PREFORMA Glossary

ANSI/NISO Z39.87-2006 (R2011) Data Dictionary

Standard created by the National Information Standard Organization that defines a set of metadata elements for raster digital images to enable users to develop, exchange, and interpret digital image files. The dictionary has been designed to facilitate interoperability between systems, services, and software as well as to support the long-term management of and continuing access to digital image collections.

Archival Information Package (AIP)

An Information Package, consisting of the Content Information and the associated Preservation Description Information (PDI), which is preserved within an OAIS.

Archive

An organization that intends to preserve information for later access and use by a Designated Community.

Assertion

Generally an Assertion is a boolean expression, i.e. it may be evaluated to true or false, used in software testing. Evaluating the Assertion involves examining some property of the item under test.

Associated Description

The information describing the content of an Information Package.

Baseline TIFF 6.0

The essentials TIFF specifications that all mainstream TIFF developers should support in their products.

Byte order

The order of the bytes that compose a digital word in computer memory. It also describes the order of byte transmission over a digital link. Words may be represented in big-endian or little-endian format.

Byte Sequence

A binary data stream read from a source, for example:

- an input stream from a file on storage device accessed through a file system;

- an input stream read from a particular URL; or

- an input stream read directly from memory.

A particular byte sequence is identified by combining its length in bytes and the SHA-1 Hash (derived from its contents).

CELLAR

Codec Encoding for LossLess Archiving and Realtime transmission, a working group formed within the Internet Engineering Task Force to standardize Matroska, FFV1, and FLAC media formats.

Challenge Brief

Document which sets forth the overall challenge for long term preservation of digital files to be addressed by the PREFORMA Research & Development actions, i.e. two strategies that empower memory institutions to gain control over the technical properties of preservation files: development of an open-source conformance checker, and establishing an ecosystem around an open source 'reference' implementation.

Codec

A codec is a device or computer program for encoding or decoding a digital data stream or signal.

Conformance checker

A PREFORMA term that defines a generalised (i.e. not concerned with a particular file format) open source validation toolset for conformance checking of digital files, intended for long-term preservation in memory institutions.

The toolset is composed of the following modules: Shell, Implementation checker, Policy checker, Reporter and Metadata fixer and its main functionalities are:

- to validate whether a file has been produced according to the specifications of a standard file format;

- to validate whether a file matches the acceptance criteria for long-term preservation by the memory institution;

- to report in human and machine readable format those properties that deviate from the standard specification and acceptance criteria;

- to perform automated fixes for simple deviations in the metadata of the preservation file.

Conformance checker reports

Reports created during examinations. These reports may provide feedback to the Suppliers during the Prototyping files and may be helpful investigating errors and other messages from the conformance checkers. Such use should however be considered as a type of "training", and consequently, an examination file whose report has been shared with a Supplier can no longer be used as an evaluation file.

PREFORMA - Future Memory Standards
PREservation FORMAts for culture information/e-archives
EC Grant agreement no: 619568

Consumer

The role played by those people or client systems, who interact with OAIS services to find preserved information of interest and to access that information in detail. This can include other OAISes, as well as internal OAIS persons or systems.

Container (format)

A container or wrapper format is a metafile format whose specification describes how different elements of data and metadata coexist in a computer file. See also: Wrapper (format).

Content

In the context of PREFORMA, the information within and functionality of a fixated and delimited information resource, e.g., text using the PDF functionality "transparency" in a PDF/A file with metadata in the XMP format.

Content Data Object

The Data Object, that together with associated Representation Information, comprises the Content Information.

Content Information

A set of information that is the original target of preservation or that includes part or all of that information. It is an Information Object composed of its Content Data Object and its Representation Information.

Context Information

The information that documents the relationships of the Content Information to its environment. This includes why the Content Information was created and how it relates to other Content Information objects.

Data

A re-interpretable representation of information in a formalized manner suitable for communication, interpretation, or processing. Examples of data include a sequence of bits, a table of numbers, the characters on a page, the recording of sounds made by a person speaking, or a moon rock specimen.

Data Submission Session

A delivery of media or a single telecommunications session that provides Data to an OAIS. The Data Submission Session format/contents is based on a data model negotiated between the OAIS and the Producer in the Submission Agreement. This data model identifies the logical constructs used by the Producer and how they are represented on each media delivery or in the telecommunication session.

Descriptive Information

The set of information, consisting primarily of Package Descriptions, which is provided to Data Management to support the finding, ordering, and retrieving of OAIS information holdings by Consumers.

Design phase

First phase of the PREFORMA Pre-Commercial Procurement. It is intended to demonstrate the feasibility of the proposed concepts for new solutions.

Designated Community

An identified group of potential Consumers who should be able to understand a particular set of information. The Designated Community may be composed of multiple user communities. A Designated Community is defined by the Archive and this definition may change over time.

Designated PREFORMA dispatcher

A PREFORMA member subject who has been designated to: receive files from providers, and storing them in the PREFORMA vault, administrate, organize and structure the PREFORMA vault, and dispatching training files from the PREFORMA vault to the PREFORMA repository.

Digital curation

Digital curation is the selection, preservation, maintenance, collection and archiving of digital assets. Digital curation establishes, maintains and adds value to repositories of digital data for present and future use. This is often accomplished by archivists, librarians, scientists, historians, and scholars.

Digital Migration

The transfer of digital information, while intending to preserve it, within the OAIS. It is distinguished from transfers in general by three attributes:

- a focus on the preservation of the full information content that needs preservation;

- a perspective that the new archival implementation of the information is a replacement for the old; and

- an understanding that full control and responsibility over all aspects of the transfer resides with the OAIS.

Digital Object

An object composed of a set of bit sequences.

Digital preservation

Digital preservation is defined by the DigitalPreservationEurope project as "a set of activities required to make sure digital objects can be located, rendered, used and understood in the future".

Dissemination files

Files used for presentation purposes by the PREFORMA Consortium and by the suppliers during the outreach phase or later by third parties for miscellaneous purposes, such as, assessing the PREFORMA Conformance checkers.

Dissemination Information Package (DIP)

An Information Package, derived from one or more AIPs, and sent by Archives to the Consumer in response to a request to the OAIS.

DPF Manager

Digital Preservation Framework Manager (DPF Manager) is an extensible, open source software project consisting of an implementation checker, policy checker, reporter, and fixer that targets preservation-level still image files (specifically TIFF files)

Dublin Core (DC)

The Dublin Core Schema is a small set of vocabulary terms that can be used to describe web resources (video, images, web pages, etc.), as well as physical resources such as books or CDs, and objects like artworks.[1] The full set of Dublin Core metadata terms can be found on the Dublin Core Metadata Initiative (DCMI) website.

EBML

Extensible Binary Meta Language (EBML) is a binary XML format. EBML supports the Matroska file format specification.

Embedded Resource

A Byte Sequence embedded into the PDF Document such as an image, font, colour profile, or attachment.

Embedded Resource Parser

A third-party tool which can parse and analyse Embedded Resources, for example a JPEG2000 validator or font validator.

Embedded Resource Report

A Machine-readable Report produced by an Embedded Resource Parser containing information about an Embedded Resource.

PREFORMA - Future Memory Standards
PREservation FORMATs for culture information/e-archives
EC Grant agreement no: 619568

Evaluation files

Files used for testing by the PREFORMA Consortium during the evaluation phase. These files must be different from those that have been used as training files during the prototyping phase, but they need to belong to the same category (i.e. they cannot introduce new failures, challenges and issues that were not included in the training set).

Examination files

Files used for testing by the PREFORMA members/network during the prototyping phase. The tests can be done locally, "in-house", or within any other environment. The files can be provided later either for training or for evaluation.

Exchangeable image file format (EXIF)

Exchangeable image file format is a standard that specifies the formats for images, sound, and ancillary tags used by digital cameras (including smartphones), scanners and other systems handling image and sound files recorded by digital cameras. The specification uses the following existing file formats with the addition of specific metadata tags: JPEG discrete cosine transform (DCT) for compressed image files, TIFF Rev. 6.0 (RGB or YCbCr) for uncompressed image files, and RIFF WAV for audio files (Linear PCM or ITU-T G.711 μ-Law PCM for uncompressed audio data, and IMA-ADPCM for compressed audio data). It is not used in JPEG 2000, PNG, or GIF.

Extended TIFF 6.0

TIFF extensions to baseline TIFF not recommended for general data interchange. Files that use such features shall be designated "Extended TIFF 6.0" files, and the particular extensions used should be documented. A Baseline TIFF 6.0 reader is not required to support any extensions.

Extensible Markup Language (XML)

Free open standards markup language that defines a set of rules for encoding documents in a format that is both human-readable and machine-readable.

Extensible Metadata Platform (XMP)

ISO standard, originally created by Adobe Systems Inc., for the creation, processing and interchange of standardized and custom metadata for digital documents and data sets. XMP standardizes a data model, a serialization format and core properties for the definition and processing of extensible metadata. It also provides guidelines for embedding XMP information into popular image, video and document file formats, such as JPEG and PDF, without breaking their readability by applications that do not support XMP.

FFmpeg

FFmpeg is a free software project that produces libraries and programs for handling multimedia data.

PREFORMA - Future Memory Standards
PREservation FORMAts for culture information/e-archives
EC Grant agreement no: 619568

FFV1

FFV1 (FF Video Codec 1) is a lossless intra-frame video codec.

File

In the context of PREFORMA, computer file storing an information resource in a format, or claims to be in a format, of the type PDF/A, TIFF/A, Matroska, which in turn can contain one or more information resources.

File provider

PREFORMA member, PREFORMA network member, or third parties who provides files to the designated PREFORMA dispatcher.

Ground truth

The ground truth determines uniquely which classes a document/file belongs to.

Human-readable Report

A report generated from a Machine-readable Report in a format suitable for human interpretation (eg. HTML or PDF) and containing messages translated according to a Language Pack.

ICC Profile

In color management, an ICC profile is a set of data that characterizes a color input or output device, or a color space, according to standards promulgated by the International Color Consortium (ICC). Profiles describe the color attributes of a particular device or viewing requirement by defining a mapping between the device source or target color space and a profile connection space (PCS).

Image File Directory (IFD)

Data structure used to store an image data information and metadata inside a TIFF file.

Image File Header (IFH)

Data structure, located in the first 8 bytes on a TIFF file, used to identify the TIFF file defines the byte order and the offset to the first Image File Directory.

Implementation Check

The execution of a discrete Validation Test for a particular file.

Implementation checker

A PREFORMA term for the component which "performs a comprehensive check of the standard specifications listed in the standard document."

PREFORMA - Future Memory Standards
PREservation FORMATs for culture information/e-archives
EC Grant agreement no: 619568

This module verifies whether a file has been produced according to the specifications of a standard file format.

Independently Understandable

A characteristic of information that is sufficiently complete to allow it to be interpreted, understood and used by the Designated Community without having to resort to special resources not widely available, including named individuals.

Information

Any type of knowledge that can be exchanged. In an exchange, it is represented by data. An example is a string of bits (the data) accompanied by a description of how to interpret the string of bits as numbers representing temperature observations measured in degrees Celsius (the Representation Information).

Information Interchange Model (IIM)

The Information Interchange Model (IIM) is a file structure and set of metadata attributes that can be applied to text, images and other media types.

It was developed in the early 1990s by the International Press Telecommunications Council (IPTC) to expedite the international exchange of news among newspapers and news agencies. Although IIM was intended for use with all types of news items a subset found broad worldwide acceptance as the standard embedded metadata used by news and commercial photographers. Information such as the name of the photographer, copyright information and the caption or other description can be embedded either manually or automatically. The full IIM specification includes a complex data structure and a set of metadata definitions

Information Package

A logical container composed of optional Content Information and optional associated Preservation Description Information. Associated with this Information Package is Packaging Information used to delimit and identify the Content Information and Package Description information used to facilitate searches for the Content Information.

Information Property

That part of the Content Information as described by the Information Property Description. The detailed expression, or value, of that part of the information content is conveyed by the appropriate parts of the Content Data Object and its Representation Information.

Information resource

Information structures, metadata, profiles, wrappers and other formats for example text, image, audio, audiovisual, such as fonts, encoding, ICC, XMP, JPEG, PNG, TIFF, WAVE, MP3, Matroska, PDF/A.

PREFORMA - Future Memory Standards
PREservation FORMAts for culture information/e-archives
EC Grant agreement no: 619568

ISO Working Group (WG)

An ISO committee managing one or more ISO standards. For example, ISO TC 171 SC 2 WG 5 owns ISO 19005 (PDF/A) while WG 8 owns ISO 32000 (the PDF specification).

Knowledge Base

A set of information, incorporated by a person or system, that allows that person or system to understand received information.

Known sources

Files which can be traced back to their original format (if converted) and/or the software that generated the file.

Language Pack

A file or set of files which specify all string constants for a given language as well as additional localisation (such as date format).

Long Term

A period of time long enough for there to be concern about the impacts of changing technologies, including support for new media and data formats, and of a changing Designated Community, on the information being held in an OAIS. This period extends into the indefinite future.

Long Term Preservation

The act of maintaining information, Independently Understandable by a Designated Community, and with evidence supporting its Authenticity, over the Long Term.

Lossless compression

Lossless compression is an algorithm that represent data without losing any information but minimizes file size.

LPCM

LPCM is a linearly uniform method used to digitally represent sampled analog signals.

Machine-readable Report

A structured report, independent of language and localization, generated for automated processing rather than human readability.

Management

PREFORMA - Future Memory Standards
PREservation FORMAts for culture information/e-archives
EC Grant agreement no: 619568

The role played by those who set overall OAIS policy as one component in a broader policy domain, for example as part of a larger organization.

Matroska

Matroska is an open standard, free container format, a file format that can hold an unlimited number of video, audio, picture, or subtitle tracks in one file.

MediaConch

MediaConch is an extensible, open source software project consisting of an implementation checker, policy checker, reporter, and fixer that targets preservation-level audiovisual files (specifically Matroska, Linear Pulse Code Modulation (LPCM) and FF Video Codec 1 (FFV1))

MediaInfo

MediaInfo is an open source program that displays technical information about media files, as well as tag information for many audio and video files.

MediaTrace

MediaTrace is a technical report that expresses the binary architecture of a file.

Memory Institution

An organization maintaining a repository of public knowledge, a generic term used about institutions such as libraries, archives, museums, sites and monument records, clearinghouses, providers of Digital Libraries and data aggregation services which serve as memories for given societies or mankind.

Metadata

Data that provides information about other data.

Metadata Encoding and Transmission Standard (METS)

Metadata standard for encoding descriptive, administrative, and structural metadata regarding objects within a digital library, expressed using the XML schema language of the World Wide Web Consortium (W3C). The standard is maintained as part of the MARC standards of the Library of Congress, and is being developed as an initiative of the Digital Library Federation (DLF). Depending on its use, a METS document could be used in the role of Submission Information Package (SIP), Archival Information Package (AIP), or Dissemination Information Package (DIP) within the Open Archival Information System (OAIS) Reference Model.

Metadata Fix

A simple alignment of the file's internal metadata with the actual validity status (as determined by the Conformance Checker) of the file.

Metadata fixer

A PREFORMA term for the component which allows for "simple corrections to metadata embedded in otherwise-conforming or non-conforming files."

This module performs automated fixes for simple deviations in the metadata of the preservation file, leaving the original bitstream untouched and created a correct copy of the object to be preserved.

Metadata Fixing Report

A Machine-readable Report generated by the Metadata Fixer containing details of Metadata Fixes carried out and any exceptions.

National Information Standards Organization (NISO)

United States non-profit standards organization that develops, maintains and publishes technical standards related to publishing, bibliographic and library applications. It was founded in 1939, incorporated as a not-for-profit education association in 1983, and assumed its current name in 1984.

NISO MIX

XML schema created by The Library of Congress' Network Development and MARC Standards Office, in partnership with the NISO Technical Metadata for Digital Still Images Standards Committee and other interested experts, for describing a set of technical data elements required to manage digital image collections. The schema provides a format for interchange and/or storage of the data specified in the Data Dictionary – Technical Metadata for Digital Still Images (ANSI/NISO Z39.87-2006). This schema is currently referred to as "NISO Metadata for Images in XML (NISO MIX)". MIX is expressed using the XML schema language of the World Wide Web Consortium. MIX is maintained for NISO by the Network Development and MARC Standards Office of the Library of Congress with input from users.

OAIS Reference Framework

ISO Standard 14721:2012. Reference Model for an Open Archival Information System. The conformance check developed in PREFORMA enables implementation of the following OAIS functions:

- Quality assurance at Ingest, validating (QA results) the successful transfer of the SIP to the temporary storage area.

- Generate AIP at Ingest, transforming one or more SIPs into one or more AIPs that conform to the Archive's data formatting standards and documentation standards.

- Archival Information Update at Ingest, providing a mechanism for updating (repackaging, transformation) the contents of the Archive.

- Monitor Designated Communities for Preservation Planning, interacting with Archive Consumers and Producers to track changes in their service requirements and available product technologies.

- Develop Preservation Strategies and Standards for preservation planning, developing and recommending strategies and standards, and for assessing risks, to enable the Archive to make informed tradeoffs as it establishes standards, sets policies, and manages its system infrastructure.

- Establishing Standards and Policies by the Administration of the Archive system and maintain them.

Open Archival Information System (OAIS)

An Archive, consisting of an organization, which may be part of a larger organization, of people and systems, that has accepted the responsibility to preserve information and make it available for a Designated Community. It meets a set of responsibilities that allows an OAIS Archive to be distinguished from other uses of the term 'Archive'. The term 'Open' in OAIS is used to imply that this Recommendation and future related Recommendations and standards are developed in open forums, and it does not imply that access to the Archive is unrestricted.

Open source license

License enabling that anyone that has adopted the software has the right to freely read, use, improve and redistribute the source code for such software. In the context of PREFORMA, all software developed in the project must be licensed under the two open source licenses "GPL v3 or later" and "MPL v2 or later" and all digital assets produced in the project must be released under the open access license Creative Commons CC-BY v4.0 and in open file formats, i.e. an open standard as defined in the European Interoperability Framework for Pan-European eGovernment Service (version 1.0 2004).

Open source project

In the context of PREFORMA, a project that develop an open source licensed conformance checker following the requirements specified in the Challenge Brief

Open Source Software (OSS)

Open Source Software (OSS) is software which as a whole is licensed by one or more licenses approved by the Open Source Initiative, see http://www.opensource.org/licenses/.

Based on the definition of open source software from Kammarkollegiet, the Swedish organisation responsible for all public procurement(Dnr. 96-34-2014, in the document "Allmänna villkor").

Organic files

PREFORMA - Future Memory Standards
PREservation FORMAts for culture information/e-archives
EC Grant agreement no: 619568

Non-normative implementation. Data actually managed by memory institutions for their preservation duties.

Package Description

The information intended for use by Access Aids.

Packaging Information

The information that is used to bind and identify the components of an Information Package. For example, it may be the ISO 9660 volume and directory information used on a CD-ROM to provide the content of several files containing Content Information and Preservation Description Information.

PDF

The Portable Document Format (PDF) is a file format used to present documents in a manner independent of application software, hardware, and operating systems.[

PDF Document

A Byte Sequence claiming conformance with ISO 32000-1:2008 (for PDF/A-2 and PDF/A-3) or to the Adobe specification of PDF 1.4 (for PDF/A-1).

PDF Document Extract

A programmatic model of a PDF Document created by parsing a PDF. The model encapsulates applicable PDF syntax, any PDF Metadata, plus details of PDF Features. The model can be serialised as a Machine-readable Report.

PDF Feature

Any property of the PDF Document or any of its structural elements such as pages, images, fonts, color spaces, annotations, attachments, etc.

PDF Features Report

A Machine-readable Report containing details about PDF Features including PDF Metadata and other available XMP metadata packages.

PDF Metadata

PDF document-level metadata stream containing the XMP package and the entries of the PDF Info dictionary.

PDF Parser

A software component that reads a PDF Document and constructs a PDF Document Extract.

PREFORMA - Future Memory Standards
PREservation FORMAts for culture information/e-archives
EC Grant agreement no: 619568

PDF Validation TWG (TWG)

The PDF Association Technical Working Group attended by industry members to discuss and decide matters pertaining to PDF Validation.

PDF/A

PDF/A is an ISO-standardized version of the Portable Document Format (PDF) specialized for use in the archiving and long-term preservation of electronic documents.

PDF/A Document

A Byte Sequence claiming conformance to a specific PDF/A Flavour.

PDF/A Flavor

PDF/A Part+Level.

Possible PDF/A Flavors are: 1a, 1b, 2a, 2b, 2u, 3a, 3b, 3u.

PDF/A Identification

The part of the XMP metadata in a PDF Document that identifies the PDF/A Part (1, 2 or 3) and conformance Level (b, a, or u) to which the PDF Document claims to conform.

PDF/A Level

Level a, b, or u conformance as defined by the PDF/A Part.

PDF/A Part

Part 1: ISO 19005-1:2005

Part 2: ISO 19005-2:2011

Part 3: ISO 19005-3:2012

PDF/A Validation

The process of testing whether the PDF Features of a PDF Document conform to the requirements for a particular PDF/A Flavor. The PDF/A Validation process generates a Validation Report.

PDF/A Validation Report

A Machine-readable Report containing the results (all errors and notifications) of PDF/A Validation.

Policy

PREFORMA - Future Memory Standards
PREservation FORMAts for culture information/e-archives
EC Grant agreement no: 619568

Institutional acceptance criteria for long-term archiving and preservation of files and metadata, including requirements beyond those specified in ISO or other 3rd party standards.

Policy Check

The execution of a discrete Policy Test for a particular file. The Policy Checker carries out Policy Tests when enforcing a particular Policy Profile.

Policy checker

A PREFORMA term for the component "which allows for adding acceptance criteria, always compliant with the standard specifications, that further differentiates the properties of the file. This might, for example, include limiting conformance to PDF/A-1b, or exclude files containing a certain type of image."

This module verifies whether a file matches the acceptance criteria for long-term preservation by the memory institution.

Policy Profile

A file that expresses institutional Policy as a set of formal Policy Tests.

Policy Profile Registry

Enables the discovery and exchange of Policy Profiles between institutions.

Policy Report

A Machine-readable Report containing the results of Policy Checks performed as defined in a Policy Profile.

Policy Test

A Test Assertion that is evaluated by examining a Feature Report to ensure a file complies with institutional Policy.

Pre-Commercial Procurement

Competition-like procurement method, which enables public sector bodies to engage with innovative businesses and other interested parties in development projects to arrive at innovative solutions that address specific public sector challenges and needs. The new innovative solutions are created through a phased procurement of development contracts to reduce risk.

PREFORMA Shell

A PREFORMA term for an interactive component: "which allows a conformance checker to interface with other systems, allowing for interfacing multiple conformance checkers at the same time".

PREFORMA - Future Memory Standards
PREservation FORMAts for culture information/e-archives
EC Grant agreement no: 619568

Aim of this module is to ensure interoperability, enabling the integration of different conformance checkers into one single application.

Preservation Description Information (PDI)

The information which is necessary for adequate preservation of the Content Information and which can be categorized as Provenance, Reference, Fixity, Context, and Access Rights Information.

Producer

The role played by those persons or client systems that provide the information to be preserved. This can include other OAISes or internal OAIS persons or systems.

Prototyping phase

Second phase of the PREFORMA Pre-Commercial Procurement. It is intended to develop prototypes from the more promising concepts delivered by the selected suppliers in the design phase.

Reference implementation

Implementation of a standard specification that is to be used as a definitive interpretation for that standard specification. Methodology or objective frame of reference to interpret and implement the standard specifications against the background of the current variations of interpretations and implementations by software vendors, aimed to determine whether a file is what it claims to be, i.e., in this context, what makes a file a valid file that conforms to the specifications of a standard file format.

Report Template

A template file that defines the layout and format of a Human-readable or Machine-readable report. These are used by the Reporter to transform Machine-readable Reports into alternative formats.

Reporter

A PREFORMA term for the component that "interprets the output of the implementation checker and policy checker and allows for defining multiple human and machine readable output formats. This might include a well-documented JSON or XML file, a human readable report on which specifications are not fulfilled, or a fool-proof report which also indicates what should be done to fix the errors."

This module reports in human and machine readable format which properties deviate from the standard specification and acceptance criteria.

Repository

In the context of PREFORMA, the official logical location where the training and demonstration files are stored and from where the Suppliers and third parties can access and use the files.

PREFORMA - Future Memory Standards
PREservation FORMAts for culture information/e-archives
EC Grant agreement no: 619568

Representation Information

The information that maps a Data Object into more meaningful concepts. An example of Representation Information for a bit sequence which is a FITS file might consist of the FITS standard which defines the format plus a dictionary which defines the meaning in the file of keywords which are not part of the standard. Another example is JPEG software which is used to render a JPEG file; rendering the JPEG file as bits is not very meaningful to humans but the software, which embodies an understanding of the JPEG standard, maps the bits into pixels which can then be rendered as an image for human viewing.

Semantic Information

The Representation Information that further describes the meaning beyond that provided by the Structure Information.

SHA-1 Hash

A cryptographic hash function that creates a digital fingerprint for a byte sequence, referred to as 'message' in cryptographic documentation.

Standard file format

File that has been produced according to the specifications of a standard file format, In the context of PREFORMA, the following standard specifications have been selected for inclusing in the Challenge Brief.

For electronic documents:

- ISO (2005). Document management — Electronic document file format for long-term preservation — Part 1: Use of PDF 1.4 (PDF/A-1).

- ISO/TC 171/SC 2, ISO 19005-1:2005. ISO (2008). Document management — Portable document format — Part 1: PDF 1.7.

- ISO/TC 171/SC 2, ISO 32000-1:2008. ISO (2011). Document management — Electronic document file format for long-term preservation — Part 2: Use of ISO 32000-1 (PDF/A-2).

- ISO/TC 171/SC 2, ISO 19005-2:2011. ISO (2012). Document management — Electronic document file format for long-term preservation — Part 3: Use of ISO 32000-1 with support for embedded files (PDF/A-3).

- ISO/TC 171/SC 2, ISO 19005-3:2012.

For images:

- ISO (2001). Electronic still-picture imaging — Removable memory — Part 2: TIFF/EP image data format.

- ISO/TC 42, ISO 12234-2:2001. ISO (2004). Graphic Technology — Prepress digital data exchange — Tag image file format for image technology (TIFF/IT). ISO/TC 130. ISO 12369:2004.

For audiovisual files:

- MKV: Matroska – Technical Details. http://www.matroska.org/technical/index.html.

- OGG: Ogg – Documentation. https://xiph.org/ogg/doc/.

- JPEG2000: ISO (2004). Information technology – JPEG 2000 image coding system: Core coding system. ISO/IEC JTC 1/SC 29, ISO/IEC 15444-1:2004.

- FFV1: FFV1 Video Codec Specification, http://www.ffmpeg.org/~michael/ffv1.html.

- Dirac: Dirac Specification Version 2.2.3 (2008), http://diracvideo.org/download/specification/dirac-spec-latest.pdf.

- LPCM: IEC (2014). Digital audio interface – Part 1: General. IEC/TC 100, IEC 60958-1 ed3.1 Consol. with am1: 2014.

Structure Information

The Representation Information that imparts meaning about how other information is organized. For example, it maps bit streams to common computer types such as characters, numbers, and pixels and aggregations of those types such as character strings and arrays.

Submission Agreement

The agreement reached between an OAIS and the Producer that specifies a data model, and any other arrangements needed, for the Data Submission Session. This data model identifies format/contents and the logical constructs used by the Producer and how they are represented on each media delivery or in a telecommunication session.

Submission Information Package (SIP)

An Information Package that is delivered by the Producer to the OAIS for use in the construction or update of one or more AIPs and/or the associated Descriptive Information.

Synthetic files

Normative implementation. Data created with the specific purpose of pinpointing some specific compliance problem or critical issue for a given preservation format. Synthetic files are synthetic in the sense that they are purposefully created to test a specific functionality of the validator.

Tagged Image for Archival (TI/A)

PREFORMA - Future Memory Standards
PREservation FORMAts for culture information/e-archives
EC Grant agreement no: 619568

ISO Recommendation to optimize the TIFF file format definition for archival purposes. The TI/A recommendation define a subset of standard TIFF tags which are either required, optional of forbidden for the purposes of long term archival. Within this context, the goal must be that; The image can be opened with standard software even in the far future and the image data does not contain features that are not documented and therefore cannot be understood and rendered correctly in the future. Conforming to the TI/A recommendation will guarantee that the essential digital information of a image file always can be read and interpreted correctly. Since TI/A is a subset of the TIFF standard, all current TIFF readers are able to correctly and completely render TI/A just out-of-the-box.

Test Assertion

A Test Assertion is a testable or measurable expression for evaluating the adherence of an implementation (or part of it) to a normative statement in a specification.

Testing phase

Third phase of the PREFORMA Pre-Commercial Procurement. It is intended to test the applications developed during the protototyping phase by the memory institutions of the consortium.

TIFF

Tagged Image File Format, abbreviated TIFF or TIF, is a computer file format for storing raster graphics images

Training files

Files used for testing by the suppliers during the prototyping phase. These must be distinguished from the evaluation files.

Transformation

A Digital Migration in which there is an alteration to the Content Information or PDI of an Archival Information Package. For example, changing ASCII codes to UNICODE in a text document being preserved is a Transformation.

Unknown sources

Files which are not traceable to a specific origin, that is, it is not known how the files were created.

Validation Class

Label assigned to documents/files according to their characteristics.

In general classes may intersect but the correct class must be separate.

Validation Model

A formal definition of all PDF Document objects, their properties, and relationships between them expressed in a custom syntax.

Validation Profile

A structured file describing the set of Validation Tests to be performed during Validation for a particular PDF/A Flavour.

Validation Test

A Test Assertion that is evaluated by examining a PDF Features Report to ensure a PDF Document complies with a requirement expressed in a specific PDF/A Flavour.

Vault

In the context of PREFORMA, the logical and physical storage location for all files submitted by a Provider to a designated PREFORMA dispatcher. The PREFORMA vault is only accessible to a designated PREFORMA dispatcher.

veraPDF API

The veraPDF application programming interface defines the operations, inputs, outputs, and types that form the protocols for interacting programmatically with the veraPDF Library.

veraPDF Command Line Interface

The command line executable(s) providing access to the veraPDF Library API and functionality via a command line interface.

veraPDF Configuration

The detailed settings which configure an invocation of the veraPDF Conformance Checker. Configuration settings are logically divided into:

- task config: settings controlling the behaviour of a component, these are reusable across executions and installations;

- installation config: settings unique to a particular installation, such as home and temp directories;

- execution config: settings unique to a particular invocation, such as files or URLs to check, names of output report files.

veraPDF conformance checker

veraPDF conformance checker is an extensible, open source software project consisting of an implementation checker, policy checker, reporter, and fixer that targets preservation-level electronic documents (specifically PDF/A files)

veraPDF Desktop Graphical User Interface (GUI-D)

An executable program that provides access to the veraPDF API and Library on a desktop computer or workstation.

veraPDF Framework

A software library that provides a lightweight framework based on open standards for use by Conformance Checker developers.

veraPDF Library

The software library that provides the functionality and APIs for PDF/A Validation, Policy Checking, Metadata Fixing, and Reporting.

veraPDF REST API

RESTful web service API that provides HTTP access to the veraPDF Library functionality. This is a REST layer on top of the veraPDF API.

veraPDF Shell

Provides the user interfaces to manage and operate the Conformance Checker, handling issues such as workflow control and scheduling.

veraPDF Web Graphical User Interface (GUI-W)

Browser based HTML user interface that calls the veraPDF Library through the REST API, which in turn calls the veraPDF API.

Wrapper (format)

A container or wrapper format is a metafile format whose specification describes how different elements of data and metadata coexist in a computer file. See also: Container (format).

XML Schema

Description of a type of XML document, typically expressed in terms of constraints on the structure and content of documents of that type, above and beyond the basic syntactical constraints imposed by XML itself.