EMERGING FORMATS

E-RECORDS DAY 2021

CAD

Computer Aided Design (CAD) is a generic term for digital design software, hardware, and files that support complex virtual modelling. CAD technology is applied in the fields of engineering, architecture, video game design and animation, archaeology, and in many other disciplines. AutoCAD files are CAD files generated from the popular, proprietary software application AutoCAD, which is owned and developed by the company Autodesk. CAD files use mathematical calculations to produce vector graphics and contain data in multiple, interrelated layers.

COMMON FORMATS

Drawing Interchange Format (.dwg): a common, proprietary format produced from AutoCAD

Digital Interchange Format (.dxf): a format owned by AutoCAD, but interoperable with other CAD software and developed for data exchange

ISO 10303-21 Standard for the Exchange of Product Model Data (STEP) file: a plain text file defined by the STEP standard that contains technical and structural metadata to facilitate transferring and representing CAD models between computer systems

ISO/IEC 19775-1 Extensible 3D (X3D): an open standard format group "intended to be a universal interchange format for integrated 3D graphics and multimedia"

RISKS

- Software versioning
- Proprietary formats
- Accessibility and display
- Linkages between multiple files for a single design

GOOGLE WORKSPACE

Google Workspace (formerly G Suite) is a group of cloud-based collaboration products that include Google docs, sheets, and slides. Data created and shared in docs, sheets, and slides are rendered in a browser when a user interacts with that Google service, and do not exist as discrete files that can be exported in their original born-cloud formats. Instead, Google born-cloud formats must be converted and exported to another format within the Google environment, such as docx, ODT, PDF, or HTML. .gsheet or .gdoc files do not contain actual data and are instead shortcuts containing links to open the files from within Google Drive using a web browser.

RISKS

- Proprietary formats
- Versioning of documents and retention of comments and suggestions
- Ambiguity of creator, owner, or record holder
- Data loss during export, such as loss of comments or charts
- Loss of metadata, such as owner and creation dates, during export
- Anonymization or de-anonymization of user data depending on export format

PDF/A

PDF/A is an ISO-standardized Portable Document Format (PDF) used for the long-term preservation of archival documents. The PDF family of documents is developed by Adobe Systems and is widely adapted. The specifications are open and available free of charge. Although the file extension .pdf is consistent with the family of products, you must save a PDF file specifically in the PDF/A format to benefit from its features. A number of versions of PDF/A are available for varied applications. The most recent version PDF/A-4 includes PDF/A-4f for PDFs with embedded files and PDF/A4e for engineering files.

BENEFITS

- Open format
- Widely adapted
- Easy "Save As" options from common document types
- Small file size
- Bulk conversions possible

POTENTIAL LIMITATIONS

- Javascript not enabled
- Executables are not enabled
- Fonts must be embedded
- Encryption is not allowed
- LZW compression
- Multimedia not allowed

WARC

The WARC (Web ARChive) format specifies a method for combining multiple digital resources into an aggregate archival file together with related information. Resources are dated, identified by URIs, and preceded by simple text headers. By convention, files of this format are named with the extension ".warc" and have the MIME type application/warc. The WARC file format is a revision and generalization of the ARC format used by the Internet Archive to store information blocks harvested by web crawlers. A WARC record consists of a record header followed by a record content block and two newlines; the header has mandatory named fields that document the date, type, and length of the record and support the convenient retrieval of each harvested resource (file).

BENEFITS

- Open standard, maintained by International Internet Preservation Consortium
- Publicly documented
- Efficient for bulk harvesting
- Compression support to reduce file size impact
- Archive format (includes multiple file types within)

POTENTIAL LIMITATIONS

- WARC is an archive format, files not readable by common end-user software
- No guarantee of contained files usability
- No guarantee of contained files accessibility (during crawl)
- Requires indexing for user access

CoSA Council of State Archivists

SERI STATE ELECTRONIC RECORDS INITIATIVE

www.statearchivists.org

info@statearchivists.org