Version: 1.1

Authors: E. Indarto (DANS), W. Steinhoff (DANS)

Reviewers: I. Hofsink (KB), R. Voorburg (KB), M. Braaksma (DANS)

Date Modified: 2014-03-06

Status: Definitief

Version:1.2

Change log: Changed URN:NBN regex.

Author: Wilko Steinhoff (DANS)

Date: 2015-12-10

# DIDL EduStandaard Normalization

All DIDL records are validated against the DIDL xml-schema before the normalization process described in this document is started. The DIDL xml-schema is a local copy of the schema described in EduStandaard DIDL afspraken[[1]](#endnote-1). In case a DIDL record does not validate against the schema, an exception is thrown and no normalization will take place.

# Root element: root element

//DIDL

The root element is ‘templated’; it is normalized to the following template:

<didl:DIDL   
 xmlns:didl="urn:mpeg:mpeg21:2002:02-DIDL-NS"   
 xmlns:dii="urn:mpeg:mpeg21:2002:01-DII-NS"   
 xmlns:dc="http://purl.org/dc/elements/1.1/"   
 xmlns:dcterms="http://purl.org/dc/terms/"   
 xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"   
 xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"   
 xsi:schemaLocation="urn:mpeg:mpeg21:2002:02-DIDL-NS   
 http://standards.iso.org/ittf/PubliclyAvailableStandards/MPEG-21\_schema\_files/did/didl.xsd  
 urn:mpeg:mpeg21:2002:01-DII-NS   
 http://standards.iso.org/ittf/PubliclyAvailableStandards/MPEG-21\_schema\_files/dii/dii.xsd">

After the complete normalization has finished, all unused namespace declarations are removed from the document and multiple equal declarations are removed (except for the first one).

Logging: None

Exception: None

Cardinality: 1::1

# Root element: DIDLDocumentId

//DIDL/@DIDLDocumentId

The DIDLDocumentId attribute is removed from the DIDL root element.

This attribute is ‘depricated’ within EduStandaard.

Logging: None

Exception: None

Cardinality: 0::0

# Top level Item

//DIDL/Item

The toplevel Item is ‘templated’. The next four items are collected from the source and ‘transferred’ into the following template:

<didl:Item>  
 <didl:Descriptor>  
 <didl:Statement mimeType="application/xml">  
 <dii:Identifier>%PIdentifier%</dii:Identifier>  
 </didl:Statement>  
 </didl:Descriptor>  
 <didl:Descriptor>  
 <didl:Statement mimeType="application/xml">  
 <dcterms:modified>%DateModified%</dcterms:modified>  
 </didl:Statement>  
 </didl:Descriptor>  
 <didl:Component>  
 <didl:Resource mimeType="%MimeType%" ref="%ResourceURI%"/>  
 </didl:Component>

### (persistent) Identifier

//DIDL/Item/Descriptor/Statement/dii:Identifier/text()

Persistent identifier format is checked against the ‘urn:nbn: regex[[2]](#endnote-2).

Template: %PIdentifier%

Logging: None

Exception: Yes, if it does not exist or does not match the regex.

Cardinality: 1::1

### (date) Modified

//DIDL/Item/Descriptor/Statement/dcterms:modified/text()

Normalized to most recent date found in all Item (Child) elements within the DIDL node:

//Item/Descriptor/Statement/dcterms:modified/text()

Template: %DateModified%

Logging: If replaced with more recent child Item date.

Exception: Yes, if initial value does not exist or it cannot be validated against ISO8601

Cardinality: 1::1

### MIME type

//DIDL/Item/Component/Resource/@mimeType

Normalization is carried out on best-effort basis: The value is trimmed/stripped from preceding/tailing spaces and matched against a ‘MIME media-types’ regex[[3]](#endnote-3), which is based upon

RFC2045[[4]](#endnote-4).

Template: %MimeType%

Logging: If it does not match the ‘MIME media-types’ regexiii.

Exception: None (DIDL schema validation error if not supplied)

Cardinality: 1::1

### Resource

//DIDL/Item/Component/Resource/@ref

The resource URL is normalized into the @ref attribute of the <resource> tag.

The first value found from the following locations is used, if it matches the URL regex[[5]](#endnote-5):

1. '//DIDL/Item/Component/Resource/@ref',
2. '//DIDL/Item/Component/Resource/text()',
3. '//Item/Item[Descriptor/Statement/rdf:type/@rdf:resource="info:eu-repo/semantics/humanStartPage"]/Component/Resource/@ref',
4. '//Item/Item[Descriptor/Statement/rdf:type/@resource="info:eu-repo/semantics/humanStartPage"]/Component/Resource/@ref'
5. '//Item/Item[Descriptor/Statement/dip:ObjectType/text()="info:eu-repo/semantics/humanStartPage"]/Component/Resource/@ref',
6. '//Item/Item[Descriptor/Statement/rdf:type/@rdf:resource="info:eu-repo/semantics/objectFile"]/Component/Resource/@ref',
7. '//Item/Item[Descriptor/Statement/rdf:type/@resource="info:eu-repo/semantics/objectFile"]/Component/Resource/@ref'
8. '//Item/Item[Descriptor/Statement/dip:ObjectType/text()="info:eu-repo/semantics/objectFile"]/Component/Resource/@ref'

Template: %ResourceURI%

Logging: None

Exception: Yes, if none is found, or does not match the URL regexv.

Cardinality: 1::1

Example of top level Item:

<didl:Item>  
 <didl:Descriptor>  
 <didl:Statement mimeType="application/xml">  
 <dii:Identifier>urn:nbn:nl:ui:28-98532222333</dii:Identifier>  
 </didl:Statement>  
 </didl:Descriptor>  
 <didl:Descriptor>  
 <didl:Statement mimeType="application/xml">  
 <dcterms:modified>2011-08-28T13:51:55Z</dcterms:modified>  
 </didl:Statement>  
 </didl:Descriptor>  
 <didl:Component>  
 <didl:Resource mimeType="text/html" ref="http://purl.utwente.nl/publications/9853"/>  
 </didl:Component>

# Second level Items

## DescriptiveMetadata Item node

This child Item-node is ‘templated’.

The first value found from the following locations is used:

1. //DIDL/Item/Item[Descriptor/Statement/rdf:type/@rdf:resource="info:eu-repo/semantics/descriptiveMetadata]
2. //DIDL/Item/Item[Descriptor/Statement/rdf:type/@resource="info:eu-repo/semantics/descriptiveMetadata"
3. //DIDL/Item/Item[Descriptor/Statement/dip:ObjectType/text()="info:eu-repo/semantics/descriptiveMetadata"]'

<didl:Item>  
 <didl:Descriptor>  
 <didl:Statement mimeType="application/xml">  
 <rdf:type rdf:resource="info:eu-repo/semantics/descriptiveMetadata"/>  
 </didl:Statement>  
 </didl:Descriptor>  
 %IdentifierDescriptor%

%DateModifiedDescriptor%

<didl:Component>  
 <didl:Resource mimeType="application/xml">  
 %Metadata%   
 </didl:Resource>  
 </didl:Component>  
</didl:Item>

Logging: If ‘descriptiveMetadata’ is not found in the first (1) location/xpath listed above (EduStandaard).

Exception: If no descriptiveMetadata Item is found at all.

Cardinality: 1::1

### Metadata (MODS)

self::Item//mods:mods

No normalization.

Template: %Metadata%

Logging: None

Exception: Yes, if no MODS is found within a DIDL descriptiveMetadata Item.

### DateModifiedDescriptor

self::Item/Descriptor/Statement/dcterms:modified/text()

If an empty value or not a valid ISO8601 date is found within the descriptiveMetadata Item, the complete dateModifiedDescriptor (<dcterms:modified>) is removed.

Template: %DateModifiedDescriptor%

Logging: If the descriptor is removed (empty value or not a valid ISO8601 date).

Exception: None

### IdentifierDescriptor

self::Item/Descriptor/Statement/dii:Identifier/text()

If an identifier is found within the descriptiveMetadata Item, the (stripped/trimmed) identifier (<dii:Identifier>) is templated into the DescriptiveMetadataItem.

Template: %IdentifierDescriptor%

Logging: None

Exception: None

## ObjectFiles Item nodes

Search for objectFiles:

The first match from the following locations is processed:

1. //didl:DIDL/didl:Item/didl:Item[didl:Descriptor/didl:Statement/rdf:type/@rdf:resource="info:eu-repo/semantics/objectFile”]
2. //didl:DIDL/didl:Item/didl:Item[didl:Descriptor/didl:Statement/rdf:type/@resource="info:eu-repo/semantics/objectFile”]
3. //didl:DIDL/didl:Item/didl:Item[didl:Descriptor/didl:Statement/dip:ObjectType/text()="info:eu-repo/semantics/objectFile”]

Each ‘objectFile’ Item node found is templated into the following template:

<didl:Item>  
 <didl:Descriptor>  
 <didl:Statement mimeType="application/xml">  
 <rdf:type rdf:resource="info:eu-repo/semantics/objectFile"/>  
 </didl:Statement>  
 </didl:Descriptor>

%DidlDescriptorTemplate% | %DIDLComponentTemplate%  
</didl:Item>

Where %DidlDescriptorTemplate% is again a template:

<didl:Descriptor>  
 <didl:Statement mimeType="application/xml">  
 %DescriptorTemplate%  
 </didl:Statement>  
 </didl:Descriptor>

and %DIDLComponentTemplate%:

<didl:Component>  
 <didl:Resource mimeType="%MimeType%" ref="%refURL%"/>  
</didl:Component>

This template is created for each of the following Descriptor elements found in a second level ObjectFile Item:

### Identifier

Value is stripped/trimmed.

self::Item/Descriptor/Statement/dii:Identifier/text()

Template: %DescriptorTemplate%

Logger: None

Exception: None

Cardinality: 0::1

Example:

<dii:Identifier>urn:NBN:nl:ui:15-1765/40473</dii:Identifier>

### AccessRights

self::Item/Descriptor/Statement/dcterms:accessRights/text()

If the accessRights[[6]](#endnote-6) value contains one of the following keys, it is normalized to it’s fully qualified name:

* ‘openaccess’ to: 'http://purl.org/eprint/accessRights/OpenAccess'
* ‘restrictedaccess’ to: 'http://purl.org/eprint/accessRights/RestrictedAccess'
* ‘closedacces’ to: 'http://purl.org/eprint/accessRights/ClosedAccess'

Template: %DescriptorTemplate%

Logger: None

Exception: If no accessRights are found, or it cannot be normalized.

Cardinality: 1::1 (EduStandaard)

Example:

<dcterms:accessRights>http://purl.org/eprint/accessRights/OpenAccess</dcterms:accessRights>

### (date) Modified

self::Item/Descriptor/Statement/dcterms:modified/text()

Validate against ISO8601, if not valid, the Descriptor is skipped/removed.

Template: %DescriptorTemplate%

Logger: None

Exception: None

Cardinality: 0::1

Example:

<dcterms:modified>2007-03-06T12:44:43Z</dcterms:modified>

### Description

self::Item/Descriptor/Statement/dc:description/text()

Template: %DescriptorTemplate%

Logger: None

Exception: None

Cardinality: 0::1

Example:

<dc:description>description example</dc:description>

### Version type (author/published)

self::Item/Descriptor/Statement/rdf:type/@rdf:resource

If the value contains one of the following keys (EduStandaard Semanticsvi), it is normalized to its fully qualified name.

* 'publishedversion' to: 'info:eu-repo/semantics/publishedVersion'
* 'updatedversion' to: 'info:eu-repo/semantics/updatedVersion'
* 'acceptedversion' to: 'info:eu-repo/semantics/acceptedVersion'
* 'submittedversion' to: 'info:eu-repo/semantics/submittedVersion'
* 'draft' to: 'info:eu-repo/semantics/draft'

Template: %DescriptorTemplate%

Logger: None

Exception: None

Cardinality: 0::1

Example:

<rdf:type rdf:resource="info:eu-repo/semantics/authorVersion"/>

### MimeType

self::Item/Component/Resource/@mimeType

Template:%DIDLComponentTemplate%, value: %MimeType%

Logging: If it does not match the ‘MIME Media-Types’ regexiii.

Exception: None (DIDL schema validation error if not supplied)

Cardinality: 1::1

See also: MimeType of toplevel Item.

### Resource

self::Item/Component/Resource/@ref

Each Resource/@ref is checked for a valid URLv.

Template:%DIDLComponentTemplate%, value: %RefURL%

Logging: None

Exception: Yes, if none is found, or does not match the URL regexv.

Cardinality: 1::1

### Multiple Resources

If more than one valid resources (mimetype and url) are found within one objectfile Item, normalization is done by duplicating each resource in a separate Item tag.

The entire normalized Item node is copied for each resource tag.

Logger: None

Cardinality: 1::\*

Example:

<didl:Component>  
 <didl:Resource mimeType="application/doc" ref="http://doc.utwente.nl/9853/1/hoes96synthesis.doc"/>

<didl:Resource mimeType="application/pdf" ref="http://doc.utwente.nl/9853/1/hoes96synthesis.pdf"/>  
</didl:Component>

Normalization result:

<didl:Item>  
 <didl:Descriptor>  
 <didl:Statement mimeType="application/xml">  
 <rdf:type rdf:resource="info:eu-repo/semantics/objectFile"/>  
 </didl:Statement>  
 </didl:Descriptor>  
 <didl:Descriptor>  
 <didl:Statement mimeType="application/xml">  
 <dcterms:modified>2009-01-13T15:52:09Z</dcterms:modified>  
 </didl:Statement>  
 </didl:Descriptor>  
 <didl:Component>  
 <didl:Resource mimeType="application/doc" ref="http://depot.knaw.nl/566/1/ hoes96synthesis.doc"/>  
 </didl:Component>  
 </didl:Item>

<didl:Item>  
 <didl:Descriptor>  
 <didl:Statement mimeType="application/xml">  
 <rdf:type rdf:resource="info:eu-repo/semantics/objectFile"/>  
 </didl:Statement>  
 </didl:Descriptor>  
 <didl:Descriptor>  
 <didl:Statement mimeType="application/xml">  
 <dcterms:modified>2009-01-13T15:52:09Z</dcterms:modified>  
 </didl:Statement>  
 </didl:Descriptor>  
 <didl:Component>  
 <didl:Resource mimeType="application/pdf" ref="http://depot.knaw.nl/566/1/ hoes96synthesis.doc"/>  
 </didl:Component>  
 </didl:Item>

Logger: None

Exception: None

Cardinality: 0::\*

## HumanStartPage/jump-off item-node

This Item node is ‘templated’.

<didl:Item>  
 <didl:Descriptor>  
 <didl:Statement mimeType="application/xml">  
 <rdf:type rdf:resource="info:eu-repo/semantics/humanStartPage"/>  
 </didl:Statement>  
 </didl:Descriptor>  
 <didl:Component>  
 <didl:Resource ref="%ResourceURI%" mimeType="%MimeType%"/>  
 </didl:Component>  
</didl:Item>

The (first) HSP Item node is searched for at the following locations:

1. '//Item/Item[Descriptor/Statement/rdf:type/@rdf:resource="info:eu-repo/semantics/humanStartPage"]’
2. '//Item/Item[Descriptor/Statement/rdf:type/@resource="info:eu-repo/semantics/humanStartPage"]’
3. '//Item/Item[Descriptor/Statement/dip:ObjectType/text()="info:eu-repo/semantics/humanStartPage"]’

Logging: If no HSP is found in the locations mentioned. (Werkafspraak KB met repositories).

Exception: None

Cardinality: 0::1

### ResourceURI

self::Item/Component/Resource/@ref

* Each Resource/@ref is checked for a valid URLv.

If no valid URL is available from the HSP Item-node found above, or in one of the following locations, the complete HSP Item node is removed.

1. '//Item/Item/Component/Resource/@ref
2. '//Item/Item[Descriptor/Statement/rdf:type/@resource="info:eu-repo/semantics/humanStartPage"]/Component/Resource/@ref'
3. '//Item/Item[Descriptor/Statement/dip:ObjectType/text()="info:eu-repo/semantics/humanStartPage"]/Component/Resource/@ref'

Template: %ResourceURI%

Logging: None

Exception: If no valid URL is found

Cardinality: 1::1

### MIME type

self::Item/Component/Resource/@mimeType

Template: %MimeType%

Logging: If it does not match the ‘MIME Media-Types’ regexiii.

Exception: None (DIDL schemavalidation error if not supplied)

Cardinality: 1::1

1. EduStandaard: [WG Metadata, Afspraak 'DIDL'](http://www.edustandaard.nl/afspraken/beheerde-afspraken/didl/) [↑](#endnote-ref-1)
2. ^[uU][rR][nN]:[nN][bB][nN]:[nN][lL](:([a-zA-Z]{2}))?:\d{2}-.+ [↑](#endnote-ref-2)
3. ^(application|audio|example|image|message|model|multipart|text|video)/.{2,}$

   The MIME Media-Types are taken from [IANA](http://www.iana.org/assignments/media-types/media-types.xhtml) [↑](#endnote-ref-3)
4. <http://tools.ietf.org/html/rfc2045> [↑](#endnote-ref-4)
5. (^(?:http|ftp|sftp)s?://(?:(?:[A-Z0-9](?:[A-Z0-9-]{0,61}[A-Z0-9])?\.)+(?:[A-Z]{2,6}\.?|[A-Z0-9-]{2,}\.?)|\d{1,3}\.\d{1,3}\.\d{1,3}\.\d{1,3})(?::\d+)?(?:/?|[/?]\S+)$, IGNORECASE) [↑](#endnote-ref-5)
6. EduStandaard: [WG Metadata, Afspraak 'Semantics'](http://www.edustandaard.nl/afspraken/beheerde-afspraken/semantics/) [↑](#endnote-ref-6)