

PAUL SCHERRER INSTITUT



Benjamin Watts :: PolLux Beamline Scientist :: Paul Scherrer Institut

XMP metadata for HDF5

- Thumbnail Previews
 - Sorting data files is much easier with representative images
- Metadata
 - Critical to usefulness of archived data
 - Allows efficient sorting of data files

Boring List of Data Files (Win10)

The screenshot shows a Windows File Explorer window titled 'thumbed'. The address bar indicates the path: 'Flask > Documents > thumbed'. The left sidebar shows the navigation pane with 'Documents' selected. The main area displays a list of 66 files, all with the .hdf5 extension, arranged in two columns. The files are sorted by name, showing a mix of 'Sample_Image' and 'Sample_Line' files with various dates and identifiers, as well as 'Sample_Stack' files.

82k_SC_48hN2_Image_L_x.hdf5	Sample_Image_2018-09-15_171_x.hdf5	Sample_Stack_2019-02-28_038_x.hdf5
82k_SC_48hN2_Image_x.hdf5	Sample_Image_2018-09-15_174_x.hdf5	Sample_Stack_2019-02-28_059_x.hdf5
82k_SC_as_Stack_x.hdf5	Sample_Image_2019-01-29_127_x.hdf5	
Detector_2019-04-02_195_x.hdf5	Sample_Image_2019-01-29_197_x.hdf5	
Focus_2018-09-14_059_x.hdf5	Sample_Image_2019-01-30_014_x.hdf5	
Focus_2018-10-01_018_x.hdf5	Sample_Image_2019-01-30_062_x.hdf5	
Focus_2018-11-26_018_x.hdf5	Sample_Image_2019-01-30_298_x.hdf5	
Focus_2018-11-26_096_x.hdf5	Sample_Image_2019-01-30_328_x.hdf5	
Focus_2018-11-26_281_x.hdf5	Sample_Image_2019-01-30_396_x.hdf5	
Focus_2019-01-31_141_x.hdf5	Sample_Image_2019-01-31_019_x.hdf5	
Focus_2019-04-02_222_x.hdf5	Sample_Image_2019-01-31_110_x.hdf5	
Focus_2019-04-02_237_x.hdf5	Sample_Image_2019-01-31_142_x.hdf5	
OSA_2017-03-29_078_x.hdf5	Sample_Image_2019-04-02_187_x.hdf5	
OSA_2018-06-13_017_x.hdf5	Sample_Image_2019-04-02_217_x.hdf5	
OSA_2018-06-13_022_x.hdf5	Sample_Image_2019-04-02_258_x.hdf5	
OSA_2018-10-01_092_x.hdf5	Sample_Image_2019-05-20_070_x.hdf5	
PLV_190402003_x.hdf5	Sample_Image_2019-05-20_090_x.hdf5	
Sample_Image_2017-07-26_292_x.hdf5	Sample_Image_2019-05-20_115_x.hdf5	
Sample_Image_2017-07-26_314_x.hdf5	Sample_Image_2019-05-21_021_x.hdf5	
Sample_Image_2017-07-26_333_x.hdf5	Sample_Line_2018-06-13_055_x.hdf5	
Sample_Image_2017-07-26_350_x.hdf5	Sample_Line_2018-06-13_057_x.hdf5	
Sample_Image_2017-07-26_351_x.hdf5	Sample_Line_2018-06-13_066_x.hdf5	
Sample_Image_2017-07-26_353_x.hdf5	Sample_Line_2018-06-13_069_x.hdf5	
Sample_Image_2018-06-13_005_x.hdf5	Sample_Line_2018-06-13_078_x.hdf5	
Sample_Image_2018-06-13_006_x.hdf5	Sample_Line_2018-09-15_115_x.hdf5	
Sample_Image_2018-06-13_007_x.hdf5	Sample_Line_2018-09-15_157_x.hdf5	
Sample_Image_2018-06-13_011_x.hdf5	Sample_Line_2018-10-01_134_x.hdf5	
Sample_Image_2018-06-13_027_x.hdf5	Sample_Line_2019-04-02_211_x.hdf5	
Sample_Image_2018-06-13_054_x.hdf5	Sample_Line_2019-04-02_218_x.hdf5	
Sample_Image_2018-06-13_071_x.hdf5	Sample_Line_2019-04-02_219_x.hdf5	
Sample_Image_2018-06-13_077_x.hdf5	Sample_Stack_2019-02-27_196_x.hdf5	
Sample_Image_2018-09-15_147_x.hdf5	Sample_Stack_2019-02-28_016_x.hdf5	

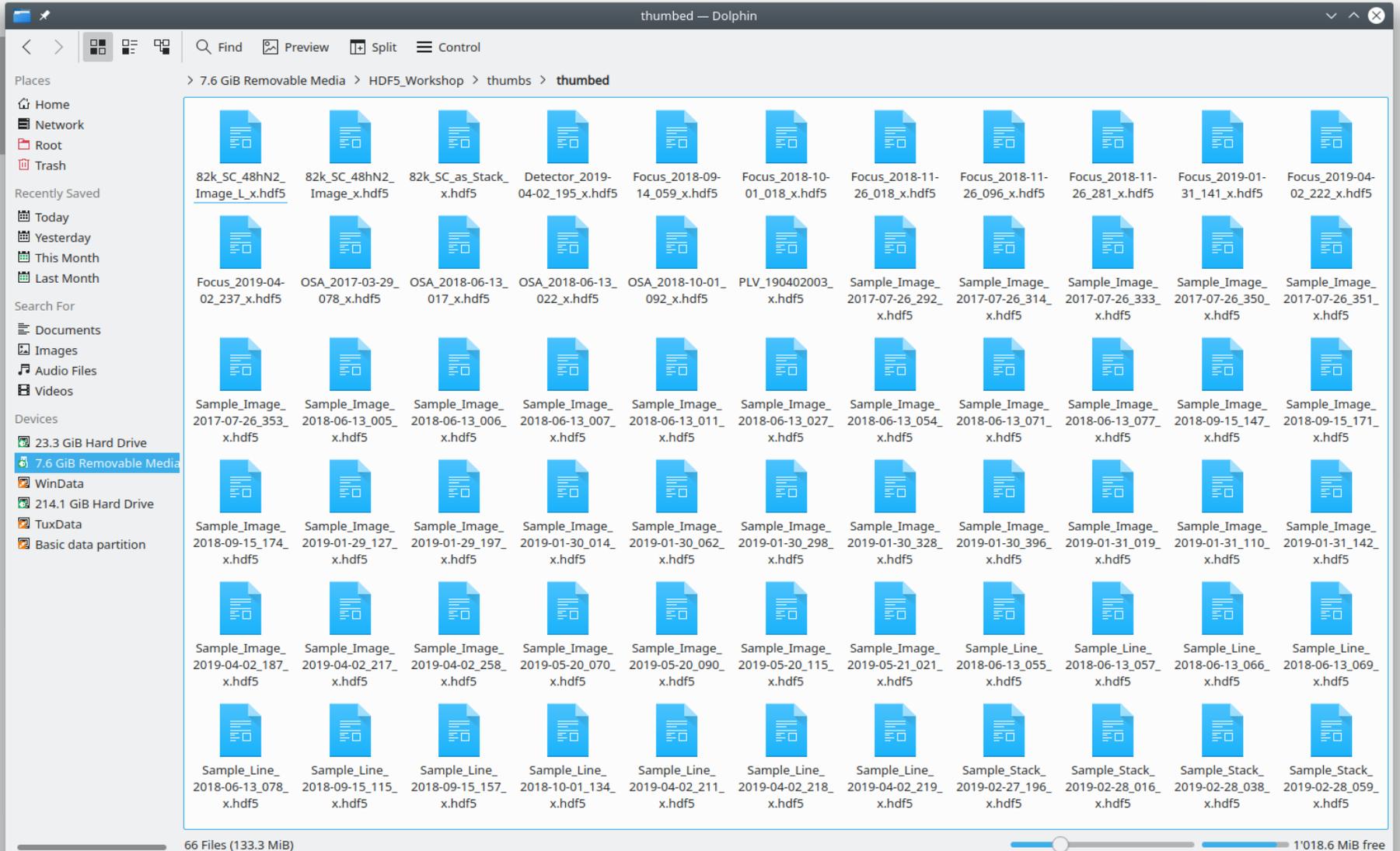
Thumbnail Previews (Win10)

The screenshot shows a Windows File Explorer window with the address bar set to 'Flask > Documents > thumbed'. The left sidebar shows the navigation pane with 'Documents' selected. The main area displays a grid of 66 thumbnail previews for HDF5 files. The thumbnails are arranged in a grid and include the following file names:

- 82k_SC_48h_N2_Image_L_x.hdf5
- 82k_SC_48h_N2_Image_x.hdf5
- 82k_SC_as_Stack_x.hdf5
- Detector_2019-04-02_195_x.hdf5
- Focus_2018-09-14_059_x.hdf5
- Focus_2018-10-01_018_x.hdf5
- Focus_2018-11-26_018_x.hdf5
- Focus_2018-11-26_096_x.hdf5
- Focus_2018-11-26_281_x.hdf5
- Focus_2019-01-31_141_x.hdf5
- Focus_2019-04-02_222_x.hdf5
- Focus_2019-04-02_237_x.hdf5
- OSA_2017-03-29_078_x.hdf5
- OSA_2018-06-13_017_x.hdf5
- OSA_2018-06-13_022_x.hdf5
- OSA_2018-10-01_092_x.hdf5
- PLV_190402003_x.hdf5
- Sample_Im age_2017-07-26_292_x.hdf5
- Sample_Im age_2017-07-26_314_x.hdf5
- Sample_Im age_2017-07-26_333_x.hdf5
- Sample_Im age_2017-07-26_350_x.hdf5
- Sample_Im age_2017-07-26_351_x.hdf5
- Sample_Im age_2017-07-26_353_x.hdf5
- Sample_Im age_2018-06-13_005_x.hdf5
- Sample_Im age_2018-06-13_006_x.hdf5
- Sample_Im age_2018-06-13_007_x.hdf5
- Sample_Im age_2018-06-13_011_x.hdf5
- Sample_Im age_2018-06-13_027_x.hdf5
- Sample_Im age_2018-06-13_054_x.hdf5
- Sample_Im age_2018-06-13_071_x.hdf5
- Sample_Im age_2018-06-13_077_x.hdf5
- Sample_Im age_2018-09-15_147_x.hdf5
- Sample_Im age_2018-09-15_171_x.hdf5
- Sample_Im age_2018-09-15_174_x.hdf5
- Sample_Im age_2019-01-29_127_x.hdf5
- Sample_Im age_2019-01-29_197_x.hdf5
- Sample_Im age_2019-01-30_014_x.hdf5
- Sample_Im age_2019-01-30_062_x.hdf5
- Sample_Im age_2019-01-30_298_x.hdf5
- Sample_Im age_2019-01-30_328_x.hdf5
- Sample_Im age_2019-01-30_396_x.hdf5
- Sample_Im age_2019-01-31_019_x.hdf5
- Sample_Im age_2019-01-31_110_x.hdf5
- Sample_Im age_2019-01-31_142_x.hdf5
- Sample_Im age_2019-04-02_187_x.hdf5
- Sample_Im age_2019-04-02_217_x.hdf5
- Sample_Im age_2019-04-02_258_x.hdf5
- Sample_Im age_2019-05-20_070_x.hdf5
- Sample_Im age_2019-05-20_090_x.hdf5
- Sample_Im age_2019-05-20_115_x.hdf5
- Sample_Im age_2019-05-21_021_x.hdf5
- Sample_Lin e_2018-06-13_055_x.hdf5
- Sample_Lin e_2018-06-13_057_x.hdf5
- Sample_Lin e_2018-06-13_066_x.hdf5
- Sample_Lin e_2018-06-13_069_x.hdf5
- Sample_Lin e_2018-06-13_078_x.hdf5
- Sample_Lin e_2018-09-15_115_x.hdf5
- Sample_Lin e_2018-09-15_157_x.hdf5
- Sample_Lin e_2018-10-01_134_x.hdf5
- Sample_Lin e_2019-04-02_211_x.hdf5
- Sample_Lin e_2019-04-02_218_x.hdf5
- Sample_Lin e_2019-04-02_219_x.hdf5
- Sample_Sta ck_2019-02-27_196_x.hdf5
- Sample_Sta ck_2019-02-28_016_x.hdf5
- Sample_Sta ck_2019-02-28_038_x.hdf5
- Sample_Sta ck_2019-02-28_059_x.hdf5

66 items

Boring List of Data Files (KDE4)



Thumbnail Previews (KDE4)

The screenshot shows the KDE4 Dolphin file manager window titled "thumbed — Dolphin". The interface includes a sidebar on the left with "Places" (Home, Network, Root, Trash), "Recently Saved" (Today, Yesterday, This Month, Last Month), "Search For" (Documents, Images, Audio Files, Videos), and "Devices" (23.3 GiB Hard Drive, 7.6 GiB Removable Media, WinData, 214.1 GiB Hard Drive, TuxData, Basic data partition). The main pane displays a grid of 66 thumbnail previews of image files, each with a filename below it. The files are organized into rows and columns, showing a variety of image types including scientific data, patterns, and abstract shapes. The status bar at the bottom indicates "66 Files (133.3 MiB)" and a progress bar showing "1'019.1 MiB free".

Thumbnail	Filename
	82k_SC_48hN2_Image_L_x.hdf5
	82k_SC_48hN2_Image_x.hdf5
	82k_SC_as_Stack_x.hdf5
	Detector_2019-04-02_195_x.hdf5
	Focus_2018-09-14_059_x.hdf5
	Focus_2018-10-01_018_x.hdf5
	Focus_2018-11-26_018_x.hdf5
	Focus_2018-11-26_096_x.hdf5
	Focus_2018-11-26_281_x.hdf5
	Focus_2019-01-31_141_x.hdf5
	Focus_2019-04-02_222_x.hdf5
	Focus_2019-04-02_237_x.hdf5
	OSA_2017-03-29_078_x.hdf5
	OSA_2018-06-13_017_x.hdf5
	OSA_2018-06-13_022_x.hdf5
	OSA_2018-10-01_092_x.hdf5
	PLV_190402003_x.hdf5
	Sample_Image_2017-07-26_292_x.hdf5
	Sample_Image_2017-07-26_314_x.hdf5
	Sample_Image_2017-07-26_333_x.hdf5
	Sample_Image_2017-07-26_350_x.hdf5
	Sample_Image_2017-07-26_351_x.hdf5
	Sample_Image_2017-07-26_353_x.hdf5
	Sample_Image_2018-06-13_005_x.hdf5
	Sample_Image_2018-06-13_006_x.hdf5
	Sample_Image_2018-06-13_007_x.hdf5
	Sample_Image_2018-06-13_011_x.hdf5
	Sample_Image_2018-06-13_027_x.hdf5
	Sample_Image_2018-06-13_054_x.hdf5
	Sample_Image_2018-06-13_071_x.hdf5
	Sample_Image_2018-06-13_077_x.hdf5
	Sample_Image_2018-09-15_147_x.hdf5
	Sample_Image_2018-09-15_171_x.hdf5
	Sample_Image_2018-09-15_174_x.hdf5
	Sample_Image_2019-01-29_127_x.hdf5
	Sample_Image_2019-01-29_197_x.hdf5
	Sample_Image_2019-01-30_014_x.hdf5
	Sample_Image_2019-01-30_062_x.hdf5
	Sample_Image_2019-01-30_298_x.hdf5
	Sample_Image_2019-01-30_328_x.hdf5
	Sample_Image_2019-01-30_396_x.hdf5
	Sample_Image_2019-01-31_019_x.hdf5
	Sample_Image_2019-01-31_110_x.hdf5
	Sample_Image_2019-01-31_142_x.hdf5
	Sample_Image_2019-04-02_187_x.hdf5
	Sample_Image_2019-04-02_217_x.hdf5
	Sample_Image_2019-04-02_258_x.hdf5
	Sample_Image_2019-04-02_070_x.hdf5
	Sample_Image_2019-05-20_090_x.hdf5
	Sample_Image_2019-05-20_115_x.hdf5
	Sample_Image_2019-05-21_021_x.hdf5
	Sample_Line_2018-06-13_055_x.hdf5
	Sample_Line_2018-06-13_057_x.hdf5
	Sample_Line_2018-06-13_066_x.hdf5
	Sample_Line_2018-06-13_069_x.hdf5
	Sample_Line_2018-09-15_078_x.hdf5
	Sample_Line_2018-09-15_115_x.hdf5
	Sample_Line_2018-09-15_157_x.hdf5
	Sample_Line_2018-10-01_134_x.hdf5
	Sample_Line_2019-04-02_211_x.hdf5
	Sample_Line_2019-04-02_218_x.hdf5
	Sample_Line_2019-04-02_219_x.hdf5

XMP – Extensible Metadata Platform



Adding Intelligence to Media

- 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

Thumbnails in File Browsers

- 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

Decimal:	137	72	77	80	13	10	26	10
Hexadecimal:	89	48	4d	50	0d	0a	1a	0a
ASCII:	\211	H	M	P	\r	\n	\032	\n

```

<?xpacket begin="" id="w5M0MpCehiHzreSzNTczkc9d"?>
<x:xmpmeta xmlns:x='adobe:ns:meta/' x:xmptk='XMP toolkit 3.0-29, framework 1.6'>
  <rdf:RDF xmlns:rdf='http://www.w3.org/1999/02/22-rdf-syntax-ns#'
    xmlns:iX='http://ns.adobe.com/iX/1.0/'>
    <rdf:Description rdf:about=''
      xmlns:xap='http://ns.adobe.com/xap/1.0/'
      xmlns:xapGImg='http://ns.adobe.com/xap/1.0/g/img/'>
      <xap:CustomKey1>value 1</xap:CustomKey1>
      <xap:CustomKey2>value 2</xap:CustomKey2>
      <xap:CustomKey3>value 3</xap:CustomKey3>
      <xap:Thumbnails>
        <rdf:Alt>
          <rdf:li rdf:parseType='Resource'>
            <xapGImg:format>PNG</xapGImg:format>
            <xapGImg:image>IMAGE_DATA_BASE64</xapGImg:image>
          </rdf:li>
        </rdf:Alt>
      </xap:Thumbnails>
    </rdf:Description>
  </rdf:RDF>
</x:xmpmeta>
<?xpacket end="w"?>

```

← Metadata key:value pairs

Thumbnail data

A standard set of metadata key:value pairs for scientific data is still needed



<https://github.com/nexusformat/hdf5xmp>

-Code for thumbnailers

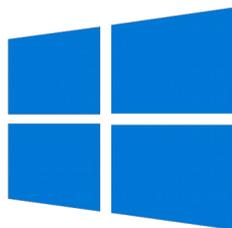
-Python code to insert/extract metadata and thumbnails

<https://github.com/nexusformat/hdf5xmp/releases>

-Binaries for thumbnailers

Support for:

- Windows 7, 8.1, 10, macOS, Gnome and KDE4

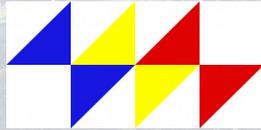


Windows 10



Many thanks to

programming
Marc Schädeli
SEMAFOR AG



funding
Swiss Light Source

