Format Registry Ontology
Rev. 2003/Mar/10 – SLA

The format registry maintains persistent, unambiguous bindings between public identifiers for digital representation formats and representation information for those formats. A format is a representation of an information model, a formal expression of exchangeable knowledge. Thus, the format registry is a typing mechanism for encoded digital content.

This specification defines two fundamental classes of formats:

1. Content stream formats
2. Physical media formats

A content stream format is a fixed byte-serialized encoding of an information model, independent of the physical medium underlying its manifestation. (The registry makes no assumption regarding the size of a byte, i.e., a byte is not necessarily an octet.) A content stream is not required to have homogeneous composition, but can be defined as an aggregation of other content streams.

A physical media format is a fixed encoding of a content stream in a tangible form on a physical storage structure.

Format representation information enables the mapping of typed formats into more meaningful concepts by capturing the significant syntactic and semantic properties of formats, with particular relevance towards the operational needs of digital repositories, including, but not limited to: object format identification, characterization, ingest validation, interchange, migration, emulation, and other archival preservation activities.

Within the registry, formats are classified functionally according to an ontology that hierarchically decomposes the format space based upon the fundamental nature of the data units representable by the formats. The top-level organization of the hierarchy is similar to the MIME media type classification.

0 Information model [ expression of exchangeable knowledge ]
1 Content stream [ byte-serialized encoding of information model ]
1.1 Logical [ data representing truth and falsity ]
1.1.1 Integer - XDR integer (RFC 1832)
1.1.1.1 Unsigned integer - XDR unsigned integer (RFC 1832)
1.2 Numeric [ data representing mathematical ordinality or cardinality ]
1.2.1 Scalar
1.2.1.1 Integer - XDR integer (RFC 1832)
1.2.1.2 Real
1.2.1.2.1 Floating point - IEEE 754
1.2.1.3 Complex
1.3 Text [ character data directly interpretable by humans ]
1.3.1 Structured text [ text with structural constraints ]
1.3.1.1 Mark-up language
1.3.1.1.1 HTML
1.3.1.1.2 ISO 8879:1986 (SGML)
1.3.1.1.2.1 LaTeX
1.3.1.1.2.2 RTF
### 1.3.1.1.1 Definition
- SGML DTD
- XML DTD
- XSD → XML

### 1.3.1.1.2 Transformation
- XSLT → XML

### 1.3.1.1.3 Presentation
- CSS
- XSL-FO → XML

### 1.3.1.2 Programming language

#### 1.3.1.2.1 Functional
- Lisp

#### 1.3.1.2.2 Declarative

##### 1.3.1.2.2.1 Interpreted
- AppleScript
- Perl
- sh

##### 1.3.1.2.2.2 Compiled
- C++
- C#
- Fortran
- Java

### 1.3.1.3 Message

#### 1.3.1.3.1 Mail
- MIME (RFC 2045)

#### 1.3.1.3.2 News
- USENET (RFC 1036)

### 1.4 Image [visual data requiring rendering technology for human interpretability]

#### 1.4.1 Still

##### 1.4.1.1 Font [character glyph data]

#### 1.4.1.1.1 Outline
- Adobe Type 1
- OpenType
- TrueType

##### 1.4.1.1.2 Raster
- SVG → XML

#### 1.4.1.2 Graphic

##### 1.4.1.2.1 Vector
- 2D
- 3D
- VRML

##### 1.4.1.2.2 Raster [rectilinear array of picture elements]
- GIF
- JFIF
- PCD
- TIFF

#### 1.4.1.3 Page description
- PDF
- PostScript
- Quark XPress

### 1.4.2 Motion
- AVI
- MPEG
- QuickTime

### 1.5 Audio [aural data requiring playback technology for human interpretability]
- AIFF
- MP3
- Real
- WAV
1.5.1 Music
- MIDI

1.6 Application [arbitrary data requiring technological mediation for human interpretability]
1.6.1 Communication
1.6.2 Database
  1.6.2.1 Hierarchical
  1.6.2.2 Relational
    1.6.2.2.1 Schema
      - SQL DDL
    1.6.2.2.2 Query
      - SQL DML
    1.6.2.2.3 Data
      - MySQL
      - Oracle
      - Postgres

1.6.3 Executable
- ELF
- EXE
- Java byte code

1.6.4 Presentation
- PowerPoint

1.6.5 Spreadsheet
- Excel

1.6.6 Word processing
- Word
- WordPerfect

1.7 Transformation [composable encodings applied against primary formats]
1.7.1 Compression [size minimizing encodings]
  1.7.1.1 Lossless
    - CCITT T.4
    - Lempel-Ziv
      - compress
      - gzip
      - LZ77
      - LZW
  1.7.1.2 Lossy
    - JPEG

1.7.2 Container [aggregations of atomic information model units]
- jar
- PKZIP
- StuffIt
- tar

1.7.3 Transfer [data encodings for transmission over a communication channel]
  1.7.3.1 7-bit safe
    - Base64
    - BinHex
    - uuencode

2 Physical media [tangible encoding of content stream on physical storage structure]
2.1 Magnetic
  2.1.1 Disk
  2.1.2 Tape
    2.1.2.1 Reel
      2.1.2.1.1 9 track
        - ANSI X3.54-1986
  2.1.2.2 Cartridge
    2.1.2.2.1 3480 class
      - ANSI X3.180-1990
  2.1.2.2.2 DLT
    - ISO/IEC 16382:2000

2003/Mar/10 – SLA
2.2 Optical
2.2.1 Disk
2.2.1.1 CD-ROM
- ISO 9660:1988
2.2.1.2 DVD
2.2.2 Film
2.3 Paper
2.3.1 Card
- Hollerith
2.3.2 Tape

1.6.3 Domain [domain-specific data]
1.6.3.1 Agriculture
1.6.3.2 Commerce
1.6.3.2.1 EDI
- ASC X12
- UN/CEFACT
1.6.3.2.2 Finance
1.6.3.3 Education
1.6.3.4 Engineering
- ISO 10303 (STEP)
1.6.3.4.1 CAD
- DXF
- IGES
1.6.3.5 Humanities
1.6.3.6 Law
1.6.3.7 Medicine
1.6.3.7.1 Pharmaceutical
1.6.3.8 Science
- HDF
- netCDF
1.6.3.8.1 Information
1.6.3.8.1.1 Archive
1.6.3.8.1.2 Computer
1.6.3.8.1.3 Library
1.6.3.8.1.3.1 Catalog
- MARC
1.6.3.8.2 Physical
1.6.3.8.2.1 Astronomy
- FITS
1.6.3.8.2.2 Biology
- PDB
1.6.3.8.2.2.1 Genetics
- GenBank
1.6.3.8.2.3.2 Zoology
1.6.3.8.2.3 Chemistry
1.6.3.8.2.4 Earth Science
1.6.3.8.2.4.1 Geography
1.6.3.8.2.4.1.1 GIS
- GeoTIFF → TIFF
- SDTS
1.6.3.8.2.4.2 Geology
1.6.3.8.2.4.3 Meteorology
1.6.3.8.2.4.4 Oceanography
1.6.3.8.2.5 Mathematics
- Mathematica
1.6.3.8.2.5.1 Statistics
- SAS
- SPSS
<table>
<thead>
<tr>
<th>Code</th>
<th>Subject</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.6.3.8.2.6</td>
<td>Physics</td>
</tr>
<tr>
<td>1.6.3.8.3</td>
<td>Social</td>
</tr>
<tr>
<td>1.6.3.8.3.1</td>
<td>Anthropology</td>
</tr>
<tr>
<td>1.6.3.8.3.2</td>
<td>Economics</td>
</tr>
<tr>
<td>1.6.3.8.3.3</td>
<td>Political Science</td>
</tr>
<tr>
<td>1.6.3.8.3.4</td>
<td>Psychology</td>
</tr>
<tr>
<td>1.6.3.8.3.5</td>
<td>Sociology</td>
</tr>
</tbody>
</table>