STANDARDS RELATED DOCUMENT

ADatP-4778.1

METADATA BINDING MECHANISM (MBM) – IMPLEMENTATION GUIDANCE

Edition A Version 1

NOVEMBER 2021

NORTH ATLANTIC TREATY ORGANIZATION

Published by the
NATO STANDARDIZATION OFFICE (NSO)

INTENTIONALLY BLANK

NORTH ATLANTIC TREATY ORGANIZATION (NATO)

NATO STANDARDIZATION OFFICE (NSO)

NATO LETTER OF PROMULGATION

10 November 2021

1. The enclosed Standard-related Document ADatP-4778.1, Edition A, Version 1 METADATA BINDING MECHANISM (MBM) – IMPLEMENTATION GUIDANCE, which has been approved in conjunction with ADatP-4778 by the nations in the Consultation, Command and Control Board (C3B), is promulgated herewith.

2. ADatP-4778.1 Edition A, Version 1 is effective upon receipt.

3. This NATO standardization document is issued by NATO. In case of reproduction, NATO is to be acknowledged. NATO does not charge any fee for its standardization documents at any stage, which are not intended to be sold. They can be retrieved from the NATO Standardization Document Database ((https://nso.nato.int/nso/) or through your national standardization authorities.

4. This publication shall be handled in accordance with C-M(2002)60.

Dimitrios SIGOULAKIS
Major General, GRC (A)
Director, NATO Standardization Office

INTENTIONALLY BLANK

TABLE OF CONTENTS

INTENTIONALLY BLANK

CHAPTER 1 INTRODUCTION

1.1. BACKGROUND

The Primary Directive on Information Management (PDIM) prescribes the application of metadata and markings in accordance with NATO policies and directives to facilitate sharing and control of NATO information.

The PDIM defines metadata as structured information that describes, explains, locates, and otherwise makes it easier to retrieve and use an information resource. The structure consists of 'elements', each of which will contain 'values'. The values relate to the resource itself, there may be controls over what the actual values can be.

Metadata is a key enabler for the effective and efficient management of information. Modern automated information systems require information resources to be labelled with metadata.

1.2. OBJECTIVE

The NATO Core Metadata Specification (NCMS) (Reference [3]) defines a set of core metadata elements to support information management in the Alliance.

This document recognizes the existence of communities of interest's specific metadata standards and aims at steering their evolution in the mid to long term and at providing a single mediation standard in the short term to achieve sharing of information among different communities of interest.

1.3. SCOPE

NCMS applies to all NATO information and to any information resource handled or processed by NATO's communications and information systems. NCMS describes information resource and supports its consistent and appropriate handling.

All NATO civil and military bodies are mandated to use NCMS.

Allies and Partners must also use NCMS when handling NATO information.

1.4. NATO METADATA REGULATORY STANDARDS

NATO has the following metadata standards:

- **ADatP-5636** NATO Core Metadata Specification defines the core set of metadata elements that must be used to support interoperable information exchange
- **ADatP-4774** Confidentiality Metadata Label Syntax provides support for the Security Layer metadata elements

- **ADatP-4778** Metadata Binding Mechanism describes how to consistently bind metadata (of any sort) to a finite data object

A number of separate, informative, Standard-related Documents (SRDs) are complementing these three metadata standards by providing implementation and other guidance, see Figure 1.

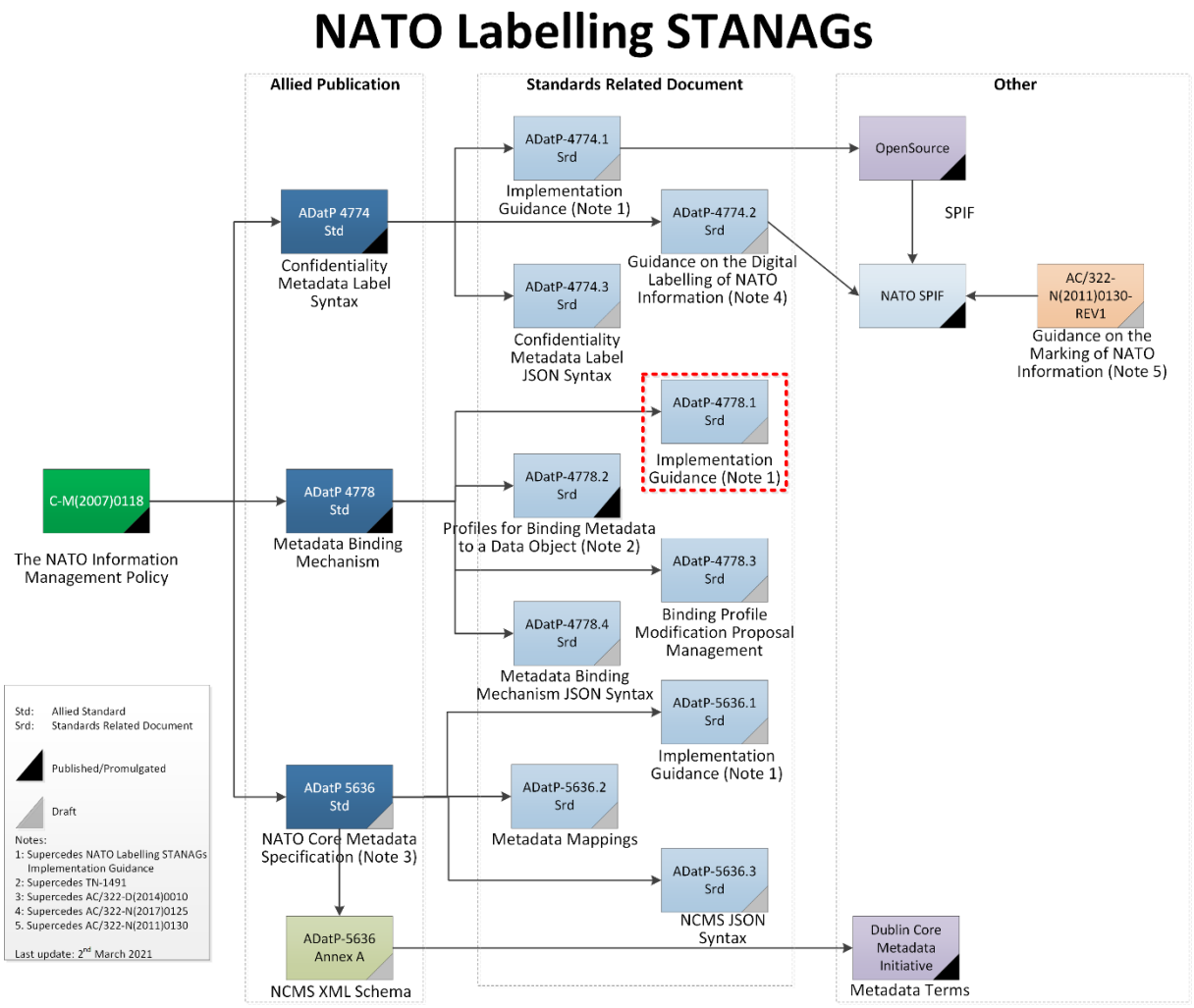

Figure 1: NATO Labelling STANAGs

This document (SRD) is the Implementation Guidance for the Metadata Binding Mechanism (highlighted in a red, dashed box in Figure 1).

CHAPTER 2 OVERVIEW

2.1. INTRODUCTION

This Implementation Guidance provides additional details for those undertaking the implementation of ADatP-4778 Metadata Binding Mechanism within their system.

This Implementation Guidance is optional and implementers of ADatP-4778 are free to follow the guidance that they feel is most appropriate to their particular requirements.

The Implementation Guidance will be periodically updated with new guidance on how to support the implementation of the Metadata Binding Mechanism based upon implementation experience.

2.2. DATA CENTRIC SECURITY

The vision of Data Centric Security (DCS) is to deliver shareable Alliance information, protected at source, controlled for life.

DCS allows for variation in how protection requirements are determined, in what way the enforcement of the protection policy is executed, and in the choice of the underlying access control model. The variation is driven by evolution in the following directions:

1. The level of detail for describing information with metadata;
    a. *Sensitivity metadata*
    b. *Common core metadata*
2. The granularity of access control;
    a. *Clearance based*
    b. *Attribute based*
3. The level of object protection;
    a. *Deny or Grant Access Control*
    b. *Cryptographic Access Control*

The DCS Vision and Strategy (Reference [2]) defines three Maturity Levels (listed below) that have been determined by the variation in the different directions of evolution:

1. Basic Labelling – the majority of new data objects are labelled
    a. Labelling and binding compliant with STANAG 4774 and STANAG 4778
    b. Guard capability to mediate release based on confidentiality labels
    c. Use and management of metadata with the NATO Enterprise
2. Enhanced Labelling – the majority of shared data objects are labelled and domain boundary release controlled.
    a. Integration with the NATO Enterprise Identify and Access Management
    b. Granular labelling of all shareable data objects, including legacy data
    c. Rich metadata compliant with STANAG 5636

   d. Alliance-wide attribute-based access control
   e. Agile response to changing security environment
   f. Metadata labels applied to non-finite data streams e.g. voice and video, with appropriate guard technology
3. Cryptographic protection – data objects controlled post-release
   a. Cryptographic protection for data objects in transit and at rest
   b. Controlled sharing of released data objects (federated digital rights management)
   c. Converged cloud platforms for multi-level data separation
   d. Increasing automation of information sharing and redaction.

Each of the Maturity Levels builds upon the foundations of the previous Maturity Level, and so all three Maturity Levels are dependent upon STANAG 4774 and 4778, and Maturity Levels 2, and 3 are dependent on STANAG 5636 and associated SRDs.

Implementation of STANAGs 4774, 4778 and 5636 across the NATO Alliance facilitates evolution of DCS in the direction of increasing the level of detail for describing information with metadata. Evolution of DCS in the directions of granularity of access control and level of object protection will require the implementation of additional standards and specifications.

This STANAG 4778 Implementation Guidance SRD thus provides guidance on the binding of all types and formats of metadata to data objects for all of the DCS Maturity Levels.

2.3. GRANULAR LABELLING

The traditional approach to labelling data objects is to apply the method that is followed when marking human-readable documents. This method determines the overall confidentiality marking of a document to be equal to the most restrictive confidentiality marking that is applied to any part of the document. Often, the overall confidentiality marking is a result of a recursive process in which sections (paragraphs and chapters) are assigned a confidentiality marking equal to the most restrictive confidentiality marking applied. The traditional approach to labelling has the disadvantage that flexible granular access control (e.g. at the paragraph level) cannot easily be enforced without fully analysing the document (which would be necessary to determine which paragraphs are the real reason for the overall confidentiality marking).

Chapter 3 provides guidance on the use of granular labelling facilitated by the usage of URI fragment identifiers and the XPath Data Model.

2.4. NATO METADATA BINDING SERVICE

The NATO Metadata Binding Service (NMBS) supports a consistent approach for labelling, specifically by:

- The provision of appropriate metadata values, as a catalog, that can be bound to a data object, which can presented to the user. For example, a catalog of
    - confidentiality labels (Reference [4]) for the originatorConfidentialityLabel metadata element; or
    - subject categories for the subjectCategory metadata element
- The provision of human-readable and machine-readable representations of the metadata values to aid the user understand and identify the type. For example, the machine-readable representation of a country may be a three-letter code, while the human-readable representation is the country name, i.e. BEL labelling value rendered as BELGIUM for marking.
- The creation of a binding of the metadata to the data object, by creating a valid STANAG 4778 binding. This includes the ability to create a cryptographic binding.
- The verification of a STANAG 4778 binding and provision of the metadata values in an appropriate form.

Chapter 4 provides an overview of the NATO Metadata Binding Service.

2.5. CANONICALIZATION

Prior to signature generation and signature verification, regardless of the cryptographic mechanism being employed, each data object is required to be converted to a canonical form that is uniquely and unambiguously representable.

The exact details for canonicalization is dependent upon the actual MIME content type for that data object. However, most MIME content types have only one representation that can be considered their canonical representation; hence, canonicalization for that data object is not required.

Chapter 5 provides some further implementation guidance for those MIME content types that require canonicalization.

INTENTIONALLY BLANK

CHAPTER 3 BINDING

3.1. INTRODUCTION

A STANAG 4778 binding provides a general mechanism that allows metadata to be bound to data objects. This chapter provides implementation guidance on:

- how to support granular labelling of data object (for example, the paragraphs within a document)
- the use of URI fragment identifier in granular labelling; and
- the use of the XPath Data Model.

3.2. GRANULAR LABELLING

The traditional approach to labelling data objects is to apply the method that is followed when marking human-readable documents. This method determines the overall confidentiality marking of a document to be equal to the most restrictive confidentiality marking that is applied to any part of the document. For example, if a document has one ‘RESTRICTED’ paragraph whereas all other content is ‘UNCLASSIFIED’ the document as a whole will become ‘RESTRICTED’. Often, the overall confidentiality marking is a result of a recursive process in which sections (paragraphs and chapters) are assigned a confidentiality marking equal to the most restrictive marking applied. The traditional approach to labelling has the disadvantage that flexible granular access control (e.g. at the paragraph level) cannot easily be enforced without fully analysing the document (which would be necessary to determine which paragraphs are the real reason for the overall confidentiality marking).

When implementing the NATO Labelling STANAGS it is recommended to apply the “top-down labelling” approach, whereby the most relaxed confidentiality label is applied to the root element of the data object and only those subsets of the data object that require more restrictive protection have the required confidentiality label bound to them. An example of this is given below with an XML data object (based on the general applicable rules for binding metadata to data objects and subsets thereof provided in STANAG 4778 Section 3.5).

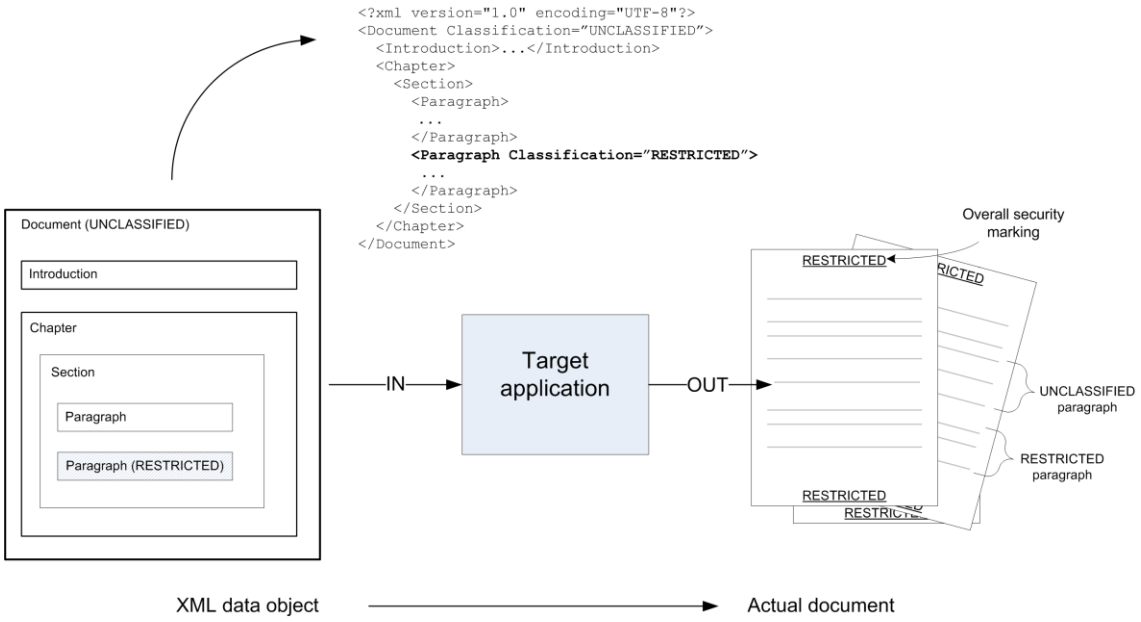

Figure 2: Granular Labelling

Only the nodes that contain 'RESTRICTED' content are labelled accordingly; the other nodes and the XML structure remain 'UNCLASSIFIED'; the target application determines the overall confidentiality marking and outputs the actual document with 'RESTRICTED' marking.

Figure 2 provides an example of top-down XML labelling and the rendering by the target application.

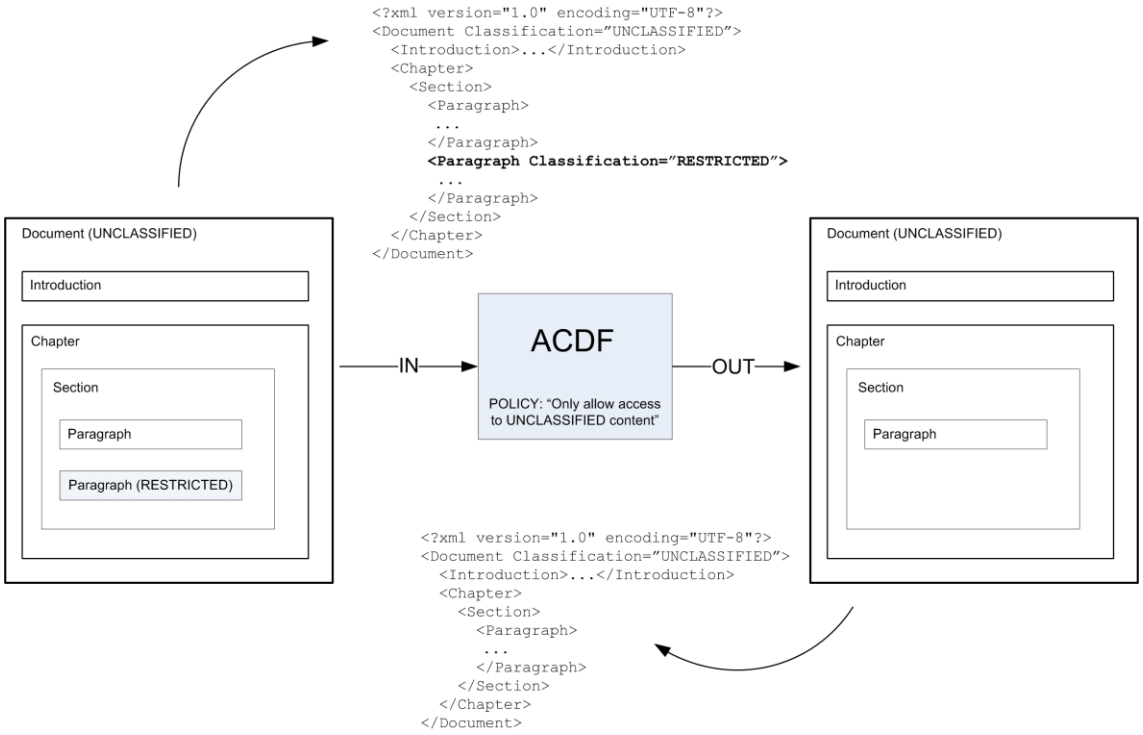

Figure 3: Granular access control

ACDF blocks access to the 'RESTRICTED' nodes and outputs the 'UNCLASSIFIED' nodes; the XML tree structure is preserved; the target application will output an unclassified document.
Figure 3 shows how granular access control is applied to the same example (in the use case that sanitisation/redaction is supported).

3.3. USE OF URI FRAGMENT IDENTIFIER

Granular labelling requirements, whereby multiple confidentiality labels are bound to a data object and subsets of that data object, will result in an application or service needing to understand the semantics of the binding that is specific to the data object format.

This section describes the NATO Labelling STANAGs method for assigning metadata (for example, confidentiality labels) to subsets (parts) of data objects.

The Uniform Resource Identifier (URI) generic syntax is specified in Reference ([6]). A component of the URI is the fragment identifier that is used to build a URI reference. A URI reference is a powerful concept that allows indirect identification of a secondary resource by reference to a primary resource. As such, the URI reference concept supported by the fragment identifier can be utilised to support the principles of "top-down labelling".

The fragment identifier is indicated by the presence of a number sign ("#") character and terminated by the end of the URI.

The significance of the fragment identifier is a function of the content type (also known as media type or MIME type). In other words, unless the content type is known the syntax and the semantics for interpreting the fragment identifier are unknown. Content types are registered on the internet and the registered list of content types are maintained at the Internet Assigned Numbers Authority (IANA, Reference ([7])). Content types that are registered with IANA can also specify how applications must interpret fragment identifiers. In order for systems performing Access Control Decision Functions (ACDF), supporting the principles of “top-down labelling” for data objects, the syntax and the semantics for interpreting the fragment identifier component of the URI must be followed based on the content type of the data object. Therefore by using the URI fragment identifier assigning confidentiality labels to subsets of data objects can be realised.

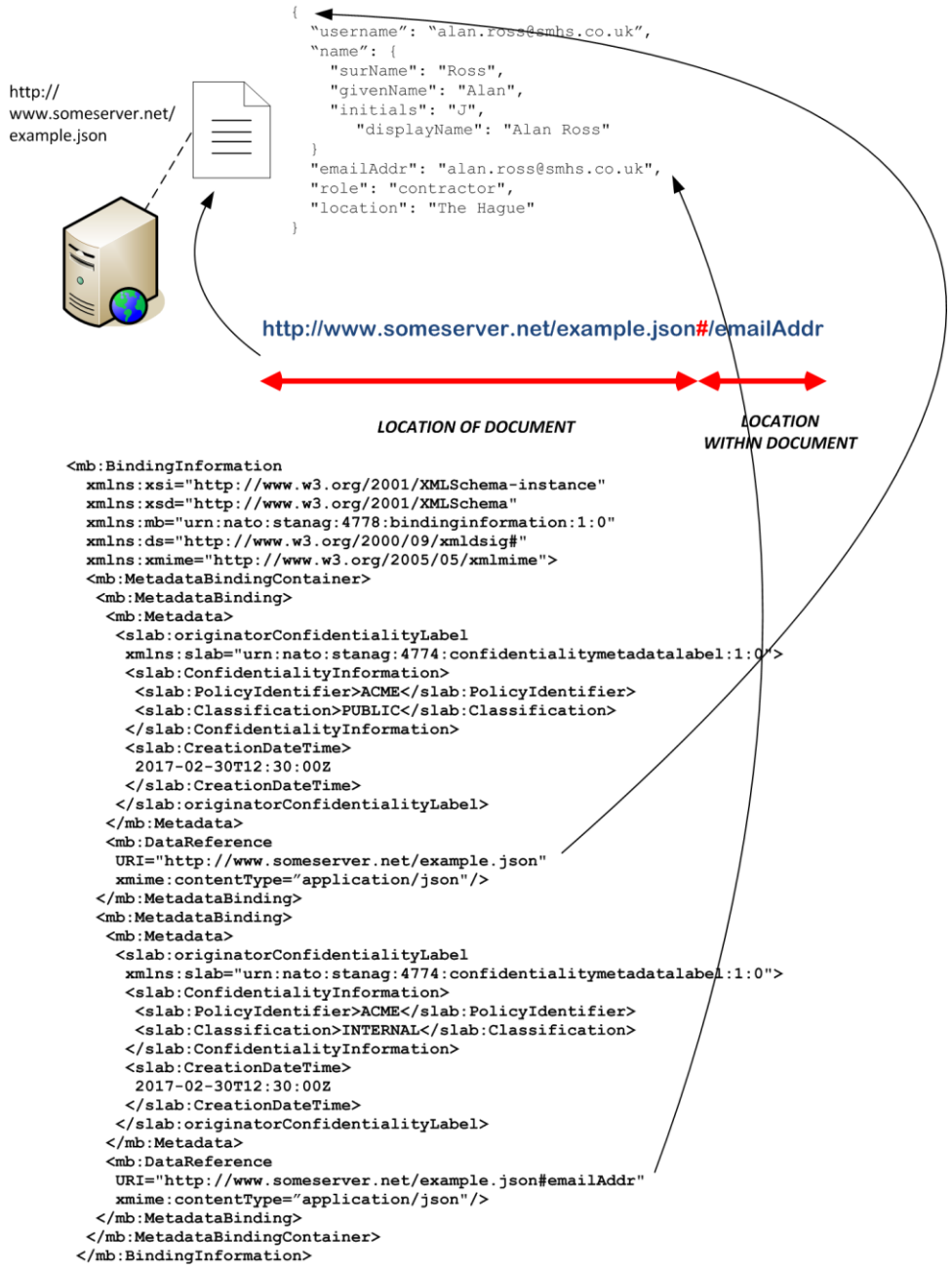

Figure 4: Example JavaScript Object Notation (JSON) document with the use of the URI fragment identifier for binding confidentiality labels to subsets of the JSON document

An example JavaScript Object Notation (JSON) document with the use of the URI fragment identifier for binding confidentiality labels (based on the ACME policy) to subsets of the JSON document is shown in Figure 4.

3.4. XPATH DATA MODEL

XML is a data format for structuring human-readable documents through the use of markup. It is an important specification for expressing structured data used for exchanging information.

The original XML specification did not specify an underlying data model. This omission was rectified by the XML Information Set (Infoset) specification, which is the normative data model for XML. XML instances provided as an XML document can be represented using the XML Infoset, which in turn can be mapped to the XPath data model.

The NATO Labelling STANAGs implementation guidance recommends the use of the XPath as the XML data model. The XML data object can be represented as an XPath node-set (i.e. the set of all nodes in an XPath tree) based on a tree-structured graph. A node in such a tree can be of exactly one of the following seven types: root; element; text; attribute; namespace; processing instruction; and, comment.

An example of an XML data object (instance) and its representation in the XPath data model is illustrated in Figure 5.

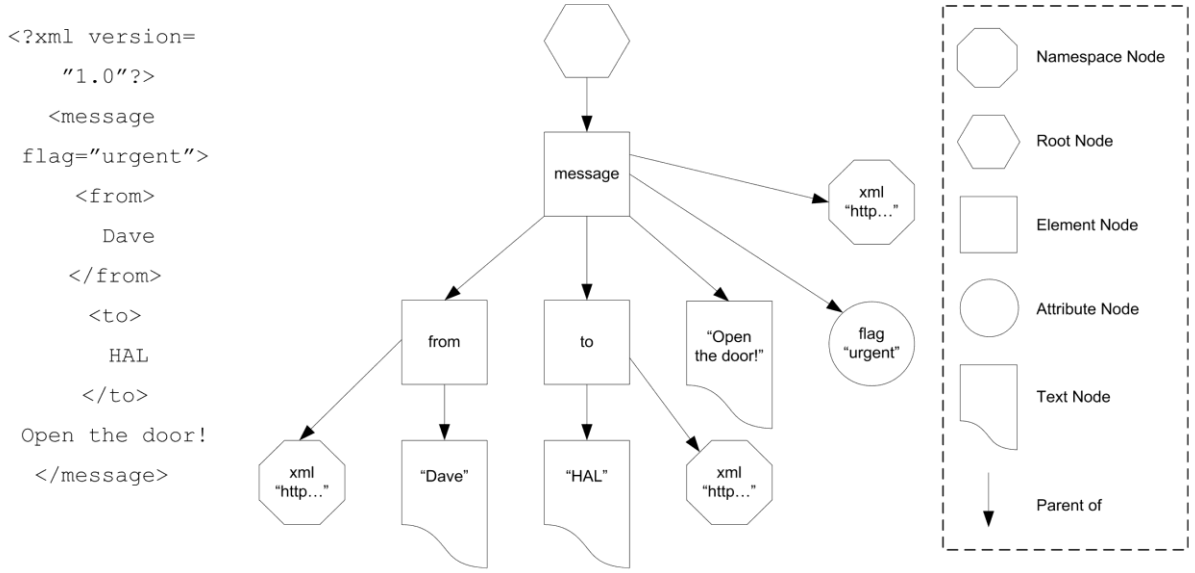

Figure 5: Example of an XML instance as an XPath data model

Example of an XML instance as an XPath data model; a textual representation is given on the left side; the XPath tree structure with its different nodes is shown in the figure on the right-hand side.

The rules for binding metadata to a node (every type) are as follows:

[Rule - 1] Root node: No classification[1] can be associated with the XPath root node. (Note: Note that the XPath 'root node' is different from the 'root element node'.

[1]'Classified information' defined by C-M(2002)49-REV1 and 'non-classified information' by CM(2002)60.

In the example in Figure 5 the element message is the ‘root element node’. The root element node is also called the document element node.)

[Rule - 2] Element node: The information represented by the expanded name of the element node has to be handled in compliance with the policies and guidelines applicable for information at the classification level associated with the element node.

[Rule - 3] Text node: The information represented by the character data of the text node has to be handled in compliance with the policies and guidelines applicable for information at the classification level associated with the text node.

[Rule - 4] Attribute node: The information represented by the expanded name as well as the string value of the attribute node has to be handled in compliance with the policies and guidelines applicable for information at the classification level associated with the attribute node.

[Rule - 5] Namespace node: The information represented by the expanded name as well as the string value of the namespace node has to be handled in compliance with the policies and guidelines applicable for information at the classification level associated with the namespace node.

[Rule - 6] Processing instruction node: The information represented by the expanded name as well as the string value of the processing instruction node has to be handled in compliance with the policies and guidelines applicable for information at the classification level associated with the processing instruction node.

[Rule - 7] Comment node: The information represented by the string value of the comment node has to be handled in compliance with the policies and guidelines applicable for information at the classification level associated with the comment node.

An example in Figure 6 is given which illustrates the Rules above in conjunction with the “top-down labelling” approach described in Granular Labelling, based on Figure 5. For this example ‘U’, ‘R’, and ‘C’ are intended to represent UNCLASSIFED, RESTRICTED and CONFIDENTIAL respectively.

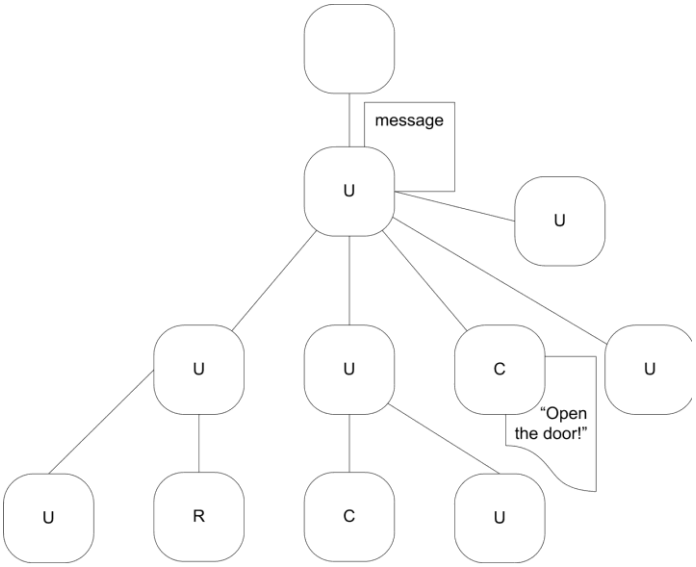

Figure 6: Different classification levels are associated with the individual nodes of the example XML instance (from Figure 5)

Classification levels are associated with the individual nodes of the example XML instance from Figure 5 this results in a tree-structured relationship between confidentiality metadata

The message (root) element node has UNCLASSIFIED confidentiality metadata bound to it. As a result all child nodes have UNCLASSIFIED confidentiality metadata bound to them. Several text nodes have also had RESTRICTED and CONFIDENTIAL confidentiality metadata bound to them replacing the inherited UNCLASSIFIED confidentiality metadata. As such, this allows for a more flexible information assurance mechanism, whereby, if the message is processed by an ACDF that removes nodes of the message that are bound to CONFIDENTIAL confidentiality metadata then only the text nodes “Open the door!” and “HAL” are removed as opposed to the complete message (based on the traditional labelling approach of labelling the overall message with the most dominant confidentiality metadata).

CHAPTER 4 NATO METADATA BINDING SERVICE

4.1. INTRODUCTION

Many communities of interest (COIs) will need to bind metadata to their own data objects in order to comply with their own, and NATO Enterprise, metadata requirements.

In order to support a consistent approach for these COIs, a service has been defined for use within the NATO Enterprise to support the creation and validation of STANAG 4778 compliant metadata bindings.

The NATO Metadata Binding Service (NMBS) supports:

- The provision of appropriate metadata values that can be bound to a data object, which can presented to the user. For example, a catalogue of confidentiality labels or subject categories.
- The provision of human-readable and machine-readable representations of the metadata values to aid the user understand and identify the type. For example, the machine-readable representation of a country may be a three-letter code, while the human-readable representation is the country name.
- The creation of a binding of the metadata to the data object, by creating a valid STANAG 4778 binding. This includes the ability to create a cryptographic binding.
- The verification of a STANAG 4778 binding and provision of the metadata values in an appropriate form.

4.2. SERVICE

The NMBS offers three operations to support the binding of metadata to a data objects These operations are:

1. Get;
2. Set; and
3. Verify.

The operations may be used by a client to generate an appropriate user interface for the user to select metadata to associate with the data object after which the selected metadata can be bound to the data object. The operations may be called multiple times with different parameters in order to generate the appropriate metadata binding.

The three operations are described in more detail below.

4.2.1. Get Operation

The Get operation provides a mechanism to obtain sets of valid metadata that a client may include within a metadata binding.

The sets of Metadata may contain both a human readable representation (c.f. ‘marking’) and a machine readable representation (c.f. ‘label’) for each piece of metadata in the format requested.

The Get operation may be invoked multiple times to determine information about all of the metadata that may be bound to a data object. For example, it may include requests to return:

- all metadata values that may be bound to the data object (for example, all confidentiality labels)
- all metadata values that a given user is allowed to bind to the data object (for example, only confidentiality labels for which the user is cleared)
- all the metadata values that may be bound to the data object for a given recipient (for example, only confidentiality labels for which the recipient is cleared).

The results of these requests can be used to generate an appropriate user interface so that the user so can make an informed selection of metadata values to bind to the data object.

4.2.2. Set Operation

The Set operation undertakes the creation of a binding between the supplied of the metadata and data object. This includes the validation of:

- the metadata values,
- the data object.

The Set operation may use appropriate credentials to provide integrity and authentication of the binding.

4.2.3. Verify Operation

The Verify operation undertakes the validation of the binding including the metadata values and data object(s).

The metadata values may be optionally transformed and mapped to an equivalent representation before being returned to the user in order to provide metadata values that the user is familiar with. For example, converting a confidentiality label that uses the US policy to an equivalent confidentiality label that uses the NATO policy.

4.3. FURTHER INFORMATION

More details of the NMBS are available in TR 2012/SPW007959/02 (Reference **Error! eference source not found.**).

INTENTIONALLY BLANK

CHAPTER 5 CANONICALIZATION

5.1. INTRODUCTION

Prior to signature generation and signature verification, regardless of the cryptographic mechanism being employed, each data object is required to be converted to a canonical form that is uniquely and unambiguously representable, i.e. independent of the surrounding context the same resulting octet stream is yielded.

The exact details for canonicalization is dependent upon the actual MIME content type for that data object. This chapter is focused on enhancing the interoperability with different cryptographic implementations by providing enhanced implementation guidance for canonicalization of certain MIME content types.

5.2. TEXT

Implementations that require to canonicalize text MIME content types SHALL comply with Secure/Multipurpose Internet Mail Extensions (S/MIME) Version 4.0 Message Specification Section 3.1.1 (Reference [8]).

5.3. JAVASCRIPT OBJECT NOTATION (JSON)

Implementations that require to canonicalize JSON content types SHALL comply with JSON Canonicalization Scheme (JCS, Reference [9]).

5.4. EXTENSIBLE MARKUP LANGUAGE (XML)

Implementations that require to canonicalize XML content types SHALL comply with the ADatP-4778.2 Chapter 2 Annex A XML Normalization section (Reference [5]).

To assist with performing the rules specified with the XML Normalization section, an XML Stylesheet (XSLT) 1.0 transform is provided and published in the NATO Metadata Registry and Repository (NMRR) at:

https://nmrr.ncia.nato.int/rest/doc/NATO/Information%20Assurance/OLP/XML_Normalisation_1.0.xsl

If the XML data object is a child node of a XML document (root or document element node), that XML data object SHALL contain all the namespaces utilized within that XML data object.

For cryptographic mechanisms other than XML Signature it is necessary to pass the output from XML Normalization XSLT transform through XML canonicalization in order to ensure that the rule *Namespace declarations SHALL appear before attribute declarations* is fulfilled. XSLT processors are not obliged to put Namespace

declarations before attributes as both are classed as xml nodes; hence are not differentiated between. XML canonicalization follows the XPath data model by putting namespace nodes before all attribute nodes.

This document RECOMMENDS that the canonicalization algorithm at

http://www.w3.org/2006/12/xml-c14n11

is used to process the output from the XML Normalization XSLT transform prior to passing into the cryptographic library[2] for signature generation or signature validation.

For XML Signature implementations it is RECOMMENDED that the XML Signature library when performing Core Signature Generation does not use a namespace prefix for the <Signature/> element and preserves whitespace when creating the XML Document containing the <Signature/> element.

The <Signature/> element SHALL NOT be passed through the XML Normalization process prior to being provided as input to a XML Signature library for Core Signature Verification[3].

5.5. CONCISE BINARY OBJECT REPRESENTATION (CBOR)

Implementations that require to canonicalize CBOR content types SHALL conform to the “Core Deterministic Encoding Requirements” as defined in section 4.2.1 of CBOR (Reference [10]).

Additionally, implementations SHALL enforce that protocol specifications are conformant with the “Additional Deterministic Encoding Considerations” as defined in section 4.2.2 of CBOR (Reference [10]).

[2] Note: This step is NOT REQUIRED for XML Signature cryptographic libraries.
[3] The <Signature/> element is created by the XML Signature library during Core Signature generation and cannot be normalized prior to its creation. As such, the <Signature/> elements that are signed can only be covered by the standard XML Signature library canonicalization facilitated by the canonicalised method specified in the <CanonicalizationMethod> element during Core Signature generation and verification.

CHAPTER 6 REFERENCE MATERIALS

6.1. REFERENCES

[1] NATO Metadata Binding Service, Technical Report 2012/SPW007959/002, April 2011.

[2] IMSM-1049-2019 (INV) “Data Centric Security Vision and Strategy for the Alliance Federation, including the NATO Enterprise” 17th May 2019.

[3] ADatP-5636, “NATO Core Metadata Specification”, Edition A, Version 1, September 2020

[4] ADatP-4774.1, “Implementation Guidance”, Standard Related Document (SRD), Edition A, Version

[5] ADatP-4778.2, “Profiles for Binding Metadata to a Data Object”, Standard Related Document (SRD), Edition A, Version 1.

[6] IETF RFC 3986, “Uniform Resource Identifier (URI): Generic Syntax”, at http://tools.ietf.org/html/rfc3986, January 2005

[7] IANA MIME Media Types, “MIME Media Types”, at http://www.iana.org/assignments/media-types

[8] IETF RFC 8551, “Secure/Multipurpose Internet Mail Extensions (S/MIME) Version 4.0 Message Specification”, at http://tools.ietf.org/html/rfc8551, April 2019

[9] IETF RFC 8785, “JSON Canonicalization Scheme”, at http://tools.ietf.org/html/rfc8785, June 2020

[10] IETF RFC 8949, “Concise Binary Object Representation (CBOR)”, at http://tools.ietf.org/html/rfc8949, December 2020

ADatP-4778.1(A)(1)