Allied Data Publication 34

(ADatP-34(J))

NATO Interoperability Standards and Profiles

Volume 3

Profile for the Long Term Preservation of NATO Digital Information of Permanent Value

DECEMBER 2016

NCI Agency

Table of Contents

This page is intentionally left blank

List of Figures

This page is intentionally left blank

1. PROFILE FOR THE LONG TERM PRESERVATION OF NATO DIGITAL INFORMATION OF PERMANENT VALUE

001. Information of permanent value shall be submitted by the NATO Information Managers in their role as Information Custodians to the NATO Archivist in one of the approved sustainable archival formats and packaged in this appendix.

002. The submission process for information of permanent value for long-term preservation is shown in Figure 1.1.

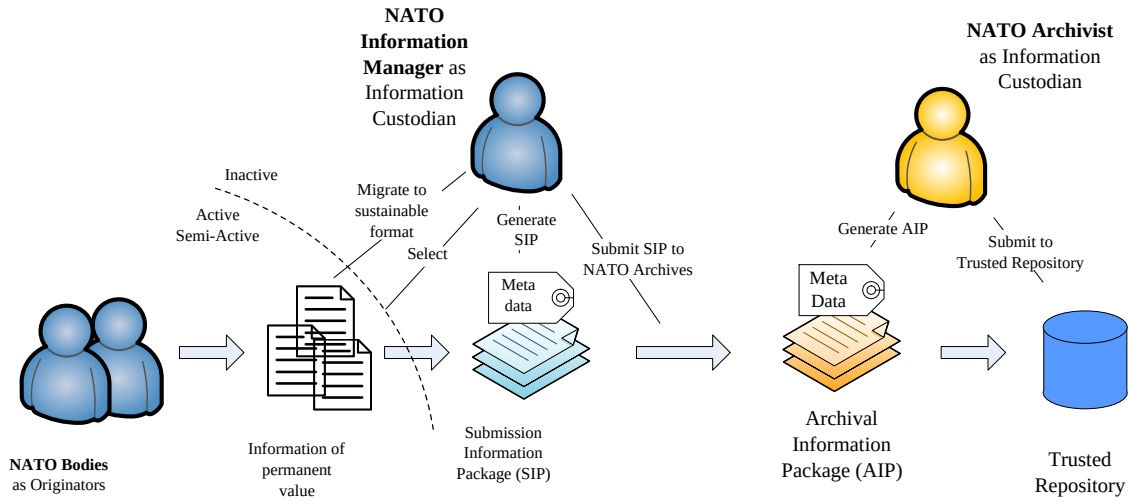

Figure 1.1. Long-term preservation

003. This profile outlines the file formats (Section 1.1) and package structures (Section 1.2) approved by the Archives Committee for the long-term preservation of NATO digital information of permanent value.

004. NATO information custodians shall provide information in these formats and structures to the NATO Archivist.

005. Further guidance on best practice will be issued in the near future. The contents of this profile shall become part of Volume 3 of the NATO Interoperability Standards and Profiles [4].

1.1. FILE FORMATS FOR LONG TERM PRESERVATION

006. The following sustainable file formats are approved by the Archives Committee for the long term preservation of NATO digital information of permanent value. The formats are ordered by content type. A brief characterization of the generic requirements for the preservation of content is included.

1.1.1. Data sets

007. Data sets are typically collections of individual values or larger coherent structures such as messages. The data set might be an export from a database or the results of an information exchange between systems.

008. There is typically a structure associated with the data set, either implicitly contained within the data set (e.g. a table structure of an Excel document or a database), or explicitly defined (e.g. as a schema definition)

Service	Standard	Implementation Guidance
Data sets (e.g. scientific data) and any structured information not fitting other content types	*Mandatory* • IETF RFC 4180 - Common Format and MIME Type for Comma-Separated Values (CSV) Files • W3C REC-xml11-20060816 - Extensible Markup Language (XML) version 1.1 (Second Edition) • W3C REC-xmlschema11-1-20120405 - XML Schema Definition Language (XSD) 1.1 Part 1: Structures • W3C REC-xmlschema11-2-20120405 - XML Schema Definition Language (XSD) 1.1 Part 2: Datatypes	**Requirements** • Preserve structured and unstructured data for future analysis • Preserve logical structure of dataset as well as syntax and semantics of elements within the dataset • Preserve data types and data structures
Database content	*Mandatory* • ISO/IEC 9075-1 - Database languages - SQL - Part 1: Framework	

1.1.2. Text

009. Documents consisting primarily of textual descriptions are the most prevalent and important category of information of permanent value in the NATO context. Text documents might also include embedded diagrams, pictures, or other non- text material. These items shall not be separated from the text and kept as part of the document.

Service	Standard	Implementation Guidance

Service	Standard	Implementation Guidance
Text documents, including common MS Office document formats (docx, xlsx, pptx)	*Mandatory* • ISO 32000-1 - Document management -- Portable document format -- Part 1: PDF 1.7	Use conformance level : PDF/A-2a **Requirements** • Preserve integrity of text, diagram and figures, pagination and navigation (formatting) • Preserve document metadata • Inclusion of fonts, layout information, and indices
Email (e.g. MS Outlook PST files)	*Mandatory* • IETF RFC 4155 - The application/mbox Media Type	**Requirements** • Preserve email content including attachments • Preserve complete mailboxes. Important messages might be exported and preserved as individual text documents.
Chat (e.g. JChat conversations)	*Mandatory* • ISO 32000-1 - Document management -- Portable document format -- Part 1: PDF 1.7 • IETF RFC 4155 - The application/mbox Media Type	Use conformance level : PDF/A-2a **Requirements** • Preserve message content, including attachments • Preserve complete dialogs per user or multi-user chat room with time-stamps. • Preserve information about users and user groups

1.1.3. Still Images

010. Still images are visual representations, including photographs, graphs, and diagrams. Still images can be divided into two main types, bitmap (or raster) images and vector images. Bitmap images are typically photographs produced by scanners and cameras at a fixed resolution, while vector images consist of scalable objects. Both types can be combined, e.g. in course of action diagrams where a bitmap image of an area can have symbology vector overlays.

Service	Standard	Implementation Guidance
Bitmap/raster images	*Mandatory* • ISO/IEC 15444-1 - JPEG 2000 image coding system: Core coding system • ISO/IEC 10918-1 - Digital compression and coding of continuous-tone still images: Requirements and guidelines • ADOBE tiff - TIFF Revision 6.0	**Requirements** • Preserve resolution (clarity, colors), scalability, and ability of render the image • Preserve image metadata • Compressibility, preference for lossless compression • Preference for larger resolution
Vector images	*Mandatory* • W3C REC-SVG11-20110816 - Scalable Vector Graphics (SVG) 1.1 Specification (Second Edition)	

1.1.4. Moving Images

011. Moving images are digital recordings of still images at a particular frame rate and resolution. A compression is often applied by only capturing the difference between adjacent frames. Moving images are typically combined with audio data and packaged into a common container.

Service	Standard	Implementation Guidance
Video files	*Mandatory* • ISO/IEC 13818-2 - Generic coding of moving pictures and associated audio information: Video • ISO/IEC 14496-2 - Coding of audio-visual objects -- Part 2: Visual • ISO/IEC 14496-10 - Coding of audio-visual objects -- Part 10: Advanced Video Coding	**Requirements** • Preserve resolution (clarity, colors), scalability, and ability of video • Preserve video metadata, including timecodes and other tagging • Compressibility, preference for lossless compression • Preference for larger resolution and higher audio bitrates

1.1.5. Sound

012. Sound files contain recordings of voice or other audio. This includes audio recordings from meetings if they contain information of permanent value.

Service	Standard	Implementation Guidance
Audio files	*Mandatory* • EBU Tech 3285 - Specification of the Broadcast Wave Format (BWF) – Version 2 • ISO/IEC 11172-3 - Coding of moving pictures and associated audio for digital storage media at up to about 1,5 Mbit/s; PCM Part 3: audio • ISO/IEC 13818-3 - Generic coding of moving pictures and associated audio information -- Part 3: Audio	**Requirements** • Preserve resolution (sampling frequency) and depth • Preserve audio metadata

1.1.6. Geospatial

013. Geospatial information is typically produced, used, and contained in geographic information systems (GIS). The information is related to the still image category, as geospatial information consists of bitmap or vector images plus additional attributes associated with particular locations depicted in the image data.

Service	Standard	Implementation Guidance
Geospatial information (e.g. GIS data)	*Mandatory* • OGC 07-147r2 - OGC KML • OGC 12-128r10 - OGC GeoPackage Encoding Standard V1.0.	**Requirements** • Preserve resolution and scalability • Preserve geospatial metadata

1.1.7. Web Archive

014. The web archive type concern the archival of entire web sites, portals, or parts of them. While some information might be contained in static web pages and is therefore easy to capture, other parts might be dynamically rendered.

015. Web archives typically contain structured textual descriptions as well as still and moving images.

Service	Standard	Implementation Guidance
Web sites and portals	*Mandatory* • ISO 28500 - Information and documentation -- WARC file format. • IETF RFC 2557 - MIME Encapsulation of Aggregate Documents, such as HTML (MHTML)	**Requirements** • Preserve structure and content of web, including scripts • Inclusion of external content might be necessary • Preserve metadata associated with content • Dynamic/interactive or userspecific content is problematic

1.2. PACKAGE STRUCTURES FOR LONG TERM PRESERVATION

016. NATO digital information of permanent value shall be processed by their Information Custodians into single digital information items with associated metadata and packaged into submission and archival information package structures [6].

1.2.1. Submission Information Package

017. NATO digital information of permanent value selected by Information Custodians for long term preservation should be delivered to the NATO Archivist as a Submission Information Package (SIP).

018. The SIP consists of two parts: the actual information packaged as a single digital information item and a set of metadata associated with this item (see Figure 1.2)

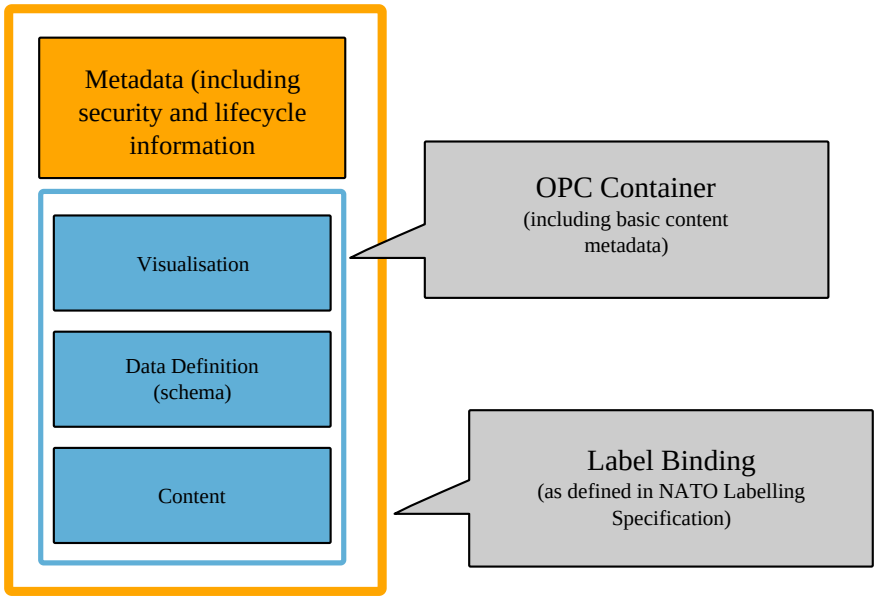

Figure 1.2. Submission Information Package structure

019. The single digital information item has the following structure:

- **Content**: Information of one of the seven types listed under Section 1.1). For certain types of content, primarily data sets (Section 1.1.1), several pieces of information might be grouped. A schema provided as part of the Data Definition can be used to describe the structure of these groupings. For other types such as documents, images, or recordings, information items shall be included individually. Items might contain other objects that should also be preserved in a sustainable format. For example, an archived email message could have text documents as attachments that should be stored in the sustainable formats listed in Section 1.1.2). Guidance on granularity and grouping will be provided by the Archives Committee.

- **Data Definition:** If the Content consists of structured data, a separate Data Definition shall be included that describes the logical structure of the Content. This is primarily applicable to Content of the types Data Set (Section 1.1.1), Geospatial (Section 1.1.6), and Web Archive (Section 1.1.7). The format of the Data Definition shall be XML Schema 1.1.

- **Visualization:** A visualization and human readable representation of contextual information is optional. The format used for the context information shall be one of those listed under Section 1.1).

020. The individual parts (Content, Data Definition and Visualization) shall be packaged as a single digital information item by using the Open Packaging Conventions [7] format.

021. The file name of the packaged single digital information item shall follow the NATO Guidance on File Naming [5]. OPC does not define an extension; the .zip extension shall be used for packages for long term preservation.

022. The SIP or AIP shall contain a basic set of metadata for the container. OPC supports a subset of six Dublin Core metadata elements (creator, description, identifier, language, subject, and title) and two Dublin Core terms (created, modified). The elements shall be filled by the Information Custodian when the OPC container for the single digital information item is created. Note that this metadata refers to the container itself, not to its contents. For example, the creation date is the date the container was created, not the creation date of the content.

023. In addition to the OPC container metadata, the Information Custodian will generate a full metadata description for the content of the SIP, including the classification of the single digital information item.

024. The SIP metadata follows the NATO Core Metadata Specification (NCMS) [1] and the NATO Labelling Specification [3]. Values for all mandatory elements shall be assigned by the Information Custodian. The NATO Archivist shall reject all submissions with incomplete metadata.

025. No information of permanent value packaged in a SIP and submitted by the Information Custodian shall be destroyed unless the SIP has been explicitly acknowledged and accepted by the NATO Archivist.

1.2.2. Archival Information Package

026. If the content of the SIP submitted by an Information Custodian for long-term preservation are accepted by the NATO Archivist, the SIP will be processed into an Archival Information Package (AIP).

027. The AIP consists of the same structure as the SIP, i.e. the single digital information item for long-term preservation packaged as an OPC container, and the NCMS-compliant metadata information bound to the container.

028. As part of the Ingest process, the metadata supplied with the SIP will be augmented by preservation metadata approved by the NATO Archivist. In addition, NATO Archivist shall become the custodian for the AIP.

029. The preservation metadata will be an extension to the NCMS metadata. The extension shall be based on the PREMIS metadata set [8].

References

[1] *NATO Core Metadata Specification*. C3B. Copyright # 2014. NATO Unclassified.

[2] *Information Management Directive for Confidentiality Labelling of NATO Information*. C3B. Copyright # 2014. NATO Unclassified.

[3] *Information Management Guidance for Confidentiality Labelling of NATO Information*. C3B. Copyright # 2014. NATO Unclassified.

[4] *NATO Interoperability Standards and Profiles, Version 8 (NISP V8)*. C3B. Copyright # 2013. NATO Unclassified, Releasable to Australia/New Zealand/Singapore..

[5] *Guidance on File Naming*. C3B. Copyright # 2010. Unclassified, Releasable to PfP..

[6] *Space data and information transfer systems – Open archival information system – Reference model, First Edition*. ISO. Copyright # 2003.

[7] *Information technology -- Document description and processing languages -- Office Open XML File Formats -- Part 2: Open Packaging Conventions*. ISO/IEC. Copyright # 2012.

[8] *PREMIS Data Dictionary for Preservation Metadata, Version 2.0*. PREMIS Editorial Committee. Copyright # 2008.

This page is intentionally left blank