MSA/MAS/AMAS Hyper-Dimensional Spectral File Format - An Update

Aaron Torpy[1], Nick Wilson[1], Colin MacRae[1], Nestor J. Zaluzec[2] and Mike Kundmann[3]

1. Microbeam Lab, CSIRO Process Science and Engineering, Box 312, Clayton South, VIC, 3169
2. Electron Microscopy Center, Argonne National Laboratory, Argonne, IL 60439 USA
3. e-Metrikos, P.O. Box 5506, Pleasanton, CA 94566, USA

The past decade has seen the widespread adoption of a new generation of analytical equipment that is capable of collecting 'hyper-dimensional' datasets. Typical hyper-dimensional datasets include hyper-spectral maps, in which a complete spectrum is collected at every point in a map (XEDS, EELS, CL, TOF-SIMS, etc.), and 'hyper-image' maps, in which an image is collected at every point in a map (EBSD, CBED, etc). Datasets with yet higher dimensionality may be generated by 3D serial sectioning, tomography, or time-resolved techniques.

As there are many techniques that now generate hyper-dimensional data, so too are there many (often proprietary) file formats used to store the datasets. At present, there is no agreed standard file format for hyper-dimensional data files that has been adopted by equipment vendors or the microanalysis software development community. Consequently, it is difficult to exchange hyper-dimensional datasets between different software packages, which hinders both data interpretation and collaboration between laboratories. The cornucopia of proprietary file formats also poses problems for long-term archiving of data.

A comparable situation existed in the early 1990's with spectrum data from XEDS and EELS, which lead to the development of the EMSA/MAS spectral file format that is now an established ISO standard [1]. Two decades later, the MSA Standards Committee is similarly developing a new standardised file format for hyper-dimensional data. A preliminary design for the proposed standard file format was presented at the M&M, ACMM and AMAS meetings in 2011-2013 [2- 5]. Feedback from researchers and vendors at those meetings has been supportive and constructive, resulting in a number of improvements and refinements to the proposed specification.

The proposed format uses two files, which are associated by file name and by a unique identifier. One file is binary, with a file extension of '.hmsa', and stores only the raw experimental data. The second file is an eXtensible Markup Language text file (.xml), which stores the relevant experimental conditions and settings, as well as a strict definition of the layout of the data in the binary file. This dual-file design has the following advantages:

- The XML file is human-readable and self-descriptive, allowing all relevant experimental conditions to be stored in a logical format that is easy to read by both humans and software.
- Binary encoding of the experimental data ensures reasonable space efficiency for hyper-dimensional datasets, and may be conveniently written and read by software.
- The structure of the binary file format is simple, unambiguous, and precisely defined within the XML file.
- Reading and writing HMSA/XML files requires no specialised libraries, and can be implemented in most programming languages on most operating systems.

A specification document has been prepared for further discussion with the microscopy community [6]. The specification defines the baseline implementation level that is required to read and write valid HMSA/XML documents, and provides a lengthy but non-exhaustive list of further optional XML entities that may be used to describe experimental conditions and other relevant metadata.

To assist the initial implementation of the HMSA file format, example HMSA datasets have been generated for a range of analytical techniques, including SEM (EDS), TEM (EELS), and EPMA (EDS, CL, WDS, BSE). Typical example datasets include the mandatory definition of the data in the binary file, plus a set of optional entities such as accelerating voltage, beam current, scan magnification, dwell time, and spectrometer calibration. Baseline example datasets are also available, which contain only the minimum information required to properly load the binary data file, and do not include any optional experimental conditions or other metadata. Finally, extended example files are provided to demonstrate a broad range of HMSA capabilities, including an extensive list of experimental settings and conditions, multiple datasets in a single file (e.g. parallel EDS and CL spectral maps), Unicode character support for non-European languages, and so forth.

Prototype software code to read and write HMSA/XML files has been developed to validate the design of the format, and to provide an initial reference implementation for developers. The reference implementation includes simple utilities to perform common hyper-dimensional data processing tasks, such as generating region of interest images from spectral maps, and integrating sum spectra from selected areas within spectral maps. These utilities include a set of reusable libraries for automating common HMSA tasks such as generating, parsing, and validating HMSA/XML documents. The source code and compiled binaries for these utilities are available on an open source basis, and are currently implemented in the C++ language for the Windows platform.

Interested parties wishing to comment or contribute to the specification development should contact the Standards Committee's chair, Nestor J. Zaluzec (zaluzec@aaem.amc.anl.gov).

References

[1] R.F. Egerton *et al.*, Proceedings of the Electron Microscopy Society of America, San Francisco Press, (1991) 526.

[2] M. Kundmann *et al.*, Microscopy and Microanalysis, **17** Suppl. 2, (2011) 860. Doi: 10.1017/S1431927611005174

[3] N. Wilson *et al.*, Proceedings of APMC 10 / ICONN 2012 / ACMM 22, 6-9 Feb., 2012, Perth, Western Australia, 773-1.

[4] N. J. Zaluzec *et al.*, *MSA/MAS/AMMS Hyper-dimensional Data File Format - An Update*, Microscopy and Microanalysis, **18**, Suppl. 3 (2012).

[5] A. Torpy *et al.*, *MSA/MAS/AMAS Hyper-dimensional Data File Format - An Update*, Proceedings of AMAS XII, 4-8 Feb., 2013, Sydney, Australia, 76-77.

[6] Available at http://www.amc.anl.gov/ANLSoftwareLibrary/MSAMASFormat/