raabe, G. Tzvetkov, M.G.C. Vernooij, and M. Ammann, *Rev. Sci. Instr.* 81, 2010, 113706.
- [O] C. Hub, M. Burkhardt, M. Halik, G. Tzvetkov and R. Fink, *J. Mater. Chem.*, 20, 2010, 4884–4887
- [P] B Watts, D Queen, A L D Kilcoyne, T Tyliczszak, F Hellman and H Ade, *J. Phys.: Conf. Ser.* **186**, 2009, 012023
- [Q] A. Vecchiola, M. Moskura, P. Haltebourg, P. Guenoun, J. Daillant, A. Madouri, A. Gianoncelli, B. Kaulich, C. Gosse, and C. Chevallard, *Proc. 15th International Conference on Miniaturized Systems for Chemistry and Life Sciences*, 2011, 1305-1307
- [R] H.W. Nho, J.Y. Kim, J. Wang, H.-J. Shin, S.-Y. Choi and T.H. Yoon, *J. Synchrotron Rad.* 21, 2014, 170–176
- [S] S. Kumar, M.D. Pickett, J.P. Strachan, G. Gibson, Y. Nishi, and R.S. Williams, *Adv. Mater.* 25(42), 2013, 6128-6132
- [T] L.R. Aramburo, S. Teketel, S. Svelle, S.R. Bare, B. Arstad, H.W. Zandbergen, U. Olsbye, F.M.F. de Groot, B.M. Weckhuysen, *J. Catal.* 307, 2013, 185–193
- [U] NeXus format (2014 (accessed December 2014)), <http://www.nexusformat.org/>
- [V] NeXus format article in preparation...
- [W] The HDF Group, HDF-5 (2014 (accessed December 2014)), <http://www.hdfgroup.org/HDF5/>
- [X] T. Habermann, A. Collette, S. Vincena, J.J. Billings, M. Gerring, K. Hinsien, W. Benger, F.R.N.C. Maia, S. Byna and P. de Buyl, 2nd Workshop on Sustainable Software for Science: Practice and Experiences, 2014, <http://dx.doi.org/10.6084/m9.figshare.1112485>
- [Y]
- [Z]