UK Location Programme

Location Information Interoperability Board

Metadata Working Group

Recommendations for XML Schemas for Encoding Metadata

DOCUMENT CONTROL

Change Summary

Version	Date	Author/Editor	Change Summary
0.1	2009-11-25	James Rapaport	First draft
0.2	2010-02-03	James Rapaport	Updates following review process. Document will become source for integrated MWG position paper document.

References

Ref.	Title/Version/Publication Date/Author
[1]	UK GEMINI Standard Version 2.0 (http://www.gigateway.org.uk/metadata/pdf/GEMINI2.pdf)
[2]	ISO 19115 – Geographic Information – Metadata
[3]	ISO / TS 19139 – Geographic Information – Metadata – XML Schema Implementation
[4]	INSPIRE Metadata Implementing Rules: Technical Guidelines based on ISO 19115 and ISO 19119 (http://inspire.jrc.ec.europa.eu/reports/ImplementingRules/metadata/MD_IR_and_ISO_20090218.pdf)
[5]	OpenGIS Catalogue Services Specification 2.0.2 – ISO Metadata Application Profile, OGC Document 07-045
[6]	XML Linking Language (XLink) Version 1.0, W3C Recommendation 27 June 2001, http://www.w3.org/TR/xlink/
[7]	ISO 19757-3 – Information Technology – Document Schema Definition Languages (DSDL) – Part 3: Rule-based validation – Schematron (http://standards.iso.org/ittf/PubliclyAvailableStandards/index.html (search for Schematron))
[8]	Technical Guidance for INSPIRE Discovery Services, Version 2.0, 2009-07-22, Network Services DT, http://inspire.jrc.ec.europa.eu/documents/Network_Services/Technical Guidance Discovery Services v2.0.pdf

Terminology

Term	Definition
CRS	Coordinate Reference System
CS-W ISO AP	*OGC* Catalogue Services Specification 2.0.2 – ISO Application Profile [5]
DIS	Draft International Standard (ISO)
EDEN	l'Equipe D'Experts en Normalisation[1]. A collaborative project of the *IGN*.
EPSG	European Petroleum Survey Group (defunct but the abbreviation is retained in the name of the EPSG Geodetic Parameter Dataset). Since 2005 *OGP* Surveying and Positioning Committee.

[1] http://eden.ign.fr/welcome

Term	Definition
GEMINI2	The UK discovery metadata profile of ISO 19115
GML	Geography Markup Language
INSPIRE	Infrastructure for Spatial Information in Europe
IGN	Institut Géographique National
ISO	International Organisation for Standardisation
LIIB	Location Information Interoperability Board (of UKLP)
MEDIN	Marine Environmental Data and Information Network[2]
Metadata Instance	Physically instantiated metadata. In the context of this document a metadata instance will be an XML encoded dataset conforming to ISO / TS 19139.
OGC	Open Geospatial Consortium
OGP	International Association of Oil and Gas Producers
SVRL	Schematron Validation Report Language
URI	Uniform Resource Identifier. URIs identify resources and may allow access to representations of the resources. A URL is a URI.
URN	Uniform Resource Name A URN is a *URI* that uses the URN scheme e.g. urn:ogc:def:crs:EPSG::4258 is the URN for ETRS89 in the *EPSG* Geodetic Parameter Dataset
XLink	*XML* Linking Language Used to include metadata element data by-reference rather than by-value.
XML	eXtensible Markup Language
XSD	*XML* Schema Definition
XSL	eXtensible Stylesheet Language
XSLT	*XSL* Transformation

[2] http://www.oceannet.org/

CONTENTS

1 INTRODUCTION

1.1 Background

1 The LIIB has recommended that UKLP discovery metadata be based on the GEMINI2 [1] profile of ISO 19115. GEMINI2 defines the content for metadata but does not express rules for encoding metadata. Metadata will be created and transmitted to a discovery system. Between creation and reception a validation stage will be necessary in order to ensure that the metadata conforms to the minimum requirements of the underlying standards.

2 XML is a technology that can be used for encoding, transmitting and storing data. In addition a range of related technologies have been developed using XML structures and grammars which can be used to validate XML. XML documents that define structures and rules are called schemas.

3 The ISO 19115 [2] standard for metadata presents a logical model for metadata and a set of metadata elements and their meaning. A set of XSD schemas defining an associated XML element structure is defined by ISO / TS 19139 [3]. GEMINI2 is a profile of ISO 19115 but it does not consider the encoding of metadata.

4 XSD schemas can only validate the structure of XML documents. Whilst validation exceptions will be raised if nodes in an XML document are not in the required order or are missing, XSD schemas cannot validate values of XML nodes and constraints that might be placed on them in a profile. Consequently a second layer of validation must be introduced using a different technology. XSL transformations or Schematron, which is based on XSL transformations, can undertake this validation.

1.2 Aims and Objectives

5 The aim of this document is to introduce and discuss the use of ISO / TC 19139 XSD schemas for encoding GEMINI2 metadata. The following objectives are identified:

- Make recommendations on the use of ISO 19139 XSD schemas for encoding GEMINI2 metadata
- Describe the ISO 19139 XSD schema sets that are available
- Describe validation mechanisms
- Introduce and describe Schematron
- Identify further areas of work required to facilitate the encoding of GEMINI2 metadata using ISO 19139 XSD schemas

1.3 Assumed Knowledge

6 It is assumed that readers will be familiar with XML and its related technologies, XSD and XSL. Readers who require background information are referred to the W3Schools introductions to the technologies:

- XML - http://www.w3schools.com/xml/xml_whatis.asp
- XSD - http://www.w3schools.com/schema/schema_intro.asp
- XSL - http://www.w3schools.com/xsl/xsl_languages.asp

1.4 Summary of Results and Recommendations

7 The results and recommendations from the discussion in this document are summarised below.

8 It is recommended that GEMINI2 metadata be encoded according to the ISO 19139 XSD schemas. The following requirements for the schemas are identified:

- Support for an ISO 19139 encoding
- Support for an ISO 19136 encoding
- Support for an ISO 19119 encoding

9 It is recommended that the ISO 19139 XSD schema set that shall be used for the validation of ISO 19139 XML encoded GEMINI2 metadata shall be those in the IGN EDEN schema repository dated 2009-03-16.[3] This schema set meets the requirements in full and is indeed wider in scope than ISO 19139, incorporating schemas for other ISO / TC 211 standards. However, this recommendation should be kept under review.

10 The ISO 19139 XSD schemas should be adopted in full. There is no need to profile them or to develop application schemas.

11 It is recommended that guidance be written setting out how to encode each GEMINI2 metadata element in ISO 19139 XML. It will follow and expand on the INSPIRE guidelines [4].

12 It is recommended that Schematron is used for the validation of ISO 19115 and GEMINI2 constraints.

13 An initial version of the GEMINI2 Schematron schema has been written and is available for testing.[4] Certain constraints in ISO 19115 must also be tested using a Schematron schema. A Schematron schema for this is available from MEDIN.[5]

1.5 Acknowledgements

14 This document has been written by James Rapaport of SeaZone Solutions Limited with funding from the Marine Environmental Data and Information Network (MEDIN).

[3] http://eden.ign.fr/xsd/isotc211/isofull/20090316

[4] https://my.huddle.net/workspace/document/8325992?workspaceid=7820584&directoryid=8294838

[5] http://www.oceannet.org/marine_data_standards/medin_approved_standards/documents/schematron_update_17nov09.zip

2 ISO 19139 SCHEMAS

2.1 Introduction

15 The ISO 19139 XML schema comprises several XSD files. This section describes the namespaces of the schemas and the various locations from which the schemas can be downloaded. Additionally, examples of XML encoding are provided in order to show some of the ways that the XML can be encoded.

16 The relationships between the separate XSD files of the ISO 19139 XML schema set can be seen in diagrammatic form in Appendix A.

2.2 Namespace Definitions

17 The ISO 19139 schemas comprise a number of namespaces. They are listed and defined in Table 1.

Table 1 – ISO 19139 Namespaces

Prefix	Definition	URI
gco	Geographic Common extensible markup language	http://www.isotc211.org/2005/gco
gmd	Geographic MetaData extensible markup language	http://www.isotc211.org/2005/gmd
gmx	Geographic Metadata XML Schema	http://www.isotc211.org/2005/gmx
gss	Geographic Spatial Schema extensible markup language	http://www.isotc211.org/2005/gss
gsr	Geographic Spatial Referencing extensible markup language	http://www.isotc211.org/2005/gsr
gts	Geographic Temporal Schema extensible markup language	http://www.isotc211.org/2005/gts
srv[6]	Service metadata	http://www.isotc211.org/2005/srv

18 Note that the service metadata namespace srv is not a part of the ISO / TS 19139 standard. Indeed the ISO 19119 standard is not referenced from ISO / TS 19139. GEMINI2 will need to use elements from the srv namespace to encode service metadata.

19 ISO 19139 also depends on other schemas for encoding certain structures. GML is used for encoding geometries and temporal extents, and XLink is used for encoding referential links. The namespaces are listed in Table 2. Note that the GML namespace has changed between GML 3.2.0 and 3.2.1. See section 2.5.1.

Table 2 – Other Namespaces relevant to ISO 19139

Prefix	Definition	URI
gml	Geography Markup Language (version 3.2.0 and earlier)	http://www.opengis.net/gml
gml	Geography Markup Language (version 3.2.1)	http://www.opengis.net/gml/3.2

[6] Schemas not released as part of ISO 19139 schemas. Note that the srv namespace is not part of the ISO / TS 19139 standard, is not currently registered by ISO / TC 211 and that the ISO / TS 19139 standard does not refer to ISO 19119 at all.

xlink	XML Linking Language	http://www.w3.org/1999/xlink

2.3 Schema Locations

20 Versions of the ISO 19139 XSD schemas can be found at a number of locations. There are subtle differences between the schema sets which arise from the use of different versions of the GML schemas and the availability of the service metadata schemas.

21 The most recent schema set is that released, unofficially, by the IGN EDEN project (Table 7). This schema set is the most complete and up-to-date. It includes ISO 19119 metadata elements in the srv namespace and GML from version 3.2.1 of the schema set.

Table 3 – ISO Publicly Available Standards

Name:	ISO Publicly Available Standards
URL:	http://standards.iso.org/ittf/PubliclyAvailableStandards/ISO_19139_Schemas/
Date:	August 2007
Comments:	This schema set uses the GML 3.2.1 (ISO 19136) XML schemas. The GML schemas are referenced in the ISO 19136 folder of the ISO Publicly Available Standards directory. The service metadata schemas, in namespace srv, are not included.

Table 4 – OGC Schema Repository (1)

Name:	OGC Schema Repository
URL:	http://schemas.opengis.net/iso/19139/20070417/
Date:	April 2007
Comments:	This schema set is identical to the schema set published in the ISO Publicly Available Standards directory. The GML schemas are referenced in the OGC Schema Repository.

Table 5 – OGC Schema Repository (2)

Name:	OGC Schema Repository
URL:	http://schemas.opengis.net/iso/19139/20060504/
Date:	June 2006
Comments:	This schema set uses the GML 3.2.0 (ISO / DIS 19136) XML schemas. The GML schemas are included with the ISO 19139 schema set. The service metadata schemas, in namespace srv, are included in this schema set. This schema set is the ISO 19139 release candidate (ISO TS RC 19139).

Table 6 – OGC Schema Repository (3)

Name:	OGC Schema Repository - APISO
URL:	http://schemas.opengis.net/csw/2.0.2/profiles/apiso/1.0.0/
Date:	August 2007
Comments:	This schema is for the CS-W 2.0.2 application profile (CS-W ISO AP [5]). The apiso.xsd schema is in fact a wrapper to include the srv namespace for service metadata with the gmd namespace. The referenced schemas are in the OGC Schema Repository in the location shown in Table 5. The base specification of the INSPIRE discovery service [8] is the CS-W ISO AP.

Table 7 – IGN EDEN Schema Repository

Name:	IGN EDEN Unofficial Repository of Schemas based on ISO / TS 19139 Encoding Rules
URL:	http://eden.ign.fr/xsd/isotc211/isofull/20090316
Date:	March 2009
Comments:	This schema set includes all the ISO / TS 19139 schemas and, in addition, service metadata in the srv namespace and ISO 19115-2 concepts in the gmi namespace. The schemas use ISO 19136 GML (version 3.2.1).

2.4 Ancillary Schemas

2.4.1 GML

22 The GML 3.2.1 (ISO 19136) schema set is available from the ISO Publicly Available Standards directory (Table 8).

Table 8 – GML 3.2.1 (ISO 19136)

Name:	ISO Publicly Available Standards
URL:	http://standards.iso.org/ittf/PubliclyAvailableStandards/ISO_19136_Schemas/
Date:	August 2007
Comments:	None.

2.4.2 XLink

23 XLink is a W3C specification which describes a structure of XML elements that allow the creation and description of links between resources [6]. Currently there is no W3C XML schema to accompany the specification so OGC have created a candidate XLink schema for use with GML. It is also used by ISO 19139 XML schemas (Table 9).

Table 9 – XLink

Name:	OGC Schema Repository
URL:	http://schemas.opengis.net/xlink/1.0.0/
Date:	December 2005
Comments:	None.

2.5 Schema Differences

2.5.1 GML Version

24 One of the main differences between the various schema sets is the version of GML that is referenced. GML is used in the ISO 19139 XML schemas to provide data types for encoding time (eg Temporal Extent) and coordinate reference systems (eg Vertical CRS). The main difference between the GML schema versions is a change in namespace:

25 GML 3.2.1 (ISO 19136) – http://www.opengis.net/gml/3.2

26 GML 3.2.0 (ISO / DIS 19136) – http://www.opengis.net/gml

27 Inspection of the official GML repository[7] indicates that the GML version went directly from 3.1.1 to 3.2.1. GML version 3.2.0 only exists in the OGC 2006 ISO 19139 schema set (Table 5). It is recommended that ISO 19139 schemas referencing the GML 3.2.1 (ISO 19136) schemas be used.

28 A valid ISO 19139 metadata instance must include a reference to the appropriate GML namespace. This will depend on the ISO 19139 schemas that are used for validation. An XML instance referencing the GML 3.2.0 namespace will appear invalid if validated against a schema referencing the GML 3.2.1 namespace, and *vice versa*. However, in order to achieve validity the GML namespace identified in the metadata instance should be changed accordingly. The validation error is not caused by an error in the structure of any GML elements in use in the metadata instance.

2.5.2 Service Metadata

29 The ISO TS RC 19139 schema set (see Table 5) includes a version of the service metadata schemas which is also used by the OGC CS-W application profile. Service metadata schemas are also available in the unofficial release from EDEN (see Table 7). However, the schemas are absent from the official ISO schema set.

2.6 Use of ISO 19139 Schemas

30 It is recommended that the ISO 19139 schemas are adopted in full. There is no need to profile them or to develop application schemas for use in encoding GEMINI2 metadata. In working in this way metadata conforming to different profiles of ISO 19115 and ISO 19119, which do not include extensions, will be interoperable with GEMINI2 metadata. This includes domain specific implementations of GEMINI2 which include additional elements, so long as those elements are chosen from the ISO 19115 and ISO 19119 element set. Where the obligation of a metadata element is changed by a profile, a constraint validation technology, such as Schematron, should be used to enforce the obligation. This implies that obligations can only be tightened so that it is not possible to make elements which are mandatory in ISO 19115 optional in a profile.

31 The discussion in section 2.3 has shown that there are a number of subtly different physical instances of the ISO 19139 XSD schemas. Therefore, we need to identify which schema set should be used in encoding GEMINI2 metadata. Let's start by listing some general requirements that are needed for encoding GEMINI2. The schema set must:

- Support an ISO 19139 encoding
- Support an ISO 19136 encoding, implying the use of GML 3.2.1
- Support an ISO 19119 encoding, implying the need for metadata elements in the srv namespace

32 These requirements are met in part by the official ISO 19139 schemas (Table 3) but this schema set includes no schemas for service metadata. This is also true of the OGC schema set released in 2007 (Table 4). The OGC schema set released in 2006 (Table 5) includes the service metadata schemas but is based on GML 3.2.0. The only schema set that meets all the requirements is the unofficial release from EDEN (Table 7).

33 For pragmatic reasons at this stage it is recommended that the EDEN schema set is used as the normative schema for ISO 19139 XML encoded GEMINI2 metadata. The general issue of the availability of multiple

[7] http://schemas.opengis.net/gml/

copies of the schema sets is well known within OGC, ISO / TC 211 and INSPIRE circles and solutions are being proposed. Therefore, this recommendation should be kept under review.

3 VALIDATION

3.1 Introduction

34 XML can be assessed against schemas to ensure that the structure of the XML conforms to the structure defined in the schema. When the structure does conform the XML is said to be schema valid. A general limitation of XML schemas is that they are grammar based which means that they do not provide a means of validating element values or domain specific profiles of ISO 19115.

35 Schematron is an XML technology for validating XML according to constraints placed on the XML. Constraints may apply to element values, or the presence of an element based on the presence or value of another element.

36 Both validation technologies are required for ISO 19139 encoded metadata.

3.2 XSD Schema Validation

37 The ISO 19139 schema comprises more than one physical XSD file so which should be used as the entry point for validating a metadata instance? It is proposed that GEMINI2 metadata instances be validated against the gmx.xsd schema.

38 In order to validate service metadata instances the srv.xsd schema must be used and compiled together with the gmx.xsd schema.

3.3 Schematron Validation

3.3.1 Introduction

39 Schematron provides another way of validating XML by looking for tree patterns and element content. Schematron works by making assertions about elements of the XML which resolve to true or false. If an assertion resolves to false it indicates a failure and the overall validation fails. The assertions are written using XPath in a Schematron schema (*.sch) which is itself expressed in XML. Schematron is designed to be used in conjunction with other validation processes and is based on XSLT and XPath so it is simple to implement.

40 The ISO / TS 19139 standard refers to Schematron as a means of testing some constraints that are set by ISO 19115 but are not testable with XSD. These constraints are set out in Table A.1 of ISO / TS 19139.

41 Schematron has been standardised by ISO as ISO 19757 Part 3. MEDIN has developed Schematron schemas for its profile of ISO 19115 using this standard.

42 The home of Schematron is http://www.schematron.com/.

3.3.2 Validation Mechanism

43 Schematron is implemented as a concatenated series of XSL transformations. The process is shown in Figure 1. The red boxes represent XSL stylesheets that are provided as part of Schematron. The blue box represents a bespoke Schematron schema. This schema is implementation specific and one will need to be written to encapsulate the constraints of GEMINI2. The blue stars represent the occurrence of an XSL transformation. The Schematron schema is first styled with the iso_abstract_expand.xsl stylesheet to produce Intermediate.sch. Schematron allows for the creation of abstract patterns which are encoded in a

schema once but used many times. These patterns must be physically instantiated at run time and this is accomplished by this step. The Intermediate.sch is a Schematron schema. It may be saved to disk or held in memory. Intermediate.sch is then styled with iso_svrl_for_xslt1.xsl to produce TheSchema.xsl. TheSchema.xsl is in fact a stylesheet which is used to style a metadata dataset, represented here by IsoMetadata.xml. This step is where the 'validation' occurs and the output is TheResult.xml which is encoded in SVRL.

Figure 1 – Schematron Stylesheet Transformation Steps

3.4 Metadata Validation

3.4.1 Process

44 With the introduction of the Schematron schema to validate the profile, the overall metadata validation workflow becomes a three stage process as shown in Figure 2. First, a candidate metadata set must be validated against the ISO 19139 schemas. If the metadata set proves to be schema valid, it can then be validated against the ISO 19139 Table A.1 Constraints Schematron schema. The Schematron schema relies on hardcoded XPath statements which will only work effectively on a schema valid XML set. Finally, if the XML is still valid it can be validated against the GEMINI2 Profile Schematron schema.

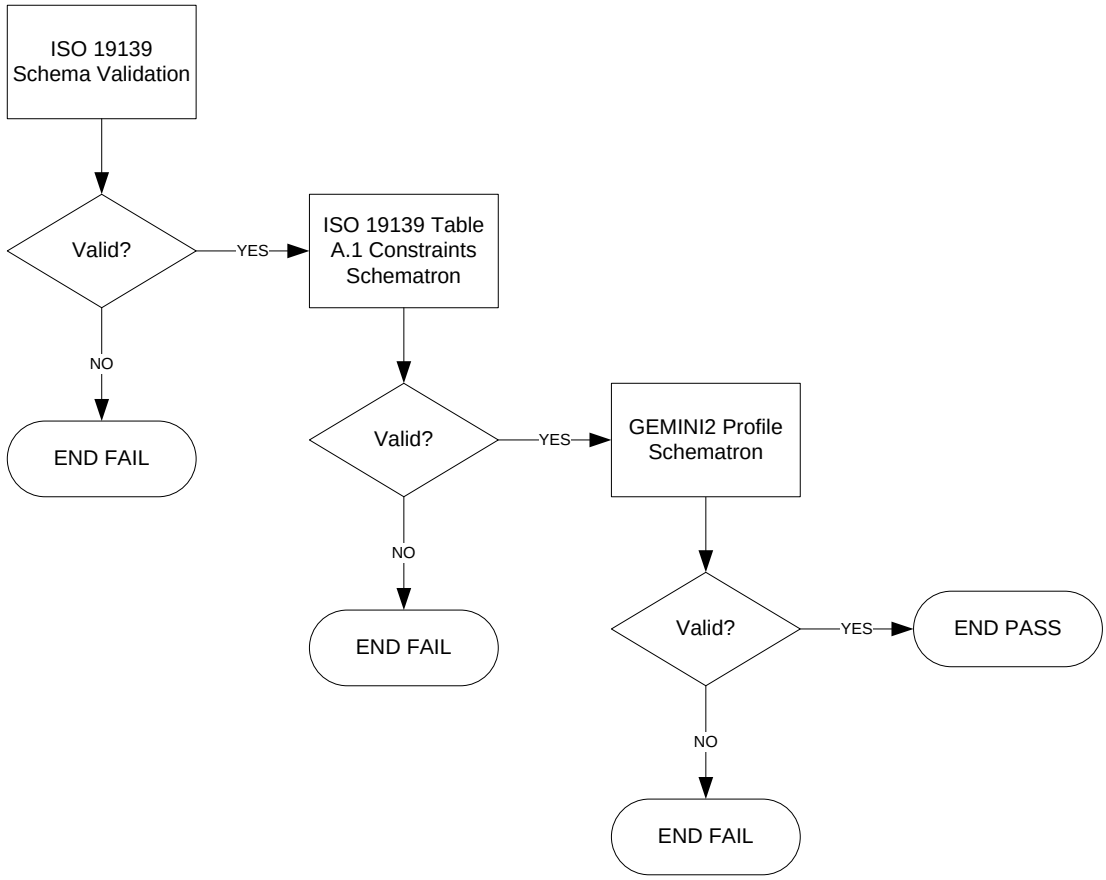

Figure 2 – Metadata Validation Workflow

3.4.2 Schematron Schemas

45 Figure 2 shows that two Schematron schemas are required for a full validation of GEMINI2 metadata: one validating for ISO constraints and one validating for the GEMINI2 constraints. A Schematron schema validating for ISO constraints is available from MEDIN.[8] An initial draft of the GEMINI2 Schematron schema is available (for testing only) in the UKLP MWG workspace.[9]

46 It should be noted that INSPIRE have made a Schematron validator available (currently only for testing) on the Geoportal.[10]

[8] http://www.oceannet.org/marine_data_standards/medin_approved_standards/documents/schematron_update_17nov09.zip

[9] https://my.huddle.net/workspace/document/8325992?workspaceid=7820584&directoryid=8294838

[10] http://www.inspire-geoportal.eu/index.cfm/pageid/48

4 ENCODING GUIDANCE

47 Tools such as XSD schemas and Schematron schemas provide a useful resource for creators of metadata instances. However, normative guidance on how to encode GEMINI2 metadata elements in ISO 19139 XML is not available. ISO 19139 XML is not simple and there are many ways of achieving similar results, some of which may not be obvious to the uninitiated.

48 It is strongly recommended that guidance is written to show how each element of GEMINI2 should be encoded in ISO 19139 XML. In addition the guidance should include fully formed and valid XML instances by way of examples. The document would be similar in scope and structure to the guidance released for INSPIRE [4].

5 INSPIRE METADATA ENCODING

49 INSPIRE does not mandate an encoding for metadata and has no plans to do so. However, technical guidelines [4] have been published which show how ISO 19139 XSD schemas can be used to encode INSPIRE metadata. The INSPIRE technical guidance for discovery services [8] indicates that the base specification for discovery services is the CS-W ISO AP [5].

50 It is understood that an ISO 19139 encoding for GEMINI2 metadata will provide for the necessary interoperability with INSPIRE.[11] Clearly when metadata are harvested from the UKLP discovery system by the INSPIRE system the metadata will need to be encoded according to the CS-W ISO AP. It does not necessarily follow that the UKLP discovery system will require that metadata instances are submitted encoded according this profile but it needs to be taken into account when considering the encoding of GEMINI2 metadata.

[11] Dr. Max Craglia, INSPIRE Metadata Lead, pers comm.

APPENDIX A

Namespace gco

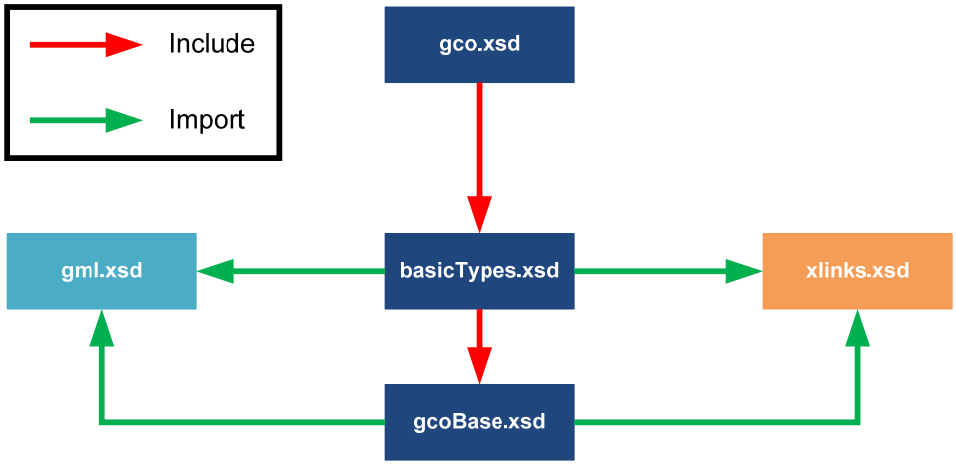

Figure 3 – Namespace gco – Schema Relationships

Namespace gmd

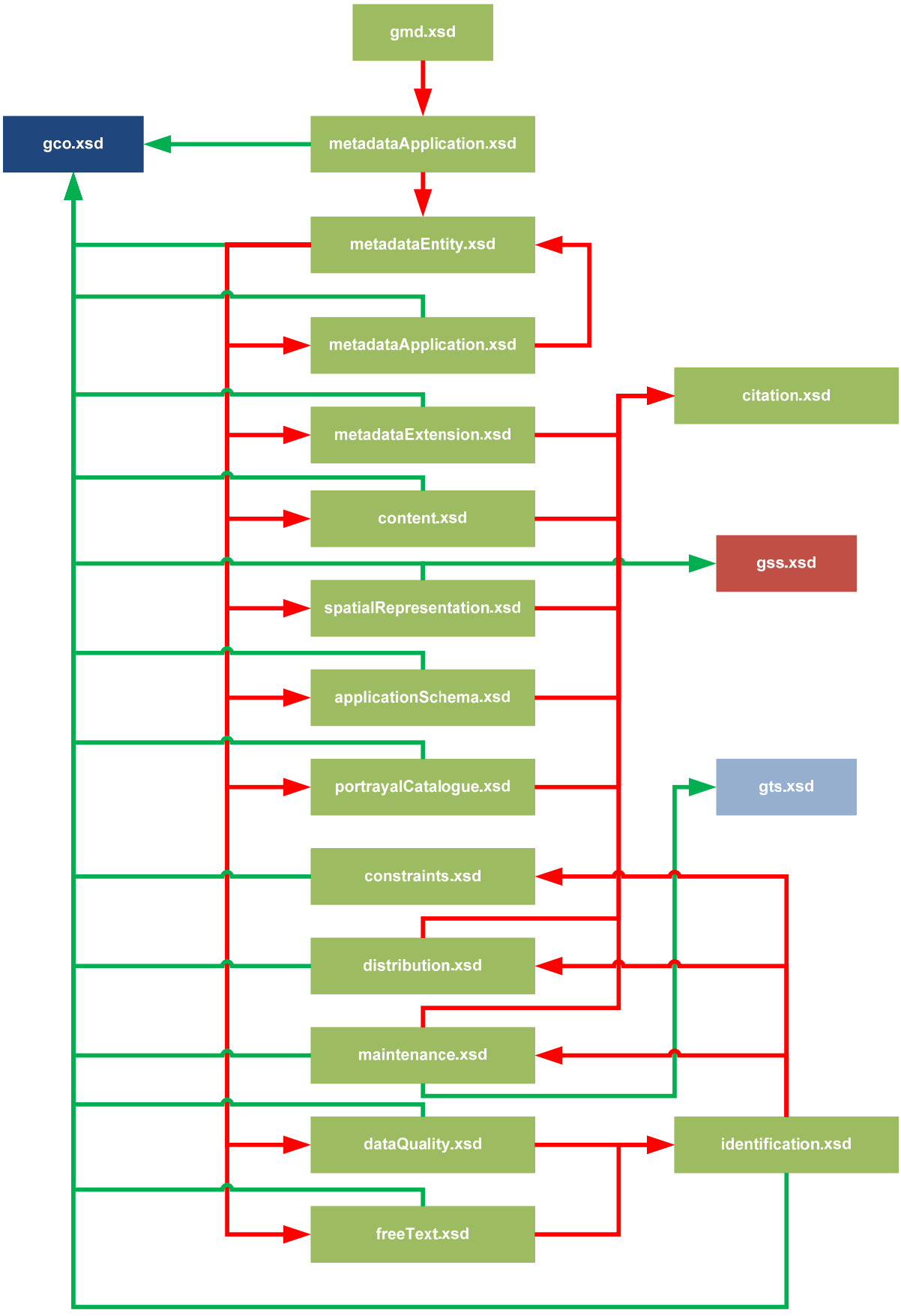

Figure 4 – Namespace gmd – Schema Relationships

Namespace gmx

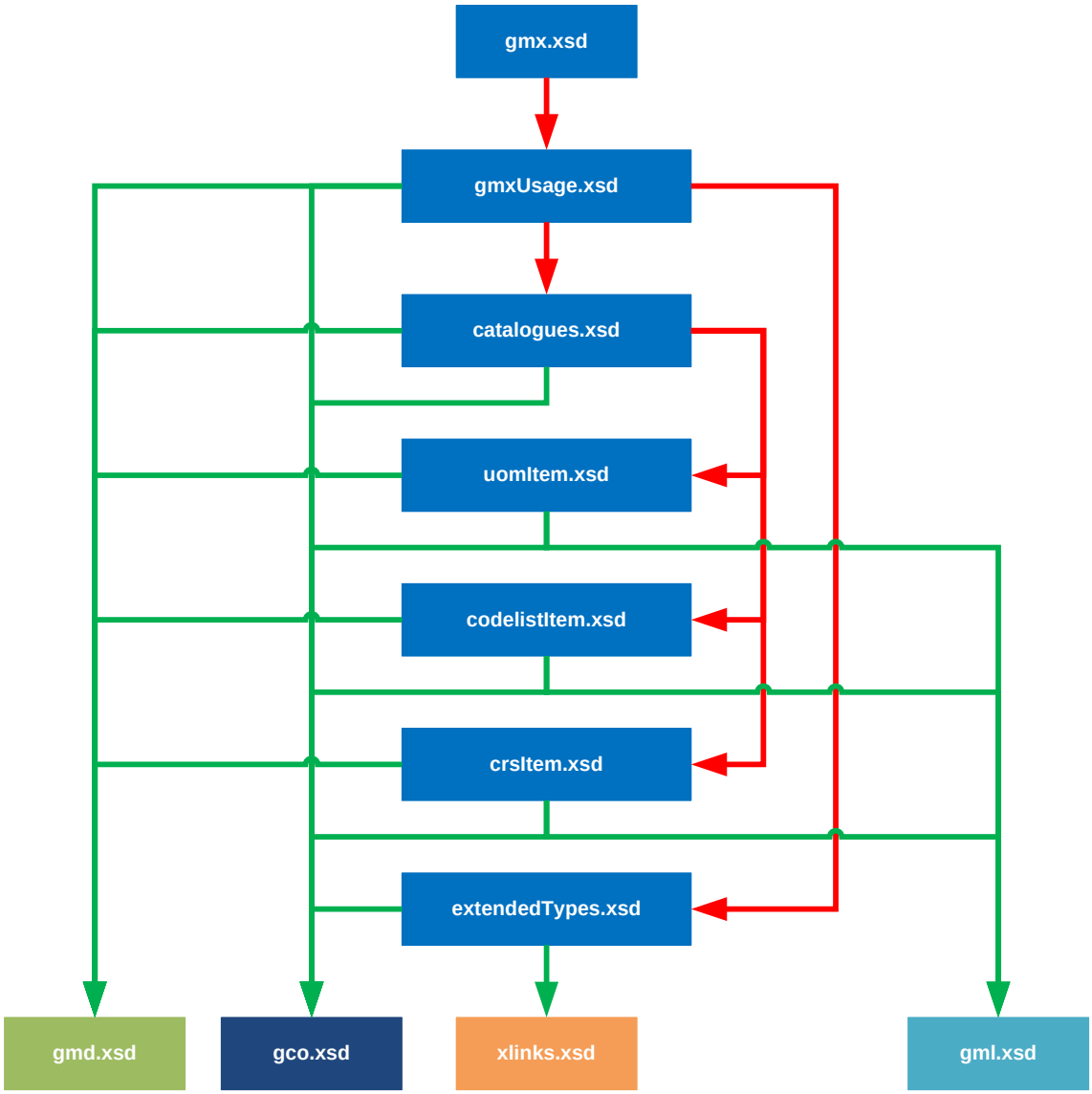

Figure 5 – Namespace gmx – Schema Relationships

Namespace gsr

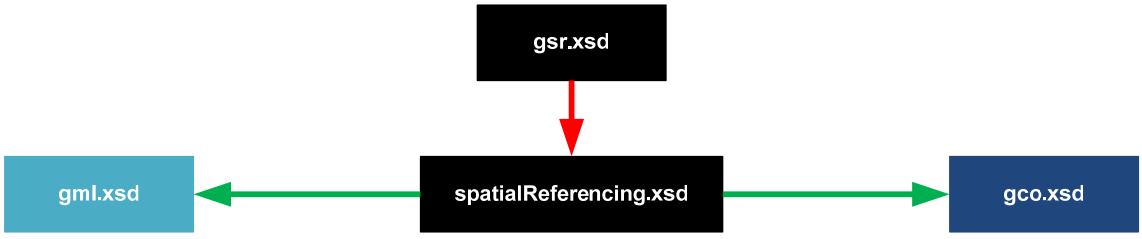

Figure 6 – Namespace gsr – Schema Relationships

Namespace gss

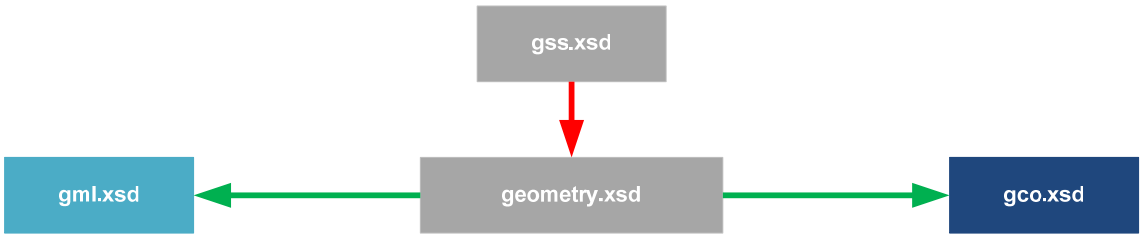

Figure 7 – Namespace gss – Schema Relationships

Namespace gts

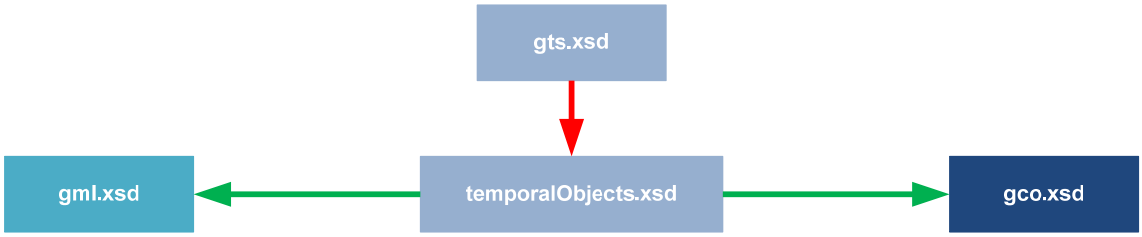

Figure 8 – Namespace gts – Schema Relationships

Namespace srv

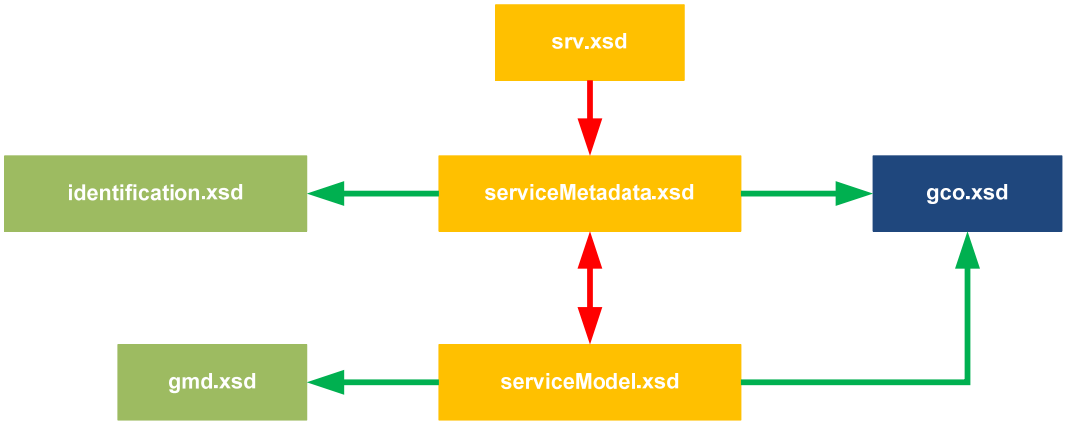

Figure 9 – Namespace srv – Schema Relationships