# Digital File Formats for Preservation: Decoded

With so many file formats out there, how can we tell which to use to ensure the long-term preservation of our materials?

Fortunately, the Library of Congress provides guidance with their yearly *Recommended Formats Statement*. Evaluation is based on 7 "sustainability factors:"

- *Disclosure*
- *Adoption*
- *Transparency*
- *Self-documentation*
- *External dependencies*
- *Impact of patents*
- *Technical protection mechanisms*

Based on these criteria, the Library recommends "preferred" and "acceptable" file formats for a variety of digital resources. This Venn diagram shows some of the most popular file formats and their usefulness for various resource types.

## Key

Underlined — Resource Type

**Bold** — Recommended Format

Standard — Acceptable Format

### Resource type by color
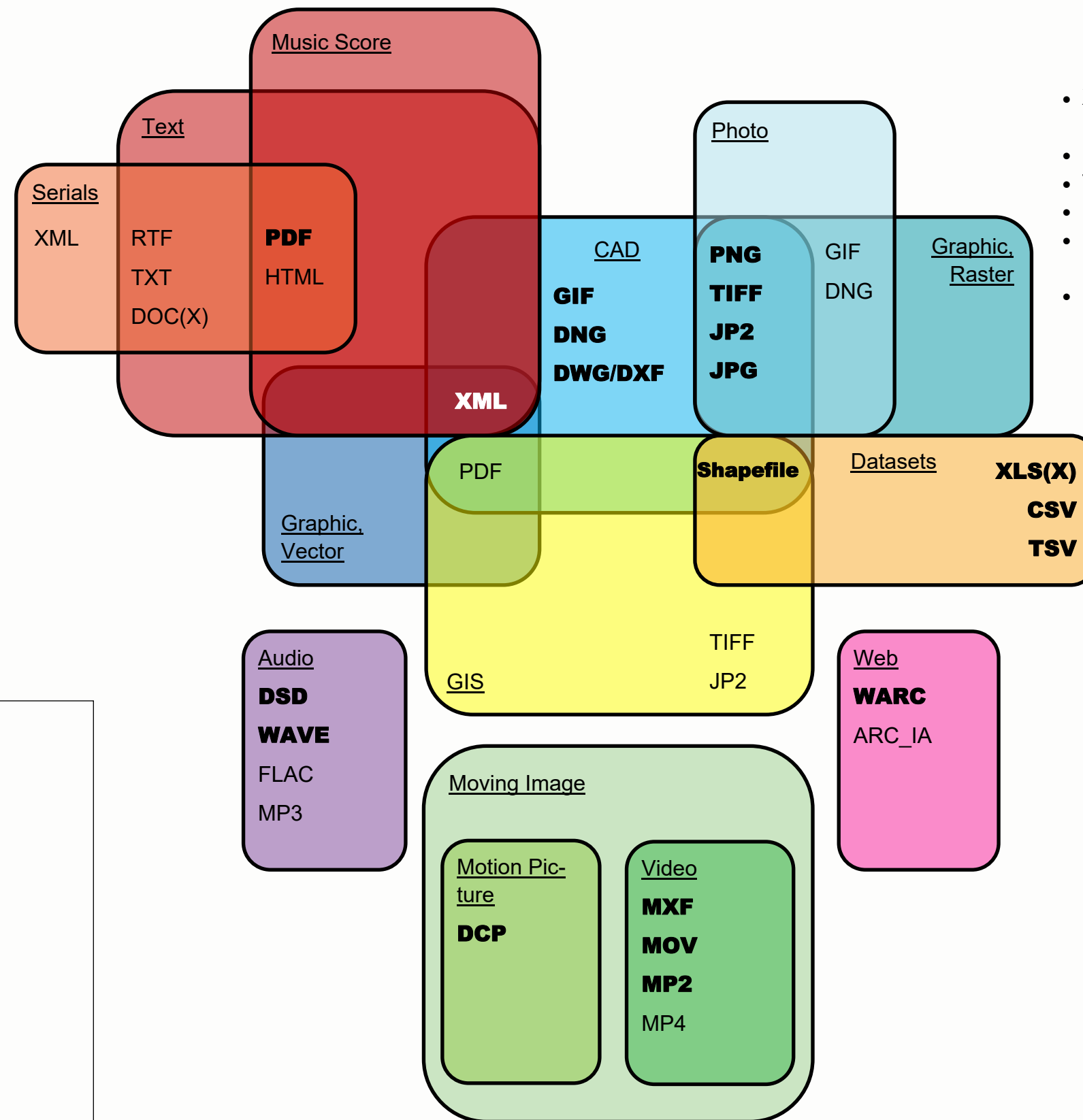
**Reds** — Textual Resources

**Orange** — Dataset Resources

**Yellow** — Geospatial Resources

**Greens** — Moving Image Resources

**Blues** — Still Image Resources

**Purple** — Audio Resources

**Pink** — Web Resources

---

### Music Score

### Text

#### Serials
XML

RTF

TXT

DOC(X)

**PDF**

HTML

**XML**

### CAD

**GIF**

**DNG**

**DWG/DXF**

### Photo

**PNG**

**TIFF**

**JP2**

**JPG**

GIF

DNG

### Graphic, Raster

### Graphic, Vector

PDF

### Shapefile

**Shapefile**

### Datasets

**XLS(X)**

**CSV**

**TSV**

### GIS

TIFF

JP2

### Audio

**DSD**

**WAVE**

FLAC

MP3

### Web

**WARC**

ARC_IA

### Moving Image

#### Motion Picture

**DCP**

#### Video

**MXF**

**MOV**

**MP2**

MP4

---

## File Extensions

### Text-based Formats

- XML: Extensible Markup Language
- RTF: Rich Text Format
- TXT: Plain text
- DOC(X): Microsoft Word
- PDF: Portable Document Format
- HTML: HyperText Markup Language

### Still Image-based Formats

- GIF: Graphics Interchange Format
- DNG: Adobe Digital Negative
- DWG/DXF: AutoCAD Formats
- PNG: Portable Network Graphics
- TIFF: Tagged Image File Format
- JPG: JPEG Encoded Images
- JP2: JPEG 2000

### Geospatial Formats

- Shapefile: Format developed by Esri (formerly the Environmental Systems Research Institute, Inc.)

### Dataset Formats

- XLS(X): Microsoft Excel
- CSV: Comma Separated Value
- TSV: Tab Separated Value

### Sound-based Formats

- DSD: Direct Stream Digital
- WAVE: Waveform Audio File Format (.wav)
- FLAC: Free Lossless Audio Codec (lossless)
- MP3: MPEG Layer 3 Audio Encoding (lossy)

### Web-based Formats

- WARC: Web ARchive File Format
- ARC_IA: Internet Archive File Format

### Moving Image-based Formats

- DCP: Digital Cinema Package
- MXF: Material Exchange Format
- MOV: QuickTime File Format
- MP2: MPEG-2 Encoding
- MP4: MPEG-4 Encoding

Source: https://www.loc.gov/preservation/resources/rfs/index.html