Metadata Extractors in Rosetta

Is One for Each Format Enough?
What’s in a namespace?

And also, in the conversion of the default [stream] and [format] sections to XML, I’d like to see an explicit namespace set, e.g. xmlns:ffprobe=http://natlib.govt.nz/blah/NDHA/FFPROBEConversion/ would translate the XML elements to <ffprobe:size> and <ffprobe:profile> etc. because with some of these values there is a likelihood of collision – droid might output size, ffprobe outputs size - which we should avoid – but also, for elements like ‘profile’ it adds clearer contextual information and additional semantic information.
For example ...

Exiftool Extract

to extract</FlashPix:Thumbna11Clip>
<FlashPix:CodePage>Windows Latin 1 (Western European)
</FlashPix:CodePage>
<FlashPix:PresentationTarget>on-screen Show
</FlashPix:PresentationTarget>
<FlashPix:Company>National Library of New Zealand
</FlashPix:Company>
<FlashPix:Bytes>21347814</FlashPix:Bytes>
<FlashPix:Paragraphs>72</FlashPix:Paragraphs>
<FlashPix:Slides>26</FlashPix:Slides>
<FlashPix:Notes>2</FlashPix:Notes>
<FlashPix:HiddenSlides>0</FlashPix:HiddenSlides>
<FlashPix:MMClips>7</FlashPix:MMClips>
<FlashPix:ScaleCrop>No</FlashPix:ScaleCrop>
<FlashPix:LinksUpToDate>No</FlashPix:LinksUpToDate>
<FlashPix:SharedDoc>No</FlashPix:SharedDoc>
<FlashPix:HyperlinksChanged>No</FlashPix:HyperlinksChanged>
<FlashPix:TitleOfParts>
< rdf:Bag>
< rdf:li>Arial</rdf:li>
< rdf:li>Times New Roman</rdf:li>
< rdf:li>Default Design</rdf:li>
< rdf:li>Microsoft PowerPoint Presentation</rdf:li>
< rdf:li>Adobe Photoshop Image</rdf:li>
< rdf:li>Slide 1</rdf:li>
< rdf:li>Slide 2</rdf:li>
< rdf:li>Slide 3</rdf:li>
< rdf:li>Slide 4</rdf:li>
< rdf:li>Slide 5</rdf:li>
< rdf:li>Five-Element Framework</rdf:li>
< rdf:li>National Digital Heritage Archive</rdf:li>
< rdf:li>Slide 6</rdf:li>
< rdf:li>Matt Shen TCV President and Executive Chief Officer Ex Libris Group, Israel</rdf:li>
< rdf:li>Slide 10</rdf:li>
</ rdf:Bag>
</ FlashPix:TitleOfParts>

Rosetta Extract

< key id="significantPropertiesValue">Slide 1
</ key>
</ record>
< record>
< key id="significantPropertiesType">ms.version
</ key>
< key id="significantPropertiesValue">Microsoft PowerPoint
</ key>
</ record>
< record>
< key id="significantPropertiesType">ms.FINumber
</ key>
< key id="significantPropertiesValue">0x3c8</key>
</ record>
< record>
< key id="significantPropertiesType">ms.magicNumber</key>
< key id="significantPropertiesValue">0xf</key>
</ record>
< record>
< key id="significantPropertiesType">ms.productVersion</key>
< key id="significantPropertiesValue">0x1f6c</key>
</ record>
< record>
< key id="significantPropertiesType">ms.modifiedDate</key>
< key id="significantPropertiesValue">2015-09-18 09:19:01</key>
</ record>
< record>
< key id="significantPropertiesType">ms.creatingApplicationName</key>
< key id="significantPropertiesValue">Microsoft Office PowerPoint</key>
</ record>
And with wav files...

FFprobe is more specific:

```
streams.stream.0.index=0
streams.stream.0.codec_name="pcm_s24le"
streams.stream.0.codec_long_name="PCM signed 24-bit little-endian"
streams.stream.0.profile="unknown"
streams.stream.0.codec="audio"
streams.stream.0.codec_tag_string="[1][0][0][0]"
streams.stream.0.codec_tag="0x0001"
streams.stream.0.sample_fmt="S32"
streams.stream.0.sample_rate="96000"
streams.stream.0.channels="1"
streams.stream.0.channel_layout="unknown"
streams.stream.0.bit_rate="2304000"
```

And EXIFTOOL extracts our XMP data:

```
<XMP-Archives_N2:Format_Extension>wav</XMP-Archives_N2:Format_Extension>
<XMP-Archives_N2:Playback_Device>Nagra 4.2</XMP-Archives_N2:Playback_Device>
<XMP-Archives_N2:Equalisation>CCIR</XMP-Archives_N2:Equalisation>
<XMP-Archives_N2:AD_Converter>Apogee Symphony I/O</XMP-Archives_N2:AD_Converter>
<XMP-Archives_N2:Equalizer>None</XMP-Archives_N2:Equalizer>
<XMP-Archives_N2:Word_Clock>Avid Sync HD</XMP-Archives_N2:Word_Clock>
<XMP-Archives_N2:Host_Computer_Manufacturer>Apple</XMP-Archives_N2:Host_Computer_Manufacturer>
<XMP-Archives_N2:Host_Computer_Name>iMac</XMP-Archives_N2:Host_Computer_Name>
<XMP-Archives_N2:Operating_System>OSX</XMP-Archives_N2:Operating_System>
<XMP-Archives_N2:Operating_System_Version>10.8.5</XMP-Archives_N2:Operating_System_Version>
<XMP-Archives_N2:Capture_Software>Adobe Audition</XMP-Archives_N2:Capture_Software>
<XMP-Archives_N2:Capture_Software_Version>CS6</XMP-Archives_N2:Capture_Software_Version>
<XMP-Archives_N2:Technician>Rosie Rowe</XMP-Archives_N2:Technician>
<XMP-Archives_N2:Technical_Notes>Clicks and pops wow and flutter in original recording</XMP-Archives_N2:Technical_Notes>
```
Hon Peter Dunne was born in Christchurch in 1954.

He graduated from the University of Canterbury in 1977 with a Master of Arts degree with Honours in Political Science, and has also studied business administration at Massey University. He was President of the University of Canterbury Students Association in 1975.

Hon Peter Dunne Leader, United Future New Zealand MP for Ohariu-Belmont

Hon Peter Dunne was born in Christchurch in 1954.

He graduated from the University of Canterbury in 1977 with a Master of Arts degree with Honours in Political Science, and has also studied business administration at Massey University. He was President of the University of Canterbury Students Association in 1975.
Thank you!

https://public.etherpad-mozilla.org/p/mde

andrea.byrne@dia.govt.nz