

Some insights into MPEG-21 DID / DIDL based on the aDORe experience

Herbert Van de Sompel

`herbertv@lanl.gov`

`http://public.lanl.gov/herbertv`

Digital Library Research & Prototyping Team
Research Library
Los Alamos National Laboratory



MPEG-21 DID / DIDL
OAI5, CERN, Switzerland, April 18th 2007
Herbert Van de Sompel

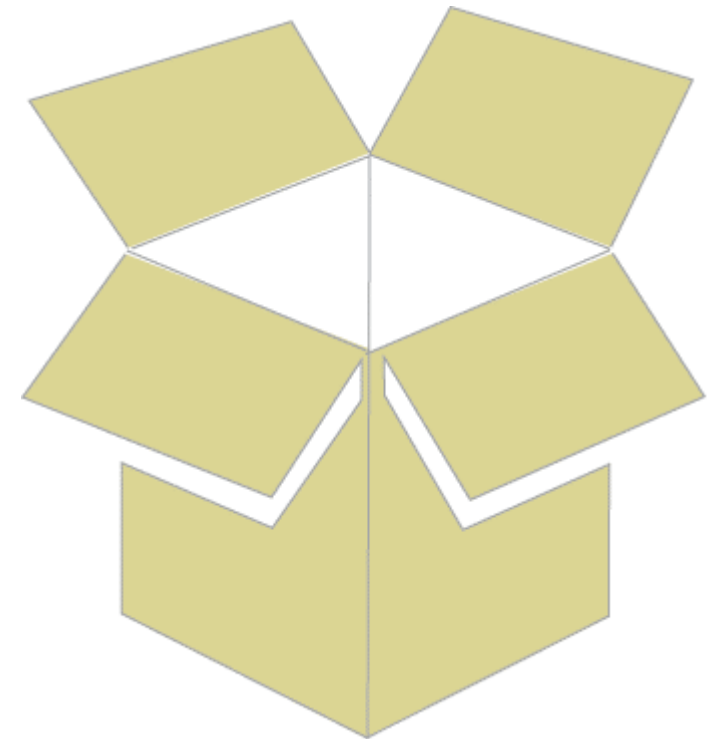
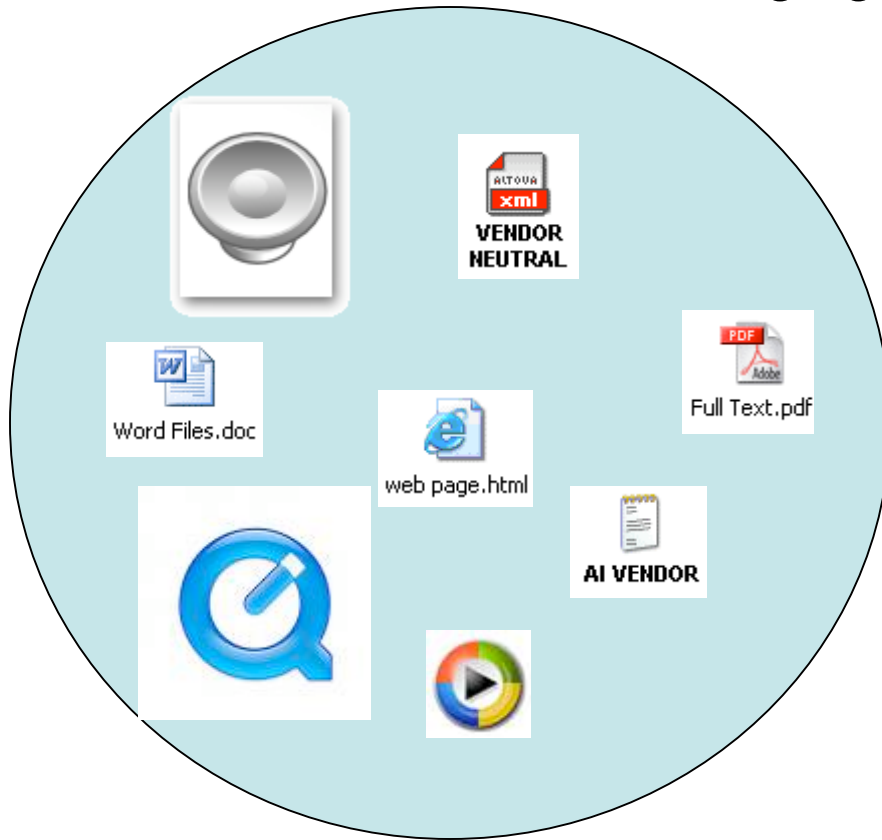


Some context

Background: LANL Research Library

- Locally load databases and electronic journals.
 - 100,000,000+ objects
- Repository started in 2004.
- Based on aDORe repository architecture:
 - Standards-based: XML, XML Schema, MPEG-21 Digital Item Declaration, the MPEG-21 Digital Item Identification, URI, info URI, OAI-PMH, NISO OpenURL, SRU, Information Environment Service Registry, Internet Archive ARC file format, OAI concepts.
 - Component-based, modular; interaction between components is protocol-based.
 - A federation of autonomous repositories; a microcosm for studying cross-repository interoperability.
 - Highly scalable
- aDORe software:
 - The repository - aDORe Archive @ <http://african.lanl.gov/aDORe/>
 - New version on its way
 - Tools to represent objects - DIDLTools @ <http://african.lanl.gov/aDORe/projects/DIDLTools/>

Packaging Requirements



About MPEG-21 DID / DIDL

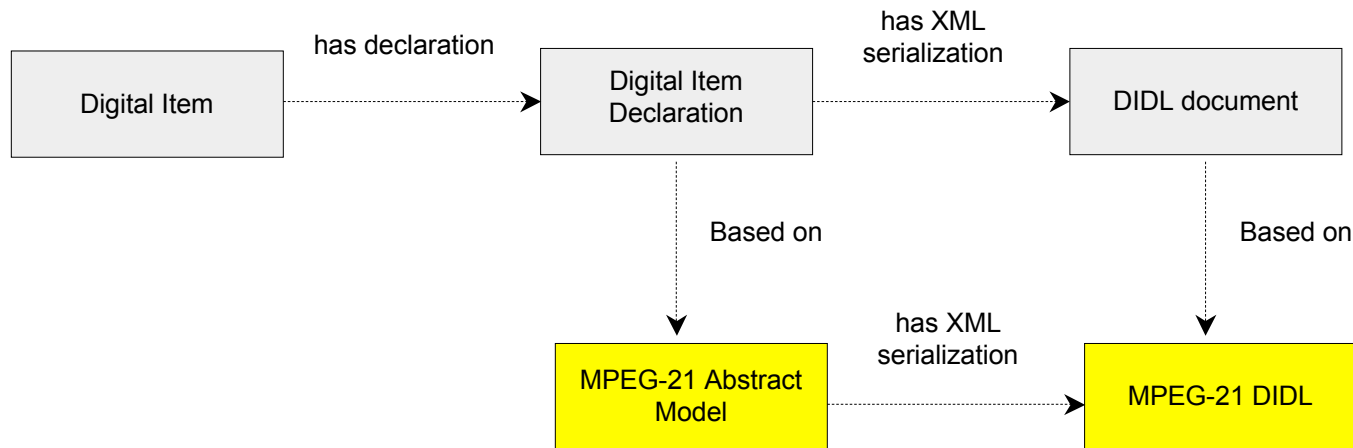
MPEG-21

- ISO/IEC TR 21000-2
- MPEG-21 is modular
 - Part 1: Vision, Technologies and Strategy
 - Part 2: DID – Digital Item Declaration
 - Part 3: DII – Digital Item Identification
 - Part 4: IPMP – Intellectual Property Management and Protection
 - Part 5: REL – Rights Expression
 - Part 6: RDD – Rights Data Dictionary
 - Part 7: DIA – Digital Item Adaptation
 - Part 8: MPEG-21 Reference Software
 - Part 9: MPEG-21 File Format
 - Part 10: DIP – Digital Item Processing
 - Part 11: Persistent Association Technology
 - Part 12: Test Bed for MPEG-21 Resource Delivery
 - Part 13: Empty
 - Part 14: Conformance
 - Part 15: Event Reporting
 - Part 16: BF – Binary Format
 - Part 17: Fragment Identification of MPEG Resources
 - Part 18: Digital Item Streaming

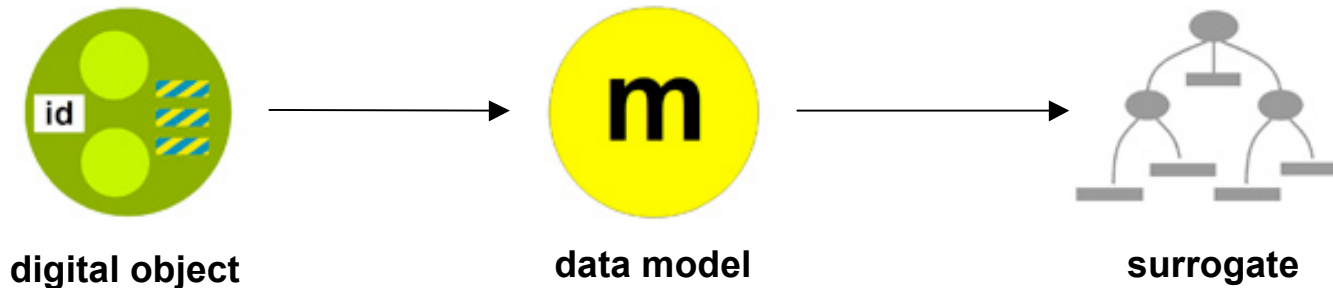
Core concepts - Digital Item & User

- Digital Item – “a structured digital object with a standard representation, identification and metadata within the MPEG-21 framework”
- User – “any entity that interacts with or makes use of the Digital Items”

MPEG-21 Digital Item Declaration & Digital Item Declaration Language



- Abstract model – MPEG-21 DID
- Instantiation of the model in XML – MPEG-21 DIDL ; DIDL definition expressed as W3C XML schema
- XML representation of a Digital Item based on DIDL - DIDL document



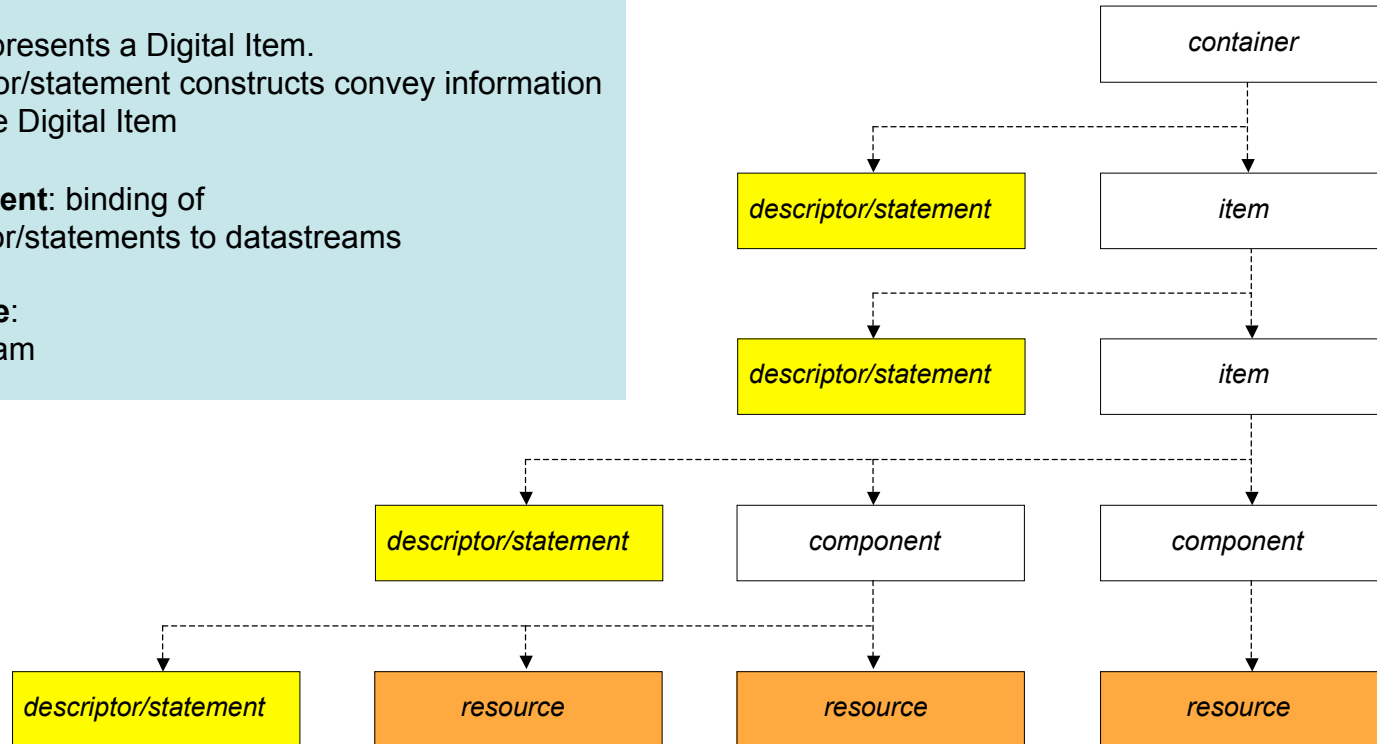
MPEG-21 DID Abstract Data Model: basic entities

container: grouping of items and descriptor/statement constructs pertaining to the container

item: represents a Digital Item. Descriptor/statement constructs convey information about the Digital Item

component: binding of descriptor/statements to datastreams

resource: datastream



Entities of MPEG-21 DID

Basic

- Container
- Item
- Component
- Descriptor
- Statement
- Resource

Additional

- Choice
- Selection
- Predicate
- Condition
- Annotation
- Assertion
- Anchor
- Fragment

Representing Digital Objects using MPEG-21 DID / DIDL



MPEG-21 DID / DIDL
OAI5, CERN, Switzerland, April 18th 2007
Herbert Van de Sompel



sample Digital Object

	Type	MIME	identifier
Digital Object	scholarly paper	N/A	DOI
Constituent datastream 1	metadata record	application/xml	PMID
Constituent datastream 2	fulltext file	application/pdf	–

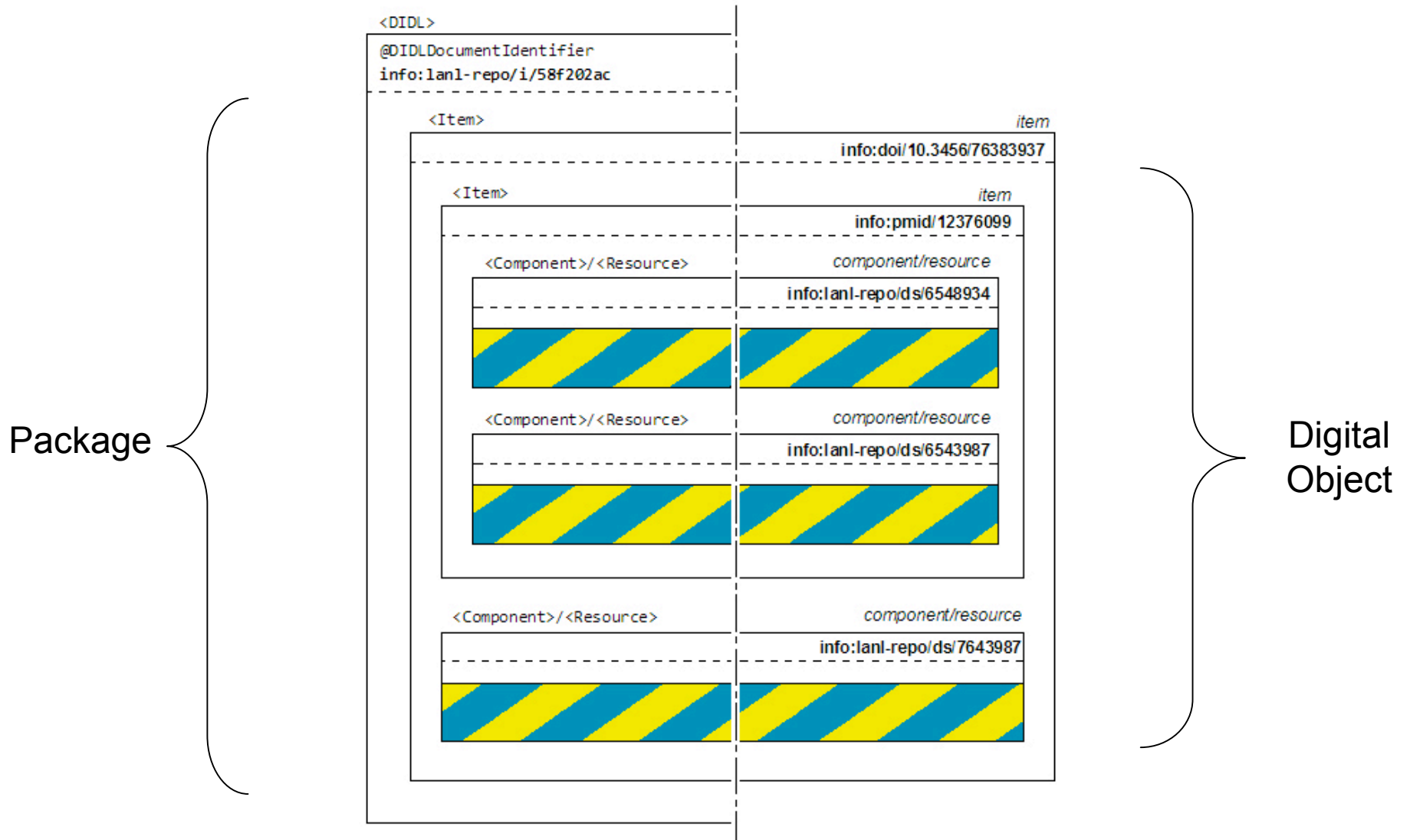
A Digital Object and its Package

- An XML Package is made available for every Digital Object
- The Package is an XML document compliant with the MPEG-21 Digital Item Declaration Language ~ DIDL document
- In aDORe, a DIDL document typically contains:
 - Related to the Digital Object represented by the DIDL document:
 - Datastreams:
 - By-Value: descriptive metadata datastream & ingest/repository related metadata
 - By-Reference: other constituent datastreams of the Digital Object
 - Descriptors to convey properties about the object and its datastreams:
 - MPEG-21 defined Descriptors (Digital Item Identifier)
 - LANL defined Descriptors (semantic type, format, creation date, ...)
 - Descriptors defined by others (JHOVE, XML Signatures, ...)
 - Related to the DIDL document itself:
 - Properties of the DIDL document including:
 - MPEG-21 defined identifier
 - LANL defined (creation/modification datetime)
 - Properties defined by others (e.g. XML Signature)

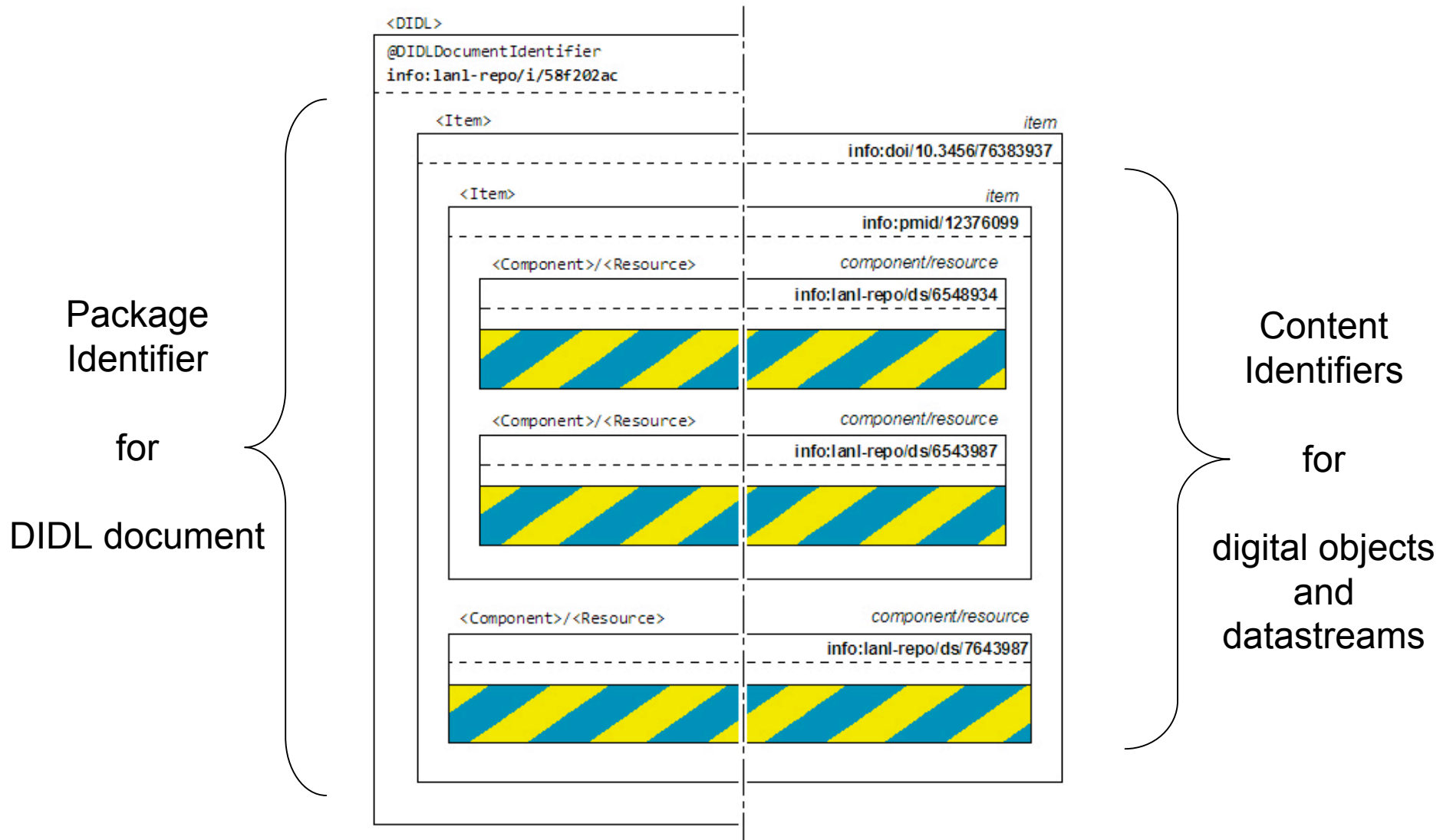
sample Digital Object

	Type	MIME	identifier
Digital Object	scholarly paper	N/A	DOI
Constituent Datastream 1	metadata record	application/xml	PMID
Constituent Datastream 2	fulltext file	application/pdf	–

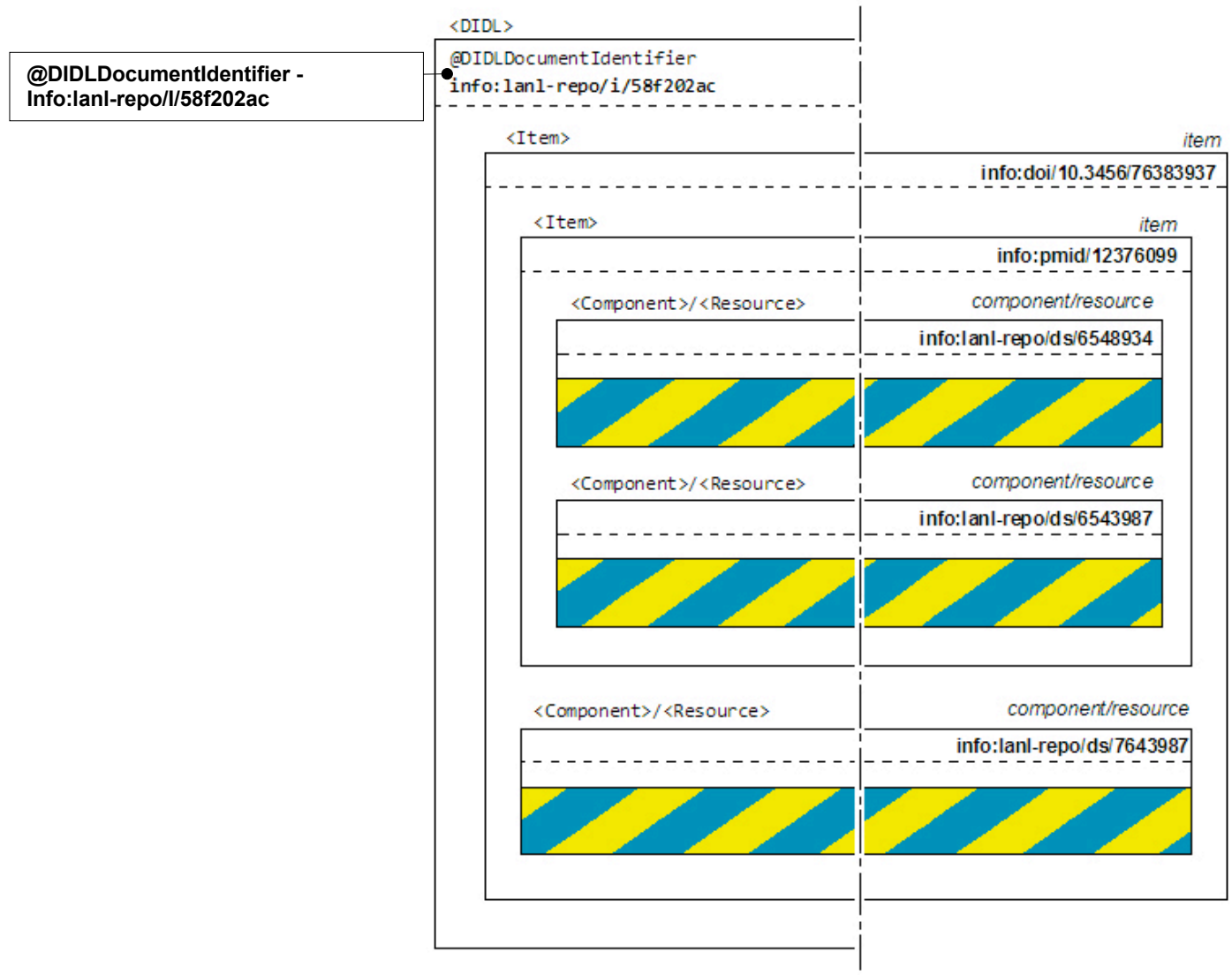
representing Digital Objects using MPEG-21 DID



Identification: digital objects, datastreams, DIDL documents

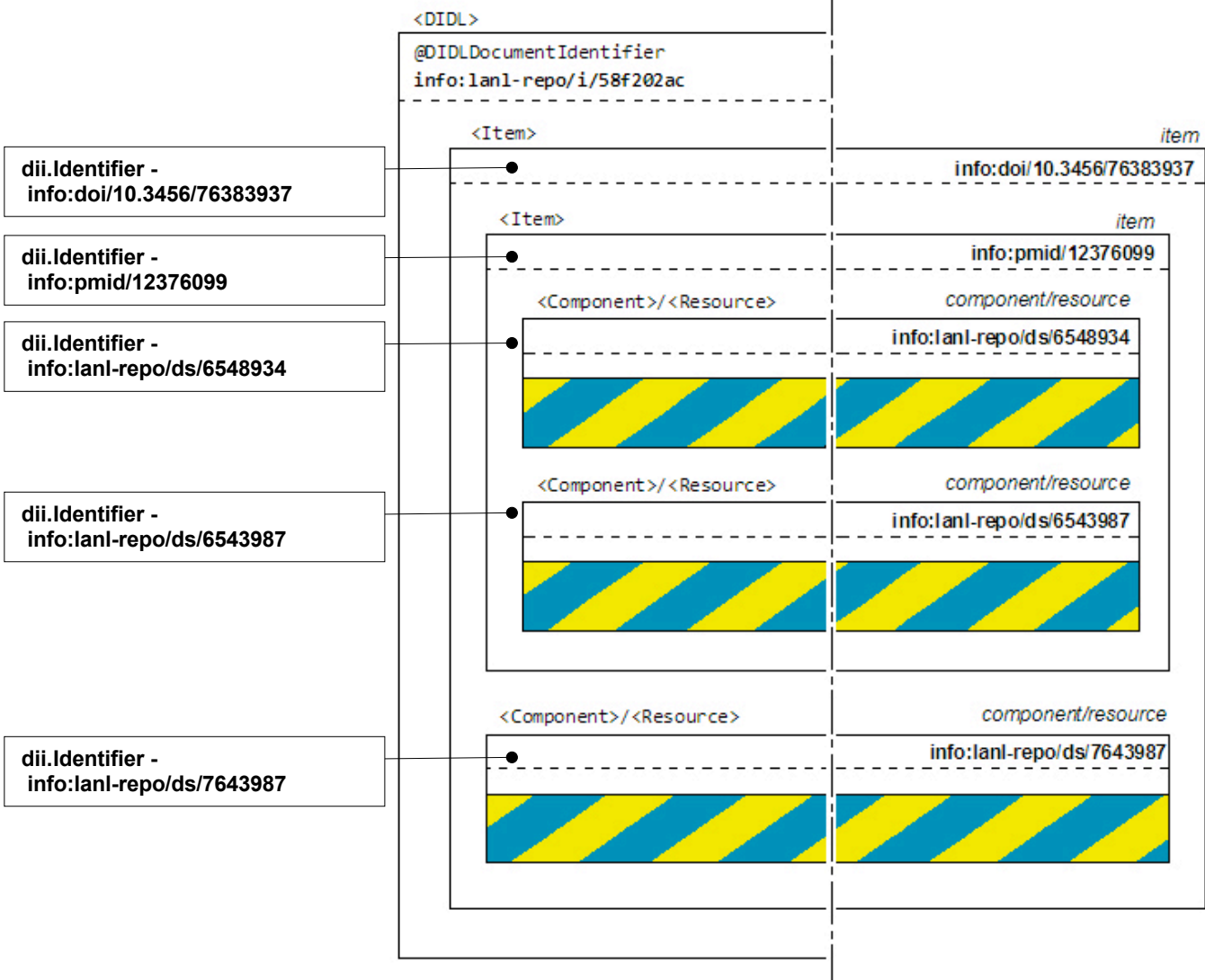


DIDL @DIDLDocumentIdentifier : Identification of DIDL document



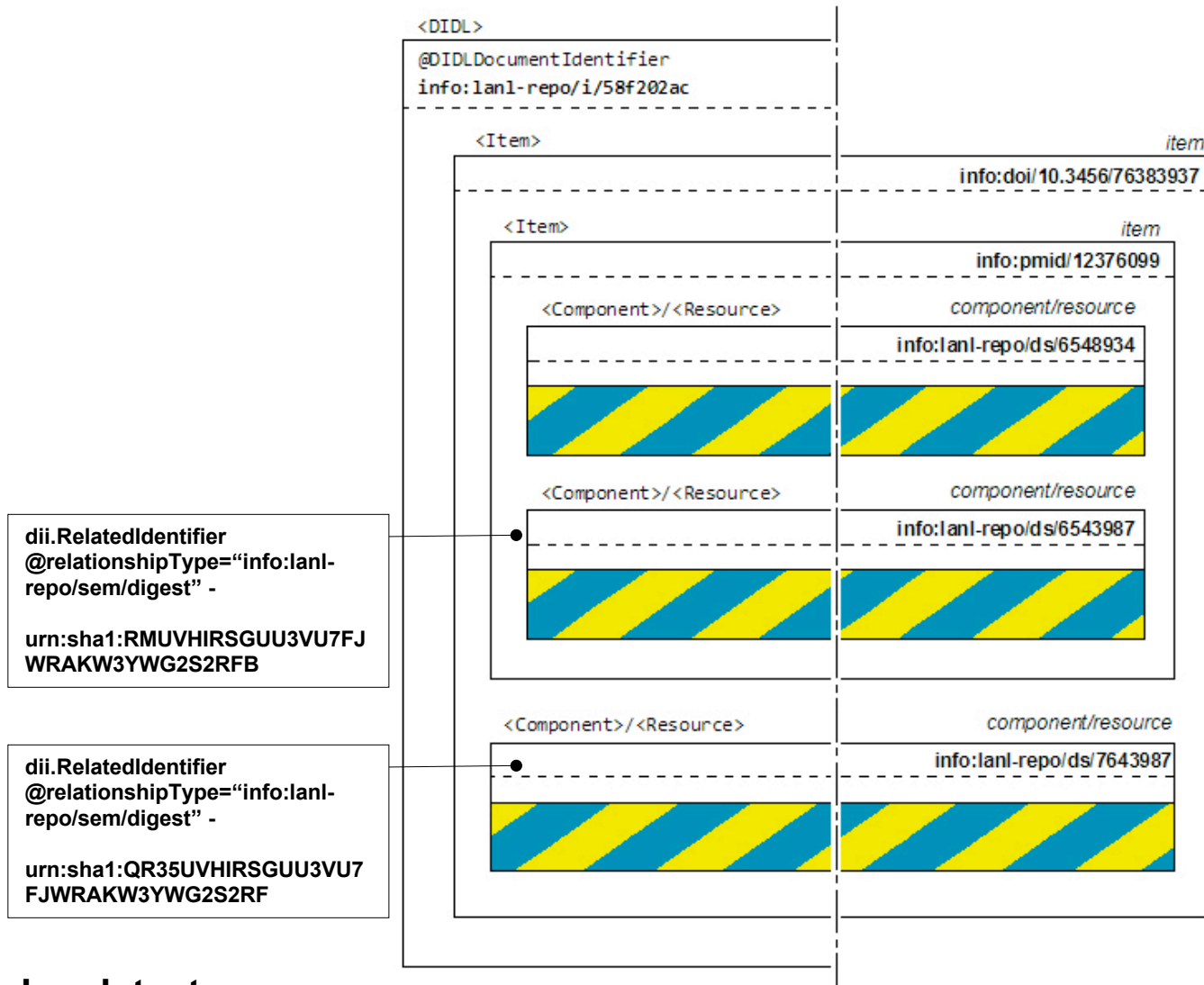
DIDL document identifier: info:lanl-repo/i/UUID

DII Identifier : Identification of digital objects and datastreams



digital object identifier: **anyURI**
 datastream identifier: `info:lanl-repo/ds/UUID`

DII RelatedIdentifier : digests for datastreams (content-based identification)



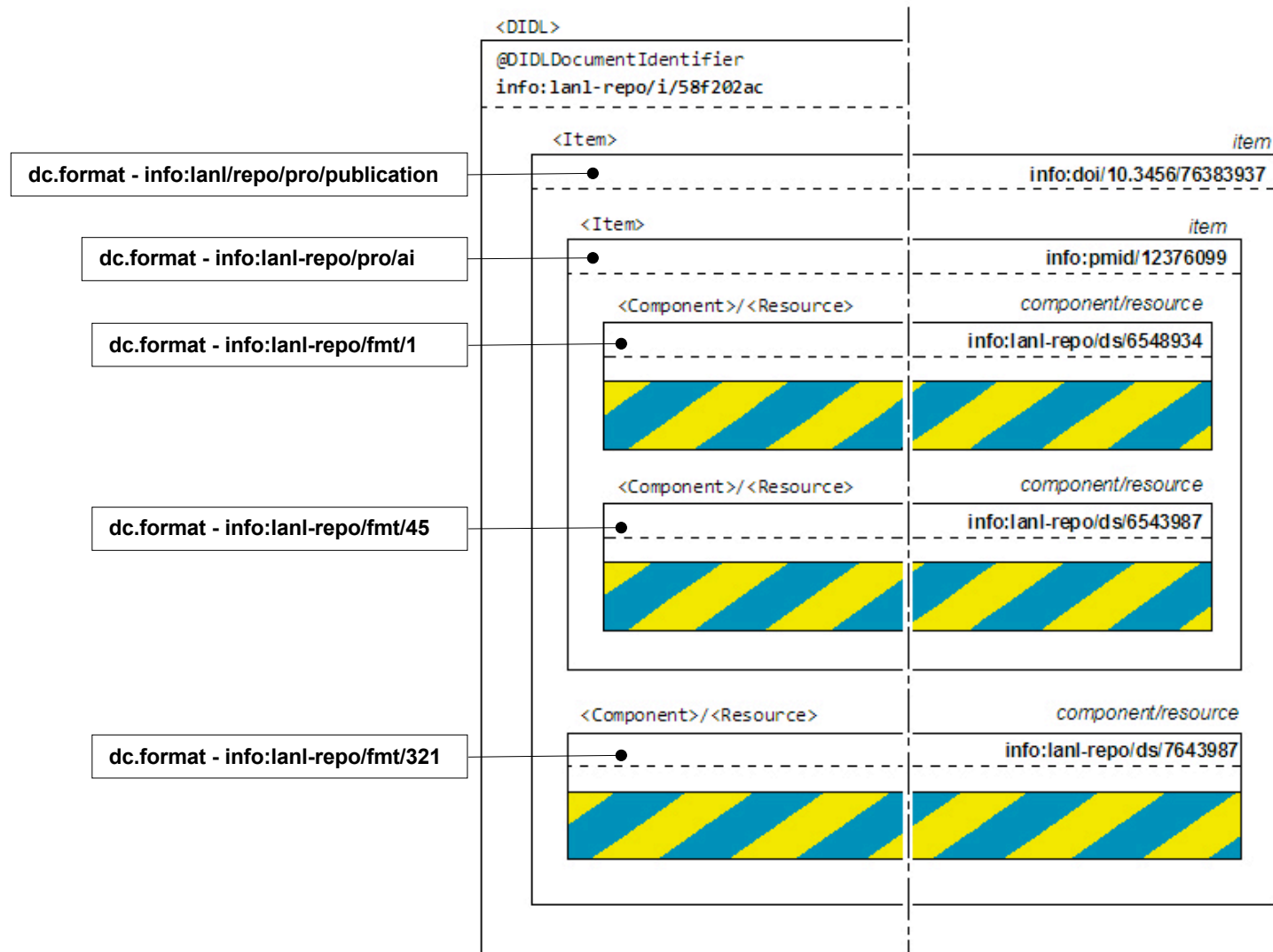
Not for by-value datastreams.
urn:sha1:sha1_digest_of_datastream

DIEXT @DIDLDocumentCreated : DIDL Creation Date



ISO 8601, seconds granularity, Z notation.

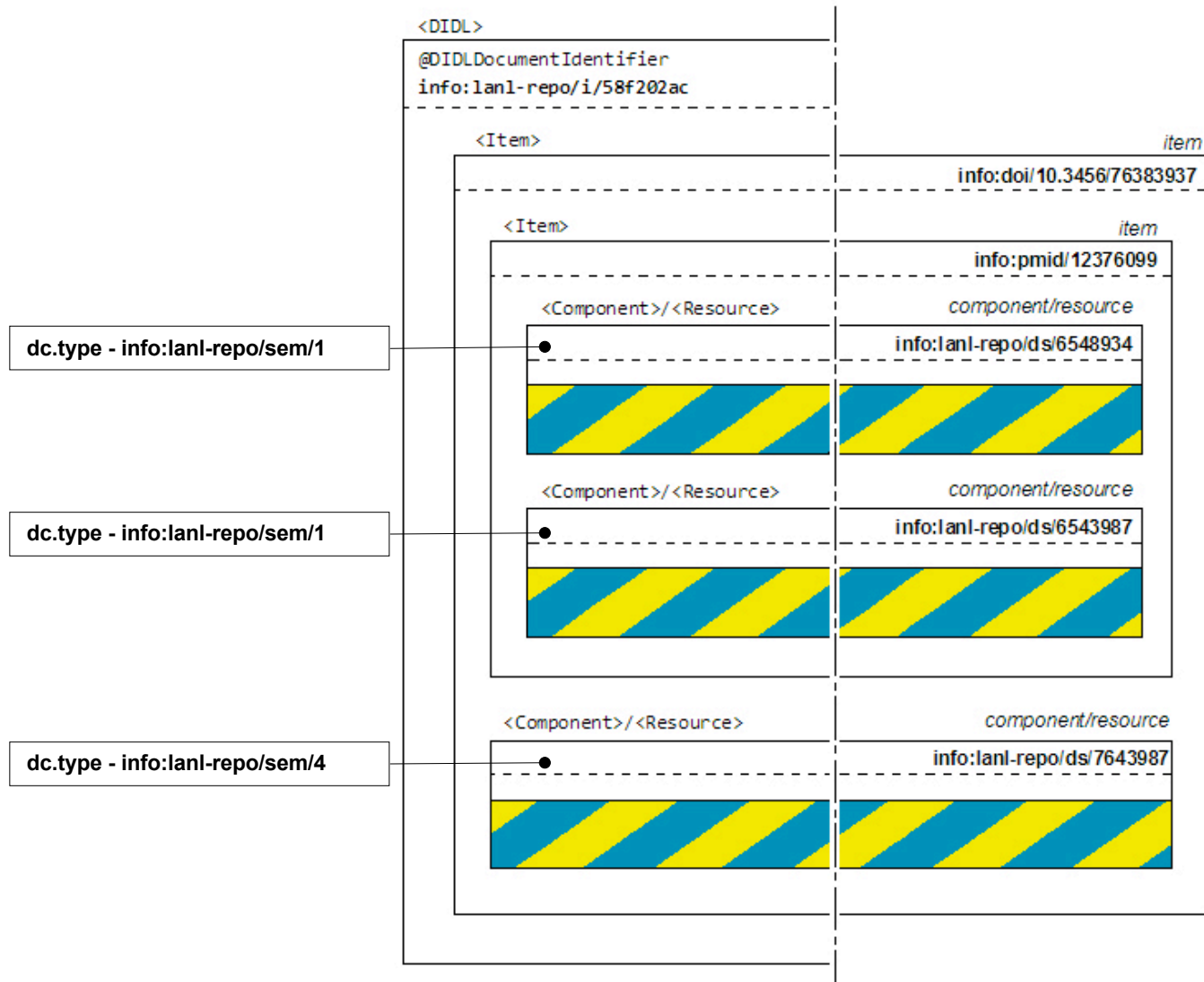
DIADM dc:format : profiles of digital objects and formats of datastream



digital object profile: info:lanl-repo/pro/string at Item level

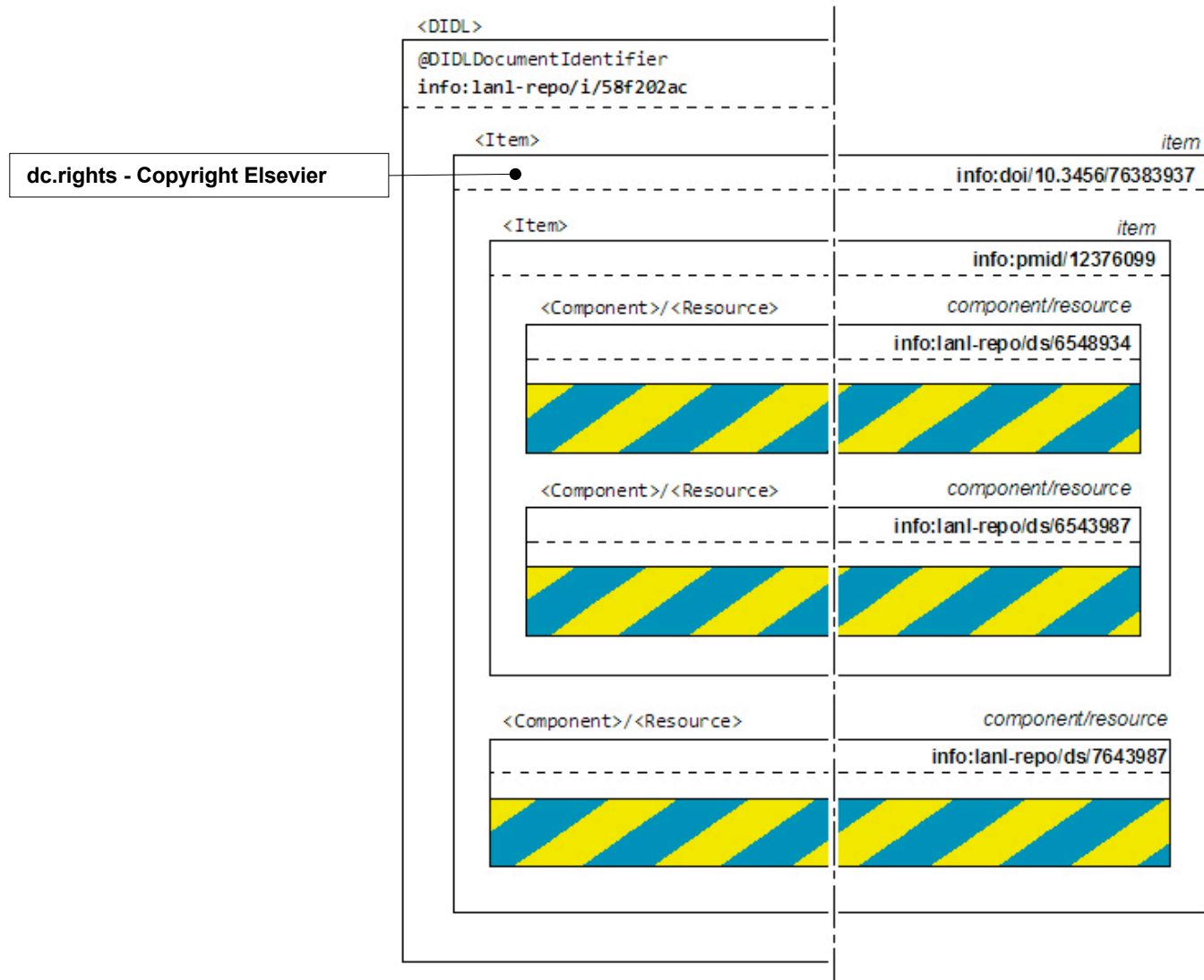
datastream media type: info:lanl-repo/fmt/integer at Component level

DIADM dc:type : datastream intellectual content type



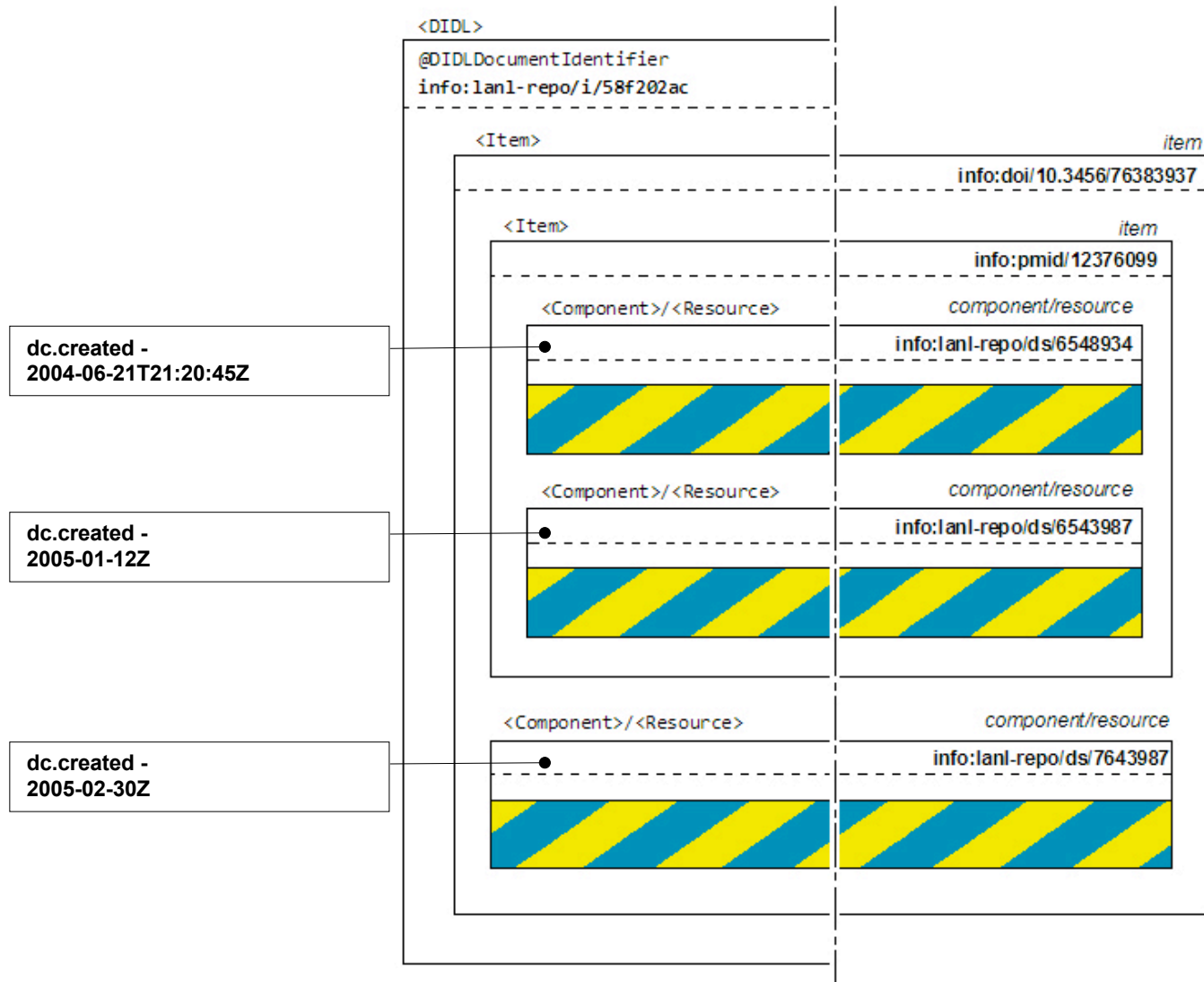
datastream semantic type: info:lanl-repo/**sem/integer** at Component level

DIADM dc:rights : copyright statements



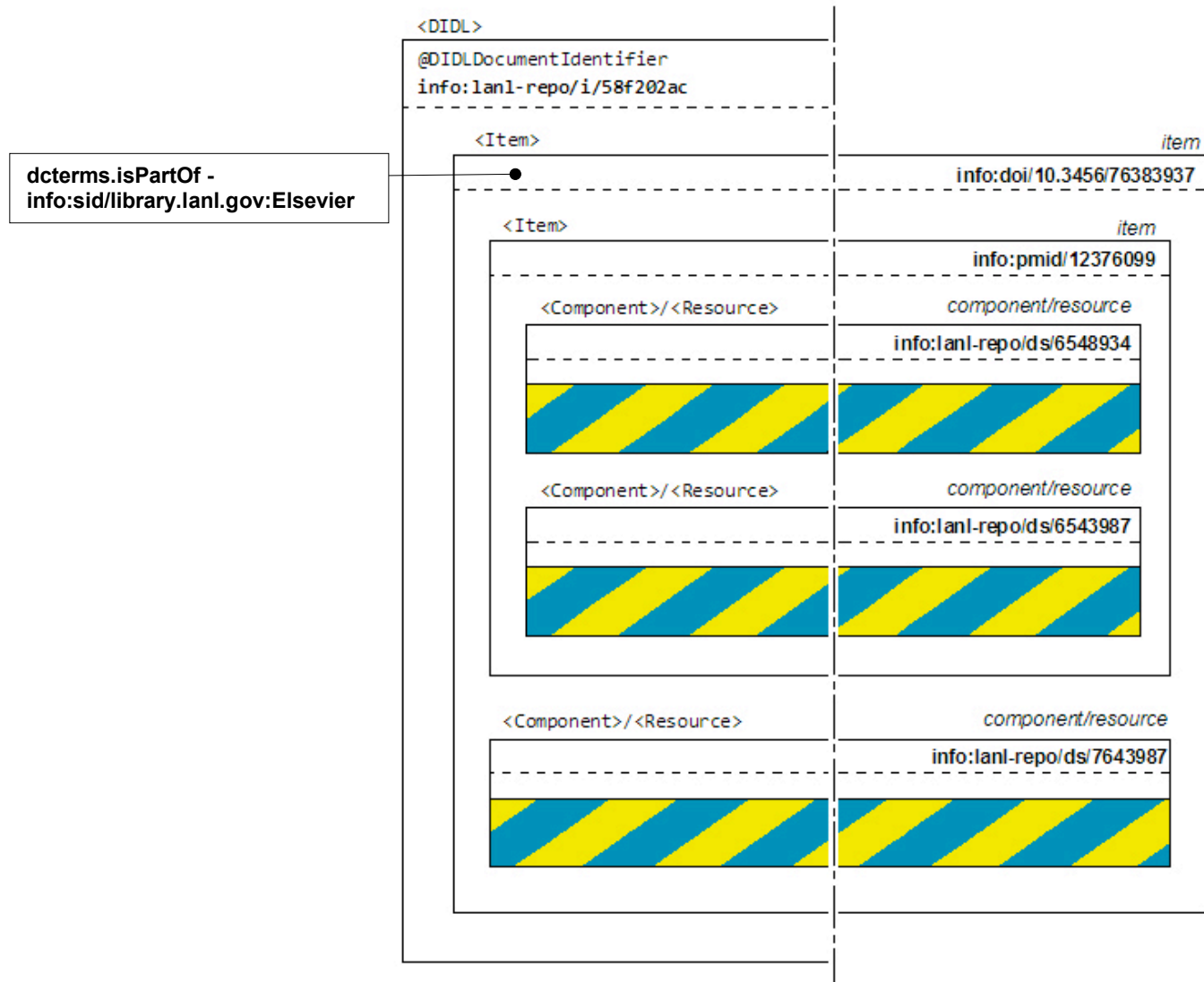
Can be Item and Component level.

DIADM dc:created : datastream creation datetime



ISO 8601, Z notation. Component level only.

DIADM dcterms:isPartOf : collection membership



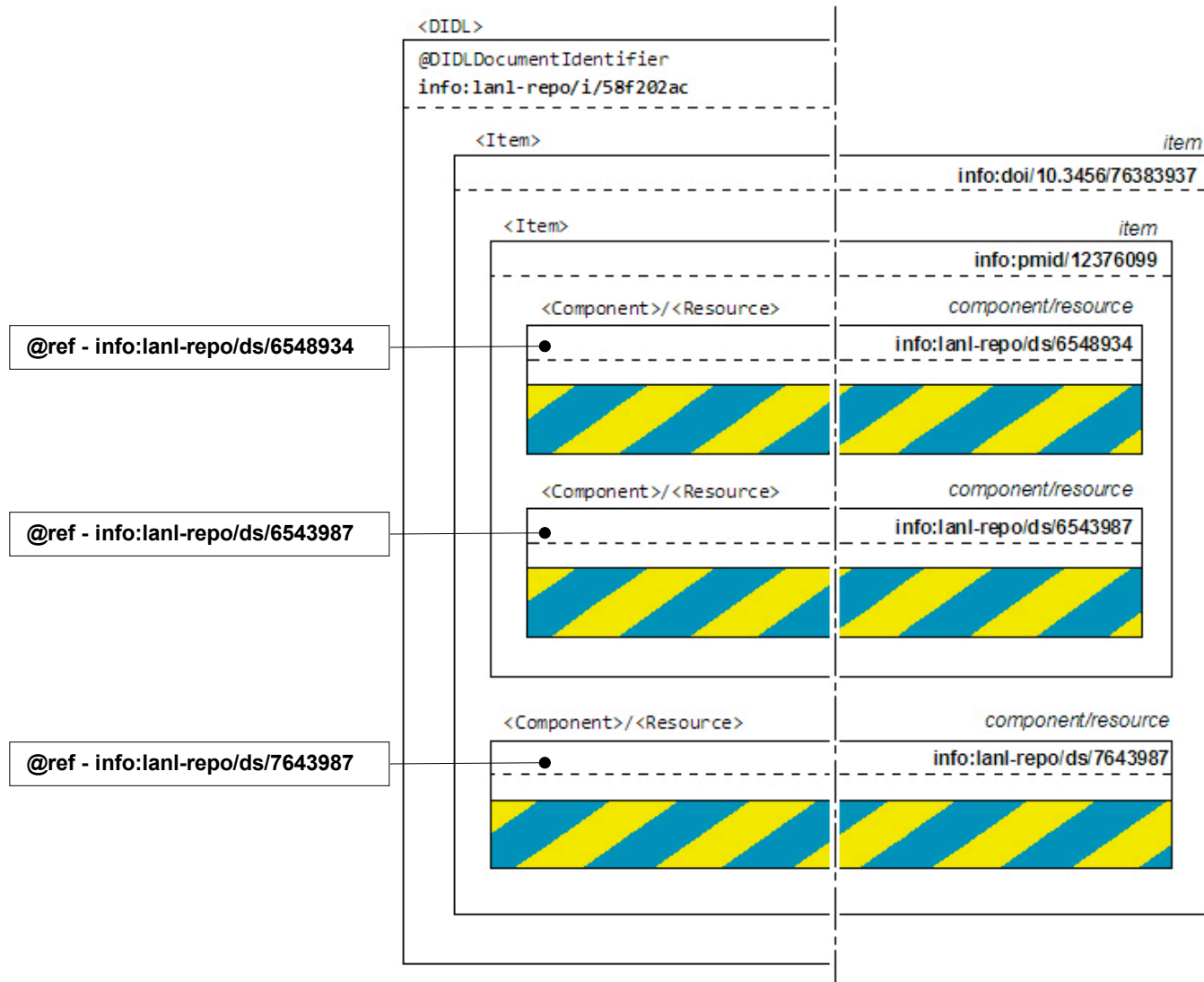
Collection information: `info:sid/library.lanl.gov:collection`. Can be Item and Component level.

DIADM dcterms:extent : datastream size



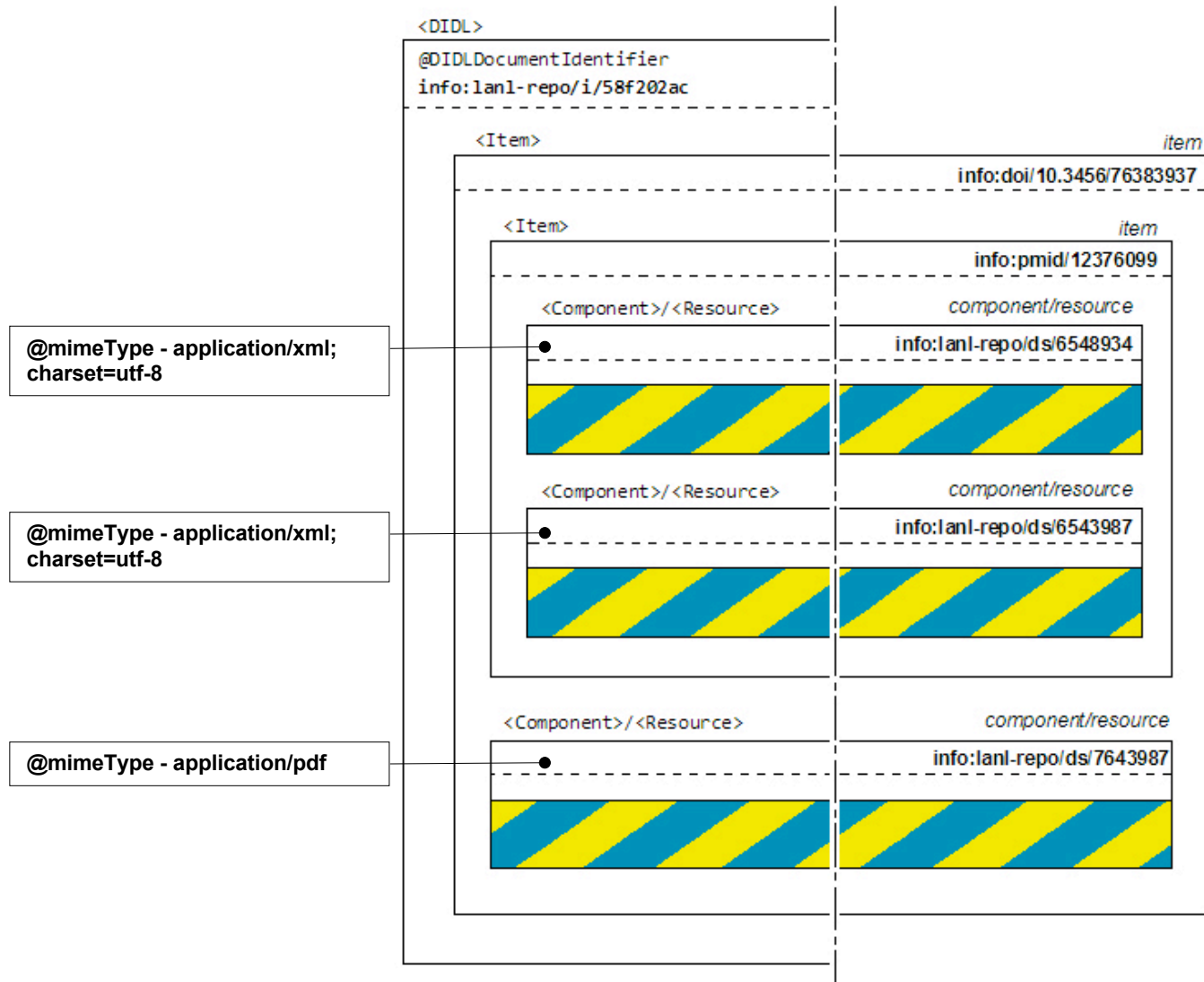
Use consistent metric, i.e. bytes. Component level only.

DIDL Resource @ref : pointing at external datastreams



@ref value = datastream identifier: info:lanl-repo/ds/UUID

DIDL Resource @mimeType : MIME types of datastreams



Use dc.format info:lanl-repo/**f**mt/**i**nTEGER to convey types of XML.

MPEG-21 DIDL and OAI-PMH?

- Yes, of course: a DIDL document is an XML document and hence can be transferred using OAI-PMH
- See:
 - Van de Sompel, Herbert, Michael L. Nelson, Carl Lagoze, Simeon Warner. Resource Harvesting within the OAI-PMH Framework. 2004. D-Lib Magazine. <http://dx.doi.org/10.1045/december2004-vandesompel>
 - Bekaert, Jeroen, and Herbert Van de Sompel. A Standards-based Solution for the Accurate Transfer of Digital Assets. 2005. D-Lib Magazine. <http://dx.doi.org/10.1045/june2005-bekaert>
 - Van de Sompel, Herbert, Jeroen Bekaert, Xiaoming Liu, Lyudmila Balakireva, Thorsten Schwander. aDORe: a modular, standards-based Digital Object Repository. 2005. The Computer Journal. <http://dx.doi.org/10.1093/comjnl/bxh114>
 - OAI-PMH and OpenURL are the core protocols to access materials in the aDORe repository