

the OAI-PMH

service provider

data provider



Core concepts in OAI-PMH

- low-barrier interoperability
- data-provider & service-provider model
- metadata harvesting model
 - OAI-PMH
 - HTTP based
 - Reply · XML Schema
 - Self contained
- shared metadata format and parallel, community-specific metadata formats
 - Dublin Core
- acceptable use
 - Community specific,
 - oai-rights

OAI-PMH data model



← resource

OAI-PMH sets

OAI-PMH identifier

entry point to all records pertaining to the resource

← item

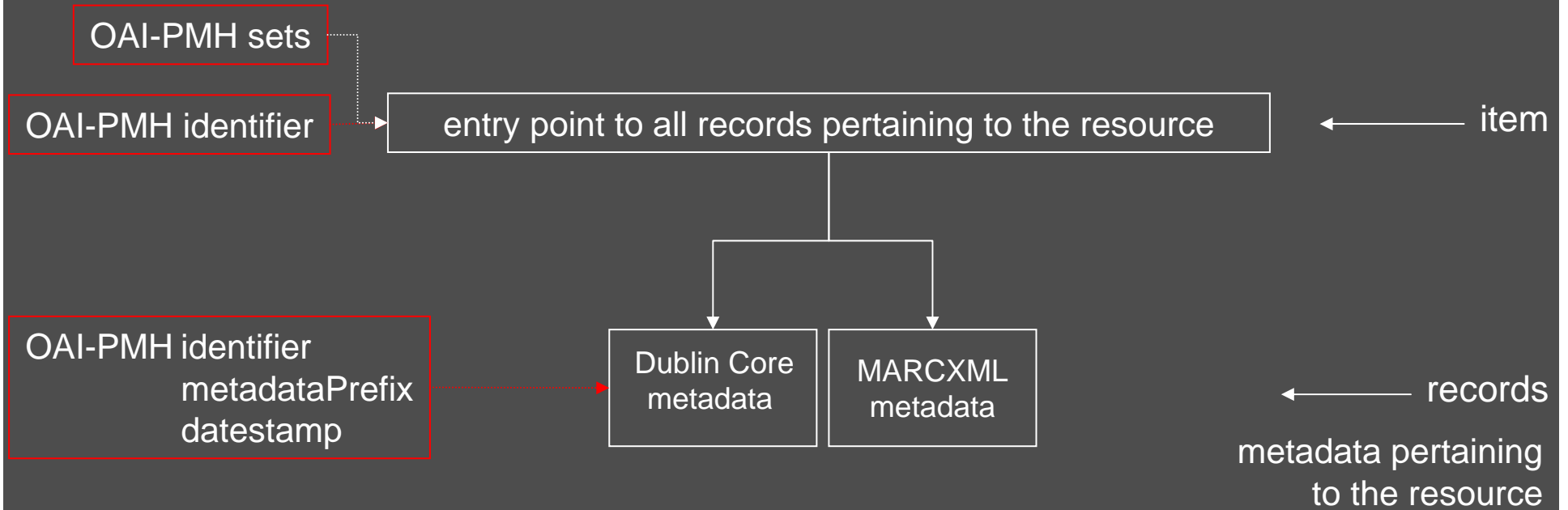
OAI-PMH identifier
metadataPrefix
datestamp

Dublin Core
metadata

MARCXML
metadata

← records

metadata pertaining
to the resource



OAI-PMH harvesting tools

service provider

data provider

h
a
r
v
e
s
t
e
r



Supporting protocol requests:

- Identify
- ListMetadataFormats
- ListSets



r
e
p
o
s
i
t
o
r
y

OAI-PMH harvesting tools

service provider

data provider

h
a
r
v
e
s
t
e
r



Supporting protocol requests:

- Identify
- ListMetadataFormats
- ListSets

Harvesting protocol requests:

- ListRecords
- ListIdentifiers
- GetRecord

r
e
p
o
s
i
t
o
r
y



OAI-PMH harvesting tools

service provider

data provider



harvesting requests

service provider

data provider

- * **from=a**
- * **until=b**
- * **set=klm**
- * **metadataPrefix=dc**

ListRecords



h
a
r
v
e
s
t
e
r



ListRecords / Time / Request
REPEAT
• Identifier
• Datestamp
• Metadata
/REPEAT

r
e
p
o
s
i
t
o
r
y



OAI-PMH identifiers

- Not (necessarily) identifier of the resource
- Each item **must** have a **globally unique** identifier
- identifiers must follow rules for valid **URIs**
- Example:
 - oai:<archiveId>:<recordId>
 - oai:etd.vt.edu:etd-1234567890
- Each identifier must resolve to a single item and always to the same item
 - Can't reuse OAI item identifiers

OAI-PMH timestamps

- Needed for every OAI record to support incremental harvesting
- Must be updated when addition or modification or deletion made in order to ensure changes are correctly propagated to harvesters
 - Also for dynamically generated metadata formats
- Different from dates within the metadata - OAI timestamp is used only for harvesting
- Can be either YYYY-MM-DD or YYYY-MM-DDThh:mm:ssZ (must be GMT timezone)