XMP metadata for HDF5
Motivation

- Thumbnail Previews
  - Sorting data files is much easier with representative images

- Metadata
  - Critical to usefulness of archived data
  - Allows efficient sorting of data files

(this slide is pretty boring without any pictures, isn't it)
Boring List of Data Files (KDE4)
Thumbnail Previews (KDE4)
XMP – Extensible Metadata Platform

- Standard designed by Adobe
- Since early 2012, XMP is also an ISO standard (16684-1)
- Used in many common file formats
  - Images: DNG GIF JPEG JPEG 2000 PNG TIFF SVG
  - Dynamic media: ASF FLV AVI MOV MP3 MPEG-2 MPEG-4 SWF
  - Video packages: AVCHD P2 Sony-HDV XDCAM
  - Adobe applications: AI INDD INDT PSD
  - Markup documents: HTML XML
  - Documents: PDF PS EPS UCF
- Expressed as an XML string
- Can be embedded in file, or written to sidecar *.xmp file
• Each operating system has a plugin system to extract and present thumbnail previews of different file types.

• Implementing previews for a new file type requires a new plugin that fetches the data and passes it to the OS.
XMP in HDF5 User Block

- HDF5 has option to leave some space near the beginning of the file that is not interpreted by the HDF5 library.
  - Referred to as the «user block»
  - Can access without using the HDF5 library
- Use magic string and a block size declaration
  - Fast identification and reading of XMP data
  - Allows other data to share the HDF5 user block

<table>
<thead>
<tr>
<th>Decimal:</th>
<th>137</th>
<th>72</th>
<th>77</th>
<th>80</th>
<th>13</th>
<th>10</th>
<th>26</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hexadecimal:</td>
<td>89</td>
<td>48</td>
<td>4d</td>
<td>50</td>
<td>0d</td>
<td>0a</td>
<td>1a</td>
<td>0a</td>
</tr>
<tr>
<td>ASCII:</td>
<td>\211</td>
<td>H</td>
<td>M</td>
<td>P</td>
<td>\r</td>
<td>\n</td>
<td>\032</td>
<td>\n</td>
</tr>
</tbody>
</table>
A standard set of metadata key:value pairs for scientific data is still needed
https://github.com/nexusformat/hdf5xmp
- Code for thumbnailers
- Python code to insert/extract metadata and thumbnails

https://github.com/nexusformat/hdf5xmp/releases
- Binaries for thumbnailers

Support for:

- Windows 7, 8.1, 10, macOS, Gnome and KDE4
Many thanks to

programming
Marc Schädeli
SEMAFOR AG

funding
Swiss Light Source