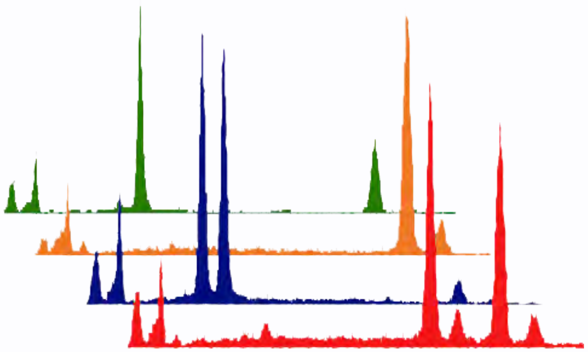
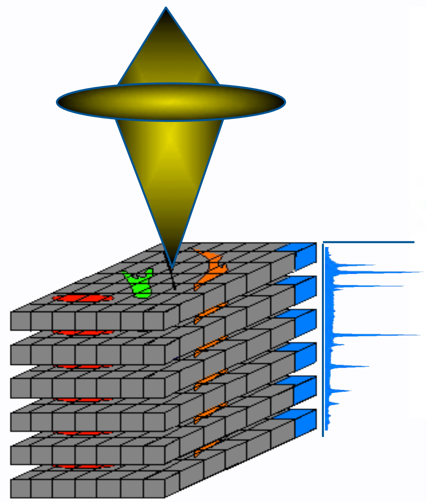
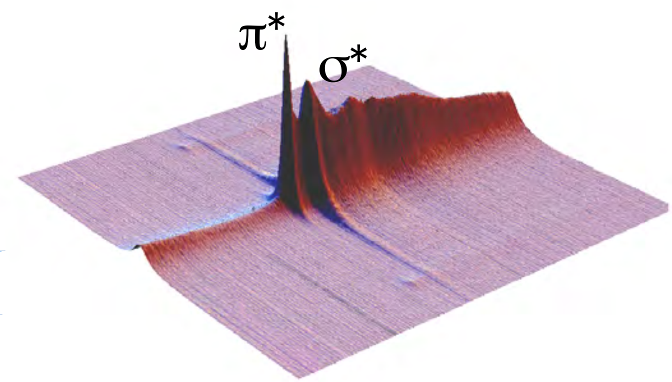
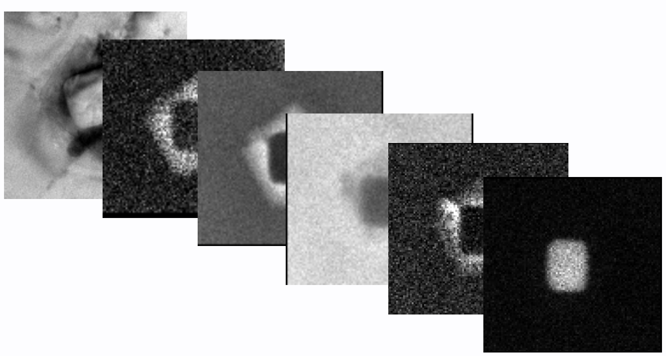
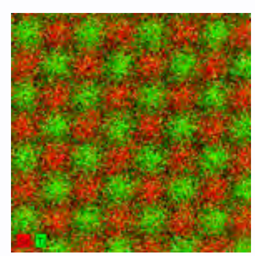
MSA/MAS/AMAS Hyper-dimensional Data File Format - An Update

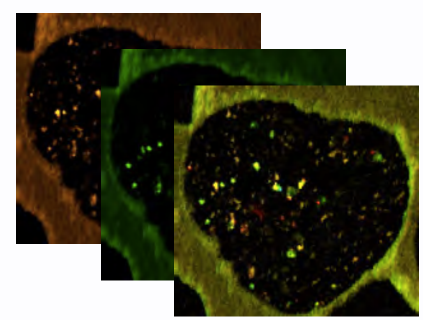












Nestor J. Zaluzec, Mike Kundmann, Nick Wilson**, Aaron Torpy***
*Argonne National Laboratory, *e-Metrikos,, **CSIRO Process Science and Engineering*
Microscopy Society of America : Standards Committee

Origins: MSA/MAS Standards Committee (1991)

Ray F. Egerton--------University of Alberta
Charles E. Fiori-------Nat. Inst. for Science & Technology
John A. Hunt-----------Lehigh University
Michael S. Isaacson-Cornell University
Earl J. Kirkland--------Cornell University
Nestor J. Zaluzec-----Argonne National Laboratory

Goal: To define a standard spectral file exchange format for use by the MSA/MAS Community

1. Represents the data exactly.

2. The format should be simple and easy to use.

3. It must NOT be tied to any particular computer, programming language or operating system

4. The format should be both human and machine (computer) readable.

5. It should be compatible with existing electronic communication networks

6. The format should support spectra of interest to the EMSA/MAS community

7. Each file should contain enough information to uniquely identify the type and origin of the spectral data and to reconstruct its significance.

8. Where possible, the format should be compatible with various commercial data plotting or analysis programs

9. The proposed format need not be the most efficient storage mechanism. Its primary goals, simplicity and ease of use.

Origins: MSA/MAS - Spectral File Format (1991)

ASCII File Format (Filename.msa)

Header Lines
*
*
Successive lines begining with MSA/MAS defined keywords. Some of which are required and some are optional
*
*
Start of Data Keyword
*
*
Experimental Data
*
*
End of Data Keyword

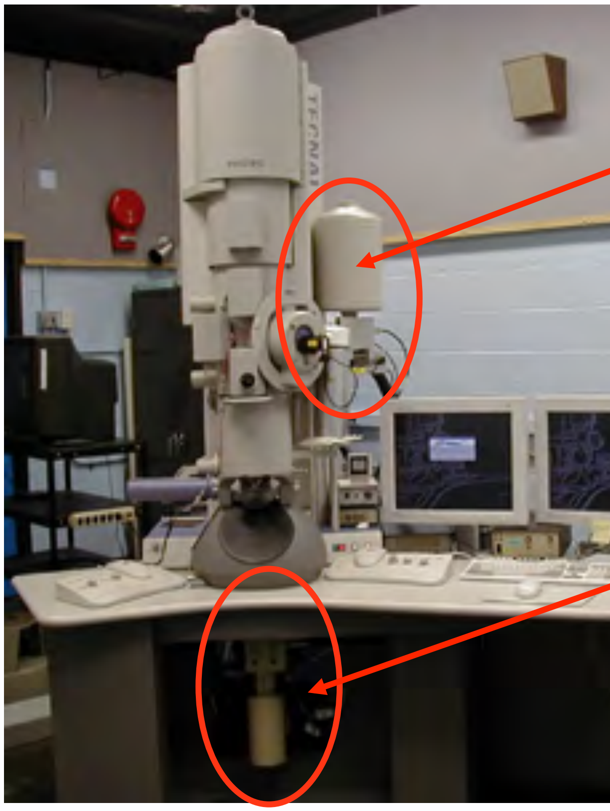
R.F. Egerton, C.E. Fiori, J.A. Hunt, M.S. Isaacson, E.J. Kirkland, N.J. Zaluzec, Proceedings of the Electron Microscopy Society of America, San Francisco Press, (1991) 526.

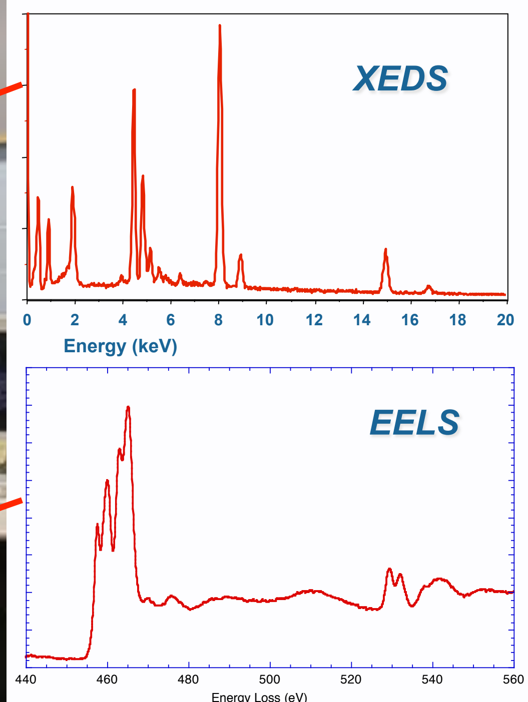
Now : ***ISO 22029:2003***.

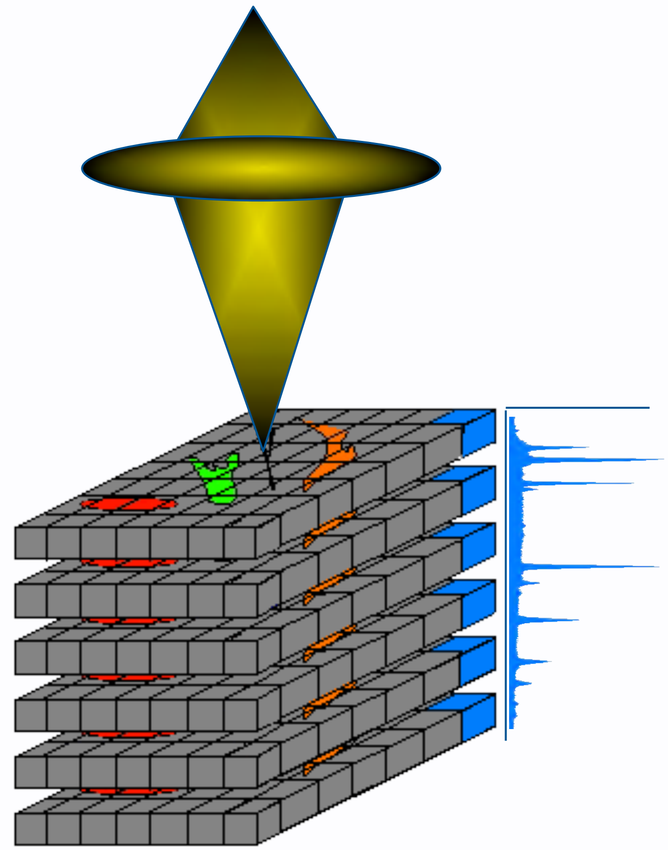
```
#FORMAT      : EMSA/MAS Spectral Data File
#VERSION     : 1.0
#TITLE       : NIO EELS OK SHELL
#DATE        : 01-OCT-1991
#TIME        : 12:00
#OWNER       : EMSA/MAS TASK FORCE
#NPOINTS     : 10.
#NCOLUMNS    : 1.
#XUNITS      : Energy Loss (eV)
#YUNITS      : Intensity
#DATATYPE    : XY
#XPERCHAN    : 3.1
#OFFSET      : 520.13
#CHOFFSET    : -168
#SIGNALTYPE  : ELS
#XLABEL      :  Energy
#YLABEL      :  Counts
#BEAMKV   -kV: 120.0
#EMISSION -uA: 5.5
#PROBECUR -nA: 12.345
#BEAMDIAM -nm: 100.0
#MAGCAM      : 100.
#CONVANGLE-mR: 1.5
#COLLANGLE-mR: 3.4
#OPERMODE    : IMAG
#THICKNESS-nm: 50.
#DWELLTIME-ms: 100.
#ELSDET      : SERIAL
#SPECTRUM    : Spectral Data Starts Here
520.13,      4066.0
523.22,      3996.0
526.32,      3932.0
529.42,      3923.0
532.51,      5602.0
535.61,      5288.0
538.70,      7234.0
541.80,      7809.0
544.90,      4710.0
#ENDOFDATA   :
```

Today's Computationally Mediated Experiments

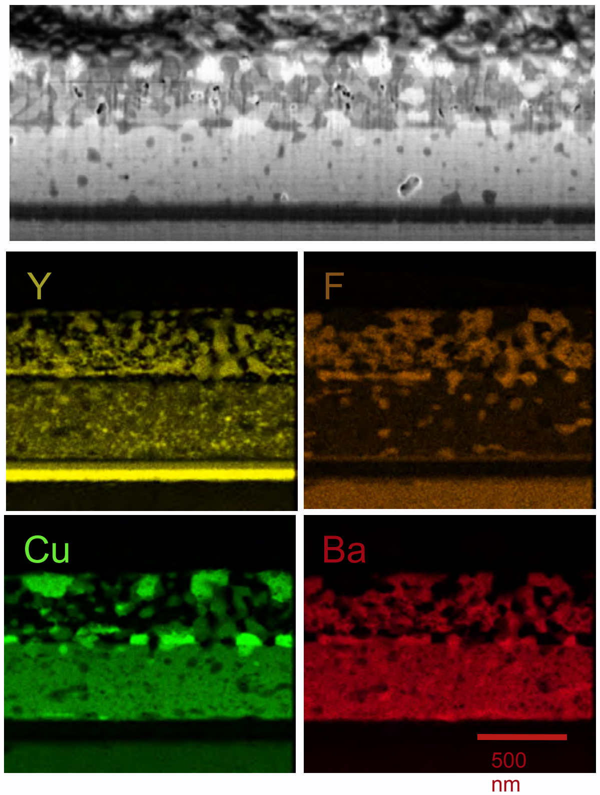
Can no longer be reasonably handled using this format



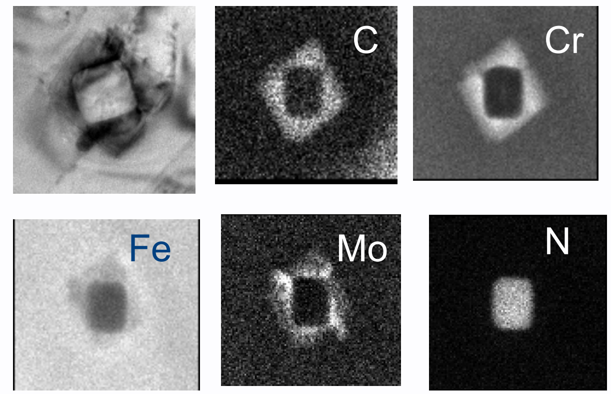




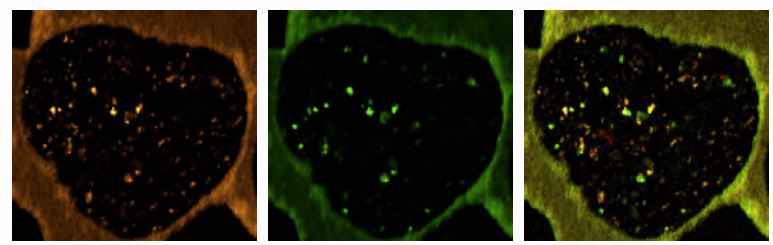
Example Hyper-Spectral Imaging



XEDS

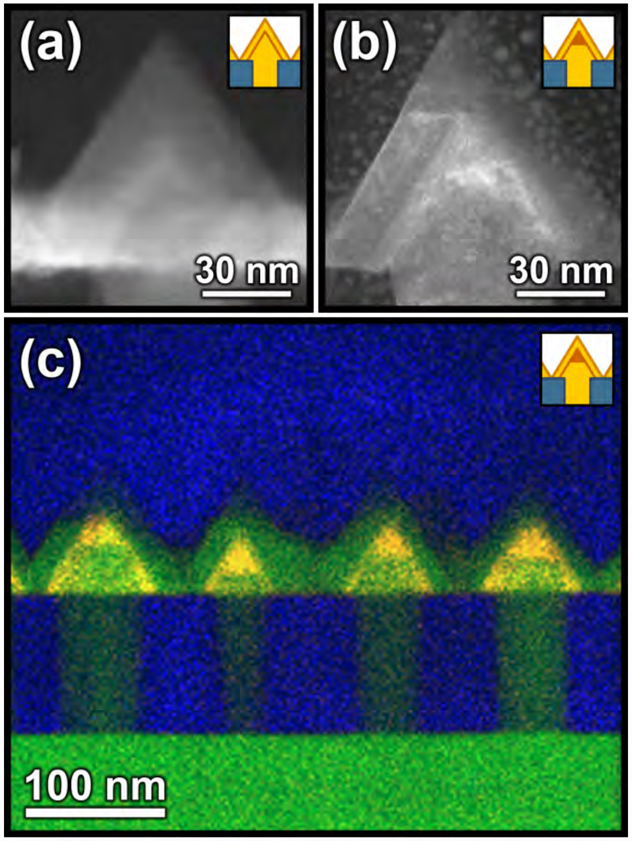


EFTEM/EELS

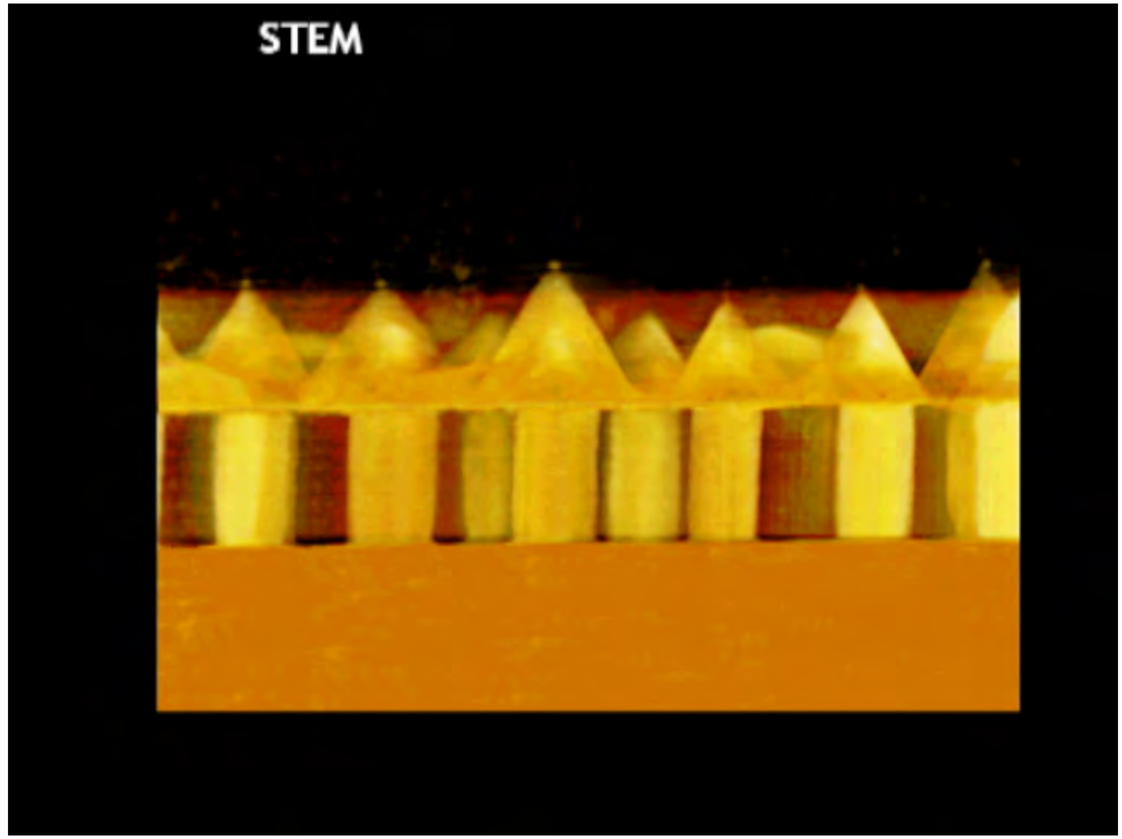


Cathodoluminescence

Hyper Spectral Tomogram of III-nitride Nanopyramid LEDs



Conventional 2D



3D Spectral Tomogram

Multi-dimensional data sets are no longer practical using the *20 year old* format designed for individual spectra.

Hyper-Dimensional Spectral File Format (2012)

- Hyper-dimensional data acquisition is now a common mode of operating in the microscopy and microanalysis community. Techniques that generate such data sets include hyper-spectral XEDS and EELS mapping (aka spectrum imaging), EFTEM, tomographic tilt series, EBSD and CBED, TEM through-focal series, and in-situ dynamic time series among many others.

- Despite the growing importance and prevalence of such large hyper-dimensional data sets, there has been, as yet, no commonly recognized standard file format to contain them. Such data are currently saved in many different program- and vendor-specific formats, some of which are proprietary.

- This poses problems for the long-term archiving of the data, as well as the sharing and comparative analysis of results between different labs and software packages.

Hyper-Dimensional Spectral File Format (2012)

Design Trade Offs

Simplicity, Transparency, Speed of Access,
Size, vs Hierarchical Databases/Repository

Must be Intuitive at the human interface level
Self-Descriptive to facilitate coding

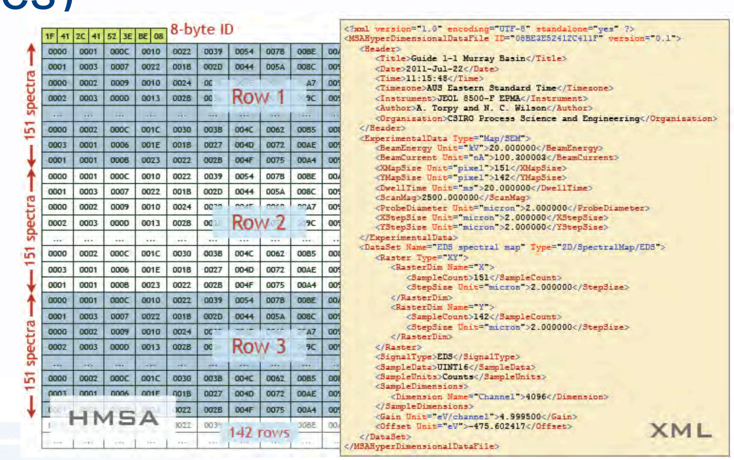
Pure Text vs Pure Binary

Text: size, I/O penalties
Binary: requires apriori knowledge of structure

Solution Implement as Dual Structure (2 Files)

XML: Descriptive Information
Binary: Raw Data Block

BaseFilename.xml
BaseFilename.hmsa



Hyper-Dimensional Spectral File Format (2012)

The following requirements were considered in the design of this file format:

- Modern experimental apparatus produce data with high dimensionality, such as a spectral maps, and 3D serial section maps. Therefore, this file format must store data of high dimensionality.
- High dimensionality data is necessarily very large, and consequently difficult and time consuming to store or transfer over networks. The file format must therefore be as compact as is reasonably practical.
- Many microanalytical techniques produce structurally similar hyper-dimensional data. To simplify implementation of common tools, this file format must use a common format to store data produced by different analytical techniques.
- The data format must preserve the scientific accuracy and meaning of the data. Therefore, the file format must store data without loss of precision, and include sufficient experimental parameters to permit the correct interpretation of the data.
- To achieve the intended mission of being a widely-supported exchange format, the file format must achieve acceptance from instrument and software vendors, and from the microanalysis community. Consequently, the file format must be useful, easy to understand, and easy to implement.
- Furthermore, as the file format is intended for exchange, it must be readable (and implementable) in any commonly available programming language/environments, The format must therefore be platform independent, and not require any proprietary or special software or hardware.

Hyper-Dimensional Spectral File Format (2012)

To satisfy these requirements, the MSA/MAS/AMAS Hyper-dimensional Data File format uses a pair of files; a simple binary file to efficiently store the experimental data, and a text-based XML file to store ancillary information, experimental settings, and a rigorous description of the layout of the binary file. The advantages of this dual format are:

- The structure of the binary file format is simple, unambiguous, and precisely defined in a human readable format within the XML file.
- High-dimensionality experimental data is binary-encoded for space efficiency, while also being easy to read and write programmatically.
- Experimental settings and conditions are stored in a human-readable and self descriptive format. Settings are stored in a hierarchical structure to logically classify related settings.
- No special libraries are required to read or write HMSA/XML files. For convenience, XML libraries may be used, and are freely available in most programming environments.
- The XML file format supports the use of the unicode character set, permitting native-language representations of names for authors, organizations, specimens, locations, etc. For maximum interoperability, the default language of the XML file is US English, and any international strings must include an alternative US English translation.

Hyper-Dimensional Spectral File Format (2012)

A simple, common container for line scans, spectrum images, tomography stacks, and other types of spectral and microscopy image series

Format ***(****BaseFilename.xml****)***

```
<MSAHyperDimensionalDataFile>
        <Header>
        *
        *
        *
        </Header>
        <ExperimentalData>
        *
        *
        *
        </ExperimentalData>

        <DataSet>
        *
        *
        *
        <DataSet>
</MSAHyperDimensionalDataFile ">
```

ID & Ownership Section

Experimental Conditions Section

Binary Data Storage Description

Hyper-Dimensional Spectral File Format (2012)

Example of a Header Section

```
<Header>
    <Analyst>    Person doing the Experiment  </Analyst>
    <Affiliation> Affiliation of the Analyst </Affiliation>
    <HostFacilty> The Host Facility where the experiment was being done </HostFacilty>
    <Location>  location of data collection host facility</Location>
    <Date>  YYYY-MM-DD </Date>
    <LocalTime> HR:MIN:SEC </LocalTime>
    <GMTzone> GMT-6.00 </GMTzone>
    <DataDescription> Description/Identification of the Data </DataDescription>
</Header>
```

Example of a Experimental Parameters Section

```
<ExperimentalParameters>

    <Instrument Name="SomeInstrument "> </Instrument>

    <OperatingMode> </OperatingMode>

    <StageCoordinates> </StageCoordinates>

    <Probe Name = "Electron | Ion | Photon | Mechanical " > </Probe>

    <Detector Name="Example" ID="D1"> </Detector>

    <OtherUserDefinedAsNeeded> </OtherUserDefinedAsNeeded>

    <DataCollectionSoftware> </DataCollectionSoftware>

    <DataCollectionSoftwareVersion> </DataCollectionSoftwareVersion>

</ExperimentalParameters>
```

Hyper-Dimensional Spectral File Format (2012)

Example of a Header Section

```
==========================================================================================
This describes the Analyst, HostLocation,  and Specimen/Sample
==========================================================================================


<Header>

    -------------- required sub elements ------------

    <Analyst>    Person doing the Experiment  </Analyst>

    <Affiliation> Affiliation of the Analyst </Affiliation>

    <HostFacilty> The Host Facility where the experiment was being done </HostFacilty>

    <Location>  location of data collection host facility</Location>

    <Date>  YYYY-MM-DD </Date>

    <LocalTime> HR:MIN:SEC </LocalTime>

    <GMTzone> GMT-6.00 </GMTzone>

    <DataDescription>
        Description/Identification of the Data and/or Sample text description
        this field is free form text and can be as long as needed
    </DataDescription>

    -------------- optional sub elements ------------

    <Client> If this experiment is being done for someone else </Client>

    <ClientAffiliation> The affiliation of the Client </ClientAffiliation>

    <Comment>    general comments about the data or specimen studied </Comment>

</Header>
```

Hyper-Dimensional Spectral File Format (2012)

Example of a Experimental Parameters Section

```xml
<ExperimentalParameters>
    <Instrument Name="SomeInstrument "> </Instrument>
    <OperatingMode> </OperatingMode>
    <StageCoordinates> </StageCoordinates>
    <Probe Name = "Electron | Ion | Photon | Mechanical " > </Probe>
    <Detector Name="Example" ID="D1"> </Detector>
    <OtherUserDefinedAsNeeded> </OtherUserDefinedAsNeeded>
    <DataCollectionSoftware> </DataCollectionSoftware>
    <DataCollectionSoftwareVersion> </DataCollectionSoftwareVersion>
</ExperimentalParameters>
```

Hyper-Dimensional Spectral File Format (2012)

Example of an Experimental Parameters Sub-Section

```
<Probe Name = "Electron | Ion | Photon | Mechanical " >
    <ProbeParameters>

    ------ note the parameters herein will vary depending upon the specific Probe type --------
    ------- These parameters describe the probe/source used  in the instrument ------------
    ------  and which may affect or be used during analysis/quantification of  the data --------

        <Type> Thermionic | ThermalAssistedFEG | FEG | SFEG | CFEG | PhotoCathode | UV | IR |
        Visible | X-ray | Laser | He-Ion | Ga-Ion | AtomicForce | MagneticForce | ScanningProbe |.... </Type>
        <Energy Units = "kV"> 30 </Energy>
        <EnergySpread Units = "eV"> 1 </EnergySpread>
        <WaveLength Units = "nm"> 10 </WaveLength>
        <WaveLengthSpread Units = "nm"> 10 </WaveLengthSpread>
        <Intensity Units = "nA"> 10  </Intensity>
        <Polarization> None </Polarization>
        <TemporalMode> Continuous | Pulsed | ...  </TemporalMode>
        <TemporalResolution Units = "ns" > 10 </TemporalMode>
        <Shape> Parallel | Gaussian | Lorentian | Focussed | ..... </Shape>
        <ConvergenceHalfAngle Units = "mR"> 10 </ConvergenceHalfAngle>
        <DiameterFWHM Units = "nm"> 10 </DiameterFWHM>
        <DiameterFWTM Units = "nm"> 10 </DiameterFWTM>
        <OtherParameterUserDefinedAsNeeded> ...  </OtherParameterUserDefinedAsNeeded>
    </ProbeParameters>
</Probe>
```

Hyper-Dimensional Spectral File Format (2012)

Example of a Data Section

```
<DataSets>
    <datum> ••• </datum>
    <datum> ••• </datum>
</DataSets>
```

Hyper-Dimensional Spectral File Format (2012)

Example of a Datum sub-Section

```xml
<datum>

    <measurementType>Spectrum</measurementType>
    <detectorID>D1</detectorID>
    <binaryLayout>

     <hyperDimensionalSampling>
        <measurementDimensionality>0|1|2|3</measurementDimensionality>
        <collectionDimensionality>0|1|2|3|4</collectionDimensionality>
        <collectionSequence>systematic</collectionSequence>
        <collectionSpacing>constant</collectionSpacing>
      </hyperDimensionalSampling>

      <binaryOffset>8</binaryOffset>
      <binaryElementType>uint</binaryElementType>
      <binaryElementSize>4</binaryElementSize>
      <binaryElementUnit>counts</binaryElementUnit>
      <binaryElementLabel>e- intensity</binaryElementLabel>

     <measurementDimension>
        <dimensionSize>2048</dimensionSize>
        <dimensionOrigin>0</dimensionOrigin>
        <dimensionStep>0.05</dimensionStep>
        <dimensionUnit>mR</dimensionUnit>
        <dimensionLabel>X scattering angle</dimensionLabel>
    </measurementDimension>

     <collectionDimension>
        <dimensionSize>64</dimensionSize>
        <dimensionOrigin>0</dimensionOrigin>
        <dimensionStep>2.0</dimensionStep>
        <dimensionUnit>nm</dimensionUnit>
        <dimensionLabel>Probe X position</dimensionLabel>
      </collectionDimension>

    </binaryLayout>

</datum>
```

Hyper-Dimensional Spectral File Format (2012)

Dimensionality of the Data is a New Issue

Define Collection & Measurement Dimensions

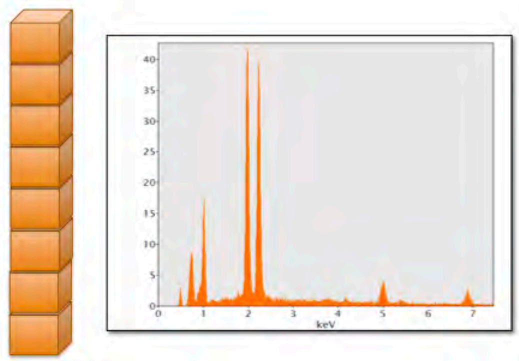
	0D measurement	1D measurement	2D measurement
0D collection	A single datum	A single spectrum acquisition	A single 2D image acquisition (e.g. micrograph, or a diffraction pattern) **
1D collection	A linescan or time sequence of single-valued data (e.g. Ti Kα counts, BSE yield, vacuum pressure.)	A linescan or time sequence of spectra.	A linescan or time sequence of 2D data.
2D collection	An X/Y map of single-valued data (e.g. a intensity of an image)**	An X/Y hyperspectral map (i.e. one spectrum per pixel)	An X/Y 'hyperimage' map (i.e. one image per pixel)
3D collection	An X/Y/Z serial section map of single valued data.	An X/Y/Z hyperspectral serial section map	An X/Y/Z hyperimage serial section map.
4D collection	A 3D Collection as a function of 4th parameter such as time	An X/Y/Z/t hyper spectral section map	A X/Y/Z/t hyperimage serial section map.

Hyper-Dimensional Spectral File Format (2012)

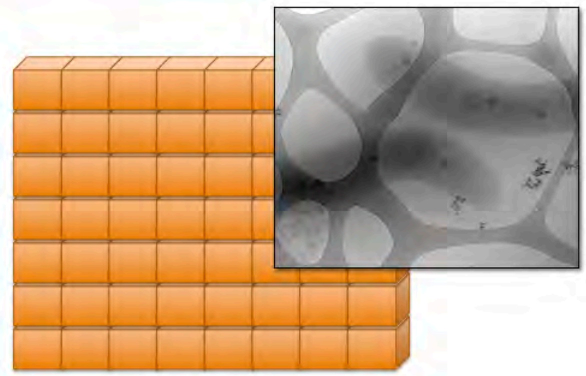
Dimensionality of the Data is a New Issue

	0D measurement	1D measurement	2D measurement
0D collection	A single datum	A single spectrum acquisition	A single 2D image acquisition (e.g. micrograph, or a diffraction pattern) **

0D M – 0D C



1D M – 0D C

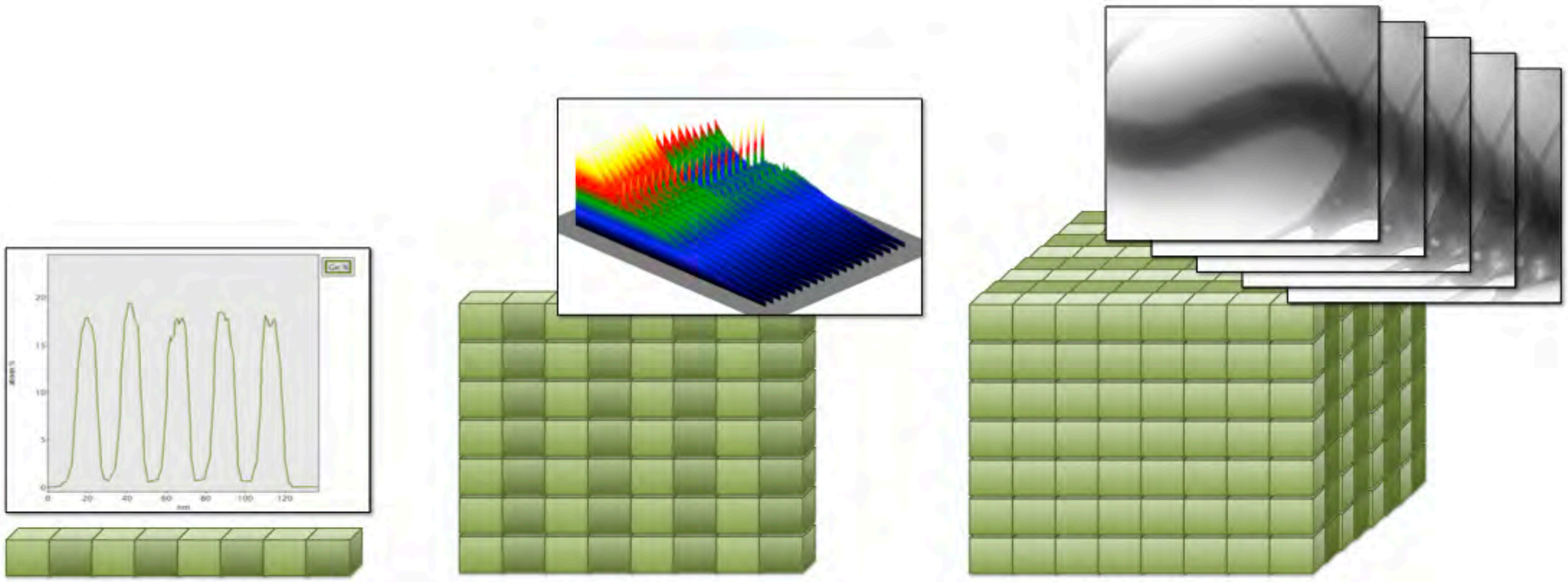


2D M – 0D C

Hyper-Dimensional Spectral File Format (2012)

Dimensionality of the Data is a New Issue

	0D measurement	1D measurement	2D measurement
1D collection	A linescan or time sequence of single-valued data (e.g. Ti Kα counts, BSE yield, vacuum pressure.)	A linescan or time sequence of spectra.	A linescan or time sequence of 2D data.



0D M –1D C

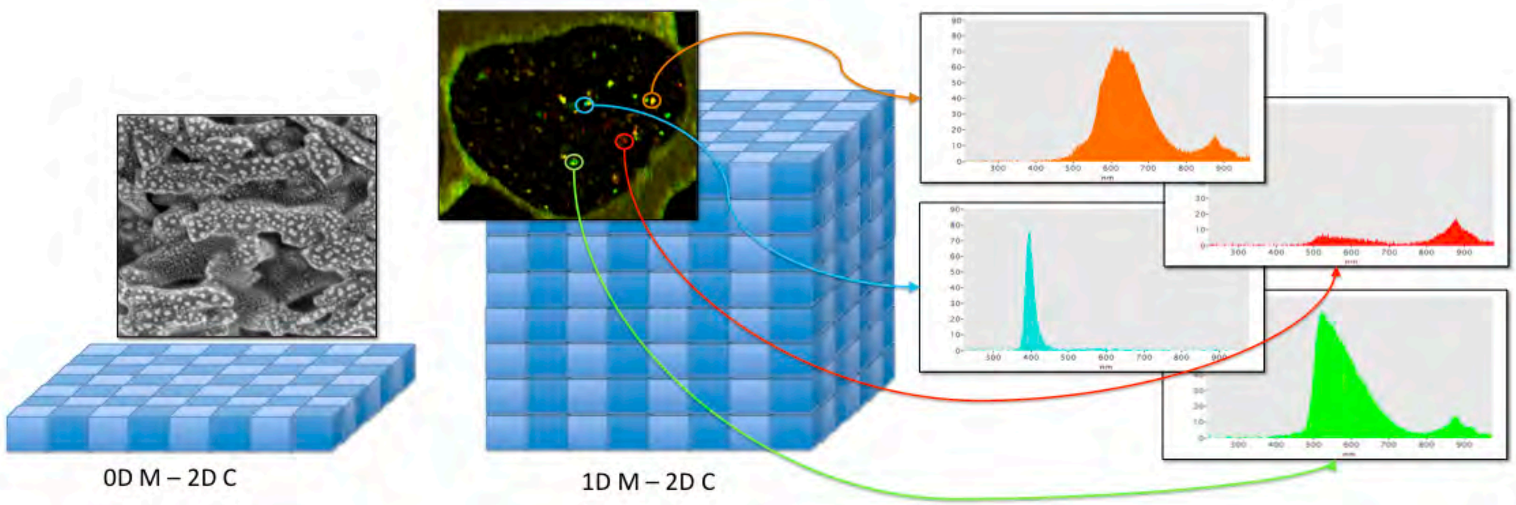
1D M – 1D C

2D M – 1D C

Hyper-Dimensional Spectral File Format (2012)

Dimensionality of the Data is a New Issue

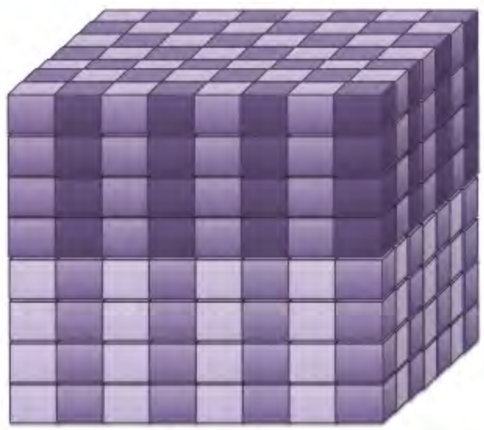
	0D measurement	1D measurement	2D measurement
2D collection	An X/Y map of single-valued data (e.g. a intensity of an image)**	An X/Y hyperspectral map (i.e. one spectrum per pixel)	An X/Y 'hyperimage' map (i.e. one image per pixel)



Hyper-Dimensional Spectral File Format (2012)

Dimensionality of the Data is a New Issue

	0D measurement	1D measurement	2D measurement
3D collection	An X/Y/Z serial section map of single valued data.	An X/Y/Z hyperspectral serial section map	An X/Y/Z hyperimage serial section map.



1D M – 3D C

Hyper-Dimensional Spectral File Format (2012)

- Not intended to replace manufacturer's proprietary format
- Does not address multi-file , entire experiments, or shared hierarchical databases
- No compression specified (but this can be done after the fact)
- The format descriptor / XML tags are being refined
- Input from the community is still being accepted
  send to: MSA Standards Committee Chair
  zaluzec@aaem.amc.anl.gov
  Detailed Specifications will be available on-line ~ Late 2012
  http://www.amc.anl.gov/ANLSoftwareLibrary/MSAMASFormat
- Will be ultimately submitted to ISO to compliment
  - ***ISO 22029:2003*** individual spectral file format

Thanks

Questions to:

zaluzec@aaem.amc.anl.gov