

Standardization of the MSA/MAS/AMAS Hyper-Dimensional File Format

Aaron Torpy¹, Mike Kundmann², Nick Wilson¹, Colin MacRae¹, and Nestor J. Zaluzec³

¹. Microbeam Lab, CSIRO Mineral Resources, Private Bag 10, Clayton South, VIC, 3169, Australia.

². e-Metrikos, P.O. Box 5506, Pleasanton, CA 94566, USA.

³. Electron Microscopy Center, Argonne National Laboratory, Argonne, IL 60439 USA.

In 2010, the Standards Committee of the Microscopy Society of America (MSA) formed a working group comprising members of the MSA, the Microbeam Analysis Society (MAS) and the Australian Microbeam Analysis Society (AMAS) to develop a standardized file format to facilitate the exchange of microscopy datasets with high dimensionality, such as hyperspectral maps. The proposed file format, known as the MSA/MAS/AMAS hyper-dimensional data file (HMSA, for short), was presented to the community for comment at the M&M 2011 meeting in Nashville, TN [1], and revisions incorporating feedback from researchers and vendors was presented at subsequent meetings [2-5]. After a further round of generalisation and simplification, a finalised specification (HMSA v1.01) was approved by the MSA Standards Committee in 2016 [6], which have now commenced the formal standardisation process with the International Standards Organisation (ISO).

The primary consideration when designing the HMSA format was to ensure the greatest ease of implementation, so as to enable the rapid and widespread adoption of the format by vendors and researchers. However, simplicity was not the only design criterion, as to be fit for purpose the format must be reasonably compact, and sufficiently flexible to support a wide range of experimental techniques and measurement modes, including those not yet envisioned. To balance these competing considerations, the HMSA format was split into two files; a simple binary format (file extension of 'HMSA') to efficiently store the hyperdimensional data in full fidelity, and an accompanying eXtensible Markup Language (XML) file [7] that defines the size, location and ordering of data in the binary HMSA file, as well as describe any experimental parameters that may be relevant.

The binary HMSA file contains one or more datasets, which are represented as regular arrays (of any dimensionality) stored as raw uncompressed data. Whilst this simple format may be easily implemented in any programming language or environment, it is also extremely flexible, as it can support spectra (1D), images (2D), hyperspectral maps or serial-section images (3D), 'hyper-image' maps with an image at every pixel (4D), and higher dimensional data. Conveniently, this structure of the binary dataset also enables efficient data processing on very large files by permitting random seeking through the dataset on disk. In the interests of simplicity, the HMSA binary file does not support internal data compression, as this may require the use of external algorithms or libraries. The binary HMSA file was deliberately designed to not require the use of external algorithms or libraries, as these may not be available in all data processing environments or operating systems. Hence, the format should be applicable to the broadest range of platforms, both now and in the future.

The XML file is used to define the layout of the binary data in the HMSA file in a human-readable format, which would not be possible if this information was stored within the binary file itself. This information may be read (or written) using basic text editor software, which allows researchers to create or interpret these files without the need for any specialised software. The XML language is an

international standard that is widely supported by data processing software and programming libraries, thereby allowing the format to be efficiently parsed and interpreted by software also.

The XML file is divided into three sections. The first section is the 'header', which may be used to record non-scientific metadata such as the creation date and owner. The second section is the 'conditions' list, which contains a set of well-defined data structures to record relevant experimental parameters to assist with the scientific interpretation of the measured data. For example, parameters relating to the configuration of an XEDS detector may be declared in a `<Detector Class="XEDS">` condition. The contents and interpretation of this condition are defined in the appendices of the HMSA specification, along with conditions for other common experimental parameters (e.g. analytical probe, specimen environment, experimental measurement modes, etc.) To ensure that the HMSA specification can expand to meet the evolving needs of the community, additional definitions of experimental conditions may be submitted to the HMSA working group for review, and accepted updates will be published on-line [6]. The third section of the XML file is the 'dataset' definition, which states the size, position, and ordering of the data in the accompanying HMSA binary file. The dataset definition follows a minimalist format that contains no optional or descriptive information, resulting in a structure that is universal for datasets of any dimensionality or experimental technique.

The detailed specification of the HMSA format is available on the MSA Standards website [6], along with links to further resources to assist in the development of HMSA-compatible software, such as example data files, software libraries, and various utilities for manipulating HMSA/XML files.

References

- [1] M. Kundmann, N. Wilson, A. Torpy, N.J. Zaluzec, *Microscopy and Microanalysis*, 17 Suppl. 2, (2011) 860. Doi: 10.1017/S1431927611005174
- [2] N. Wilson, A. Torpy, C. MacRae, M. Kundmann, N. J. Zaluzec, *Proceedings of APMC 10 / ICONN 2012 / ACMM 22*, 6-9 Feb., 2012, Perth, Western Australia, 773-1.
- [3] N. J. Zaluzec, M. Kundmann, N. Wilson, A. Torpy, *Microscopy and Microanalysis*, 18, Suppl. 3 (2012).
- [4] A. Torpy, N. C. Wilson, C. M. MacRae, N. J. Zaluzec, and M. Kundmann, *Proceedings of AMAS XII*, 4-8 Feb., 2013, Sydney, Australia, pp76-77.
- [5] A. Torpy, N. C. Wilson, C. M. MacRae, N. J. Zaluzec, and M. Kundmann, *Microscopy and Microanalysis* 19, Suppl. 2 (2013), pp830-831. DOI: 10.1017/S1431927613006144
- [6] HMSA specification is available at http://www.microscopy.org/resources/scientific_data/
- [7] XML is a standard maintained by the World Wide Web Consortium at www.w3.org/TR/xml/