



DATA MARKET
AUSTRIA

A Semantic Catalogue for the Data Market Austria

Victor Mireles
Semantic Web Company



The Data Market Austria (DMA)

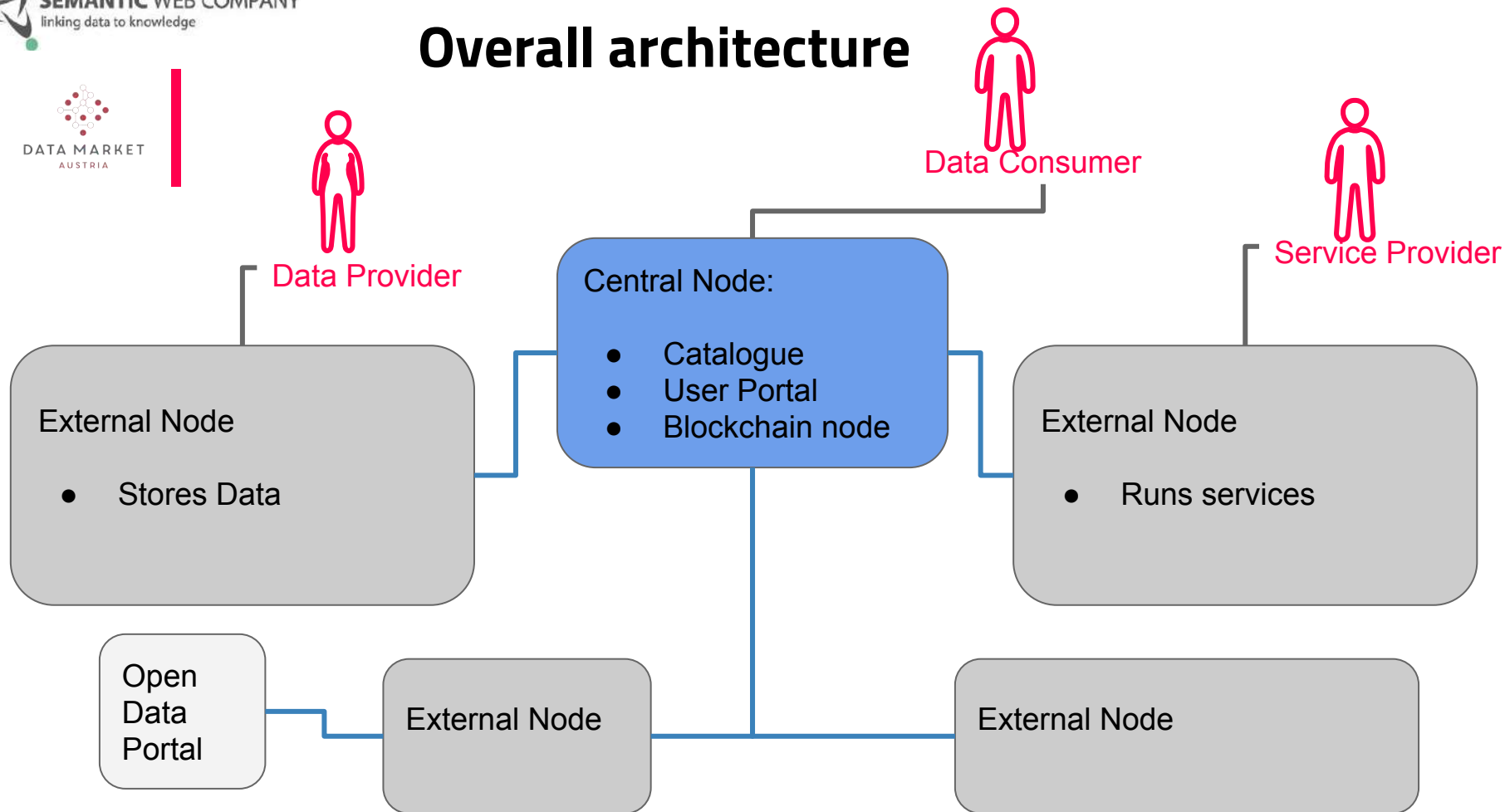
- Data is the new gold
 - Business and society benefit from an agile data economy
 - Data providers, processors, consumers

The Data Market Austria (DMA)

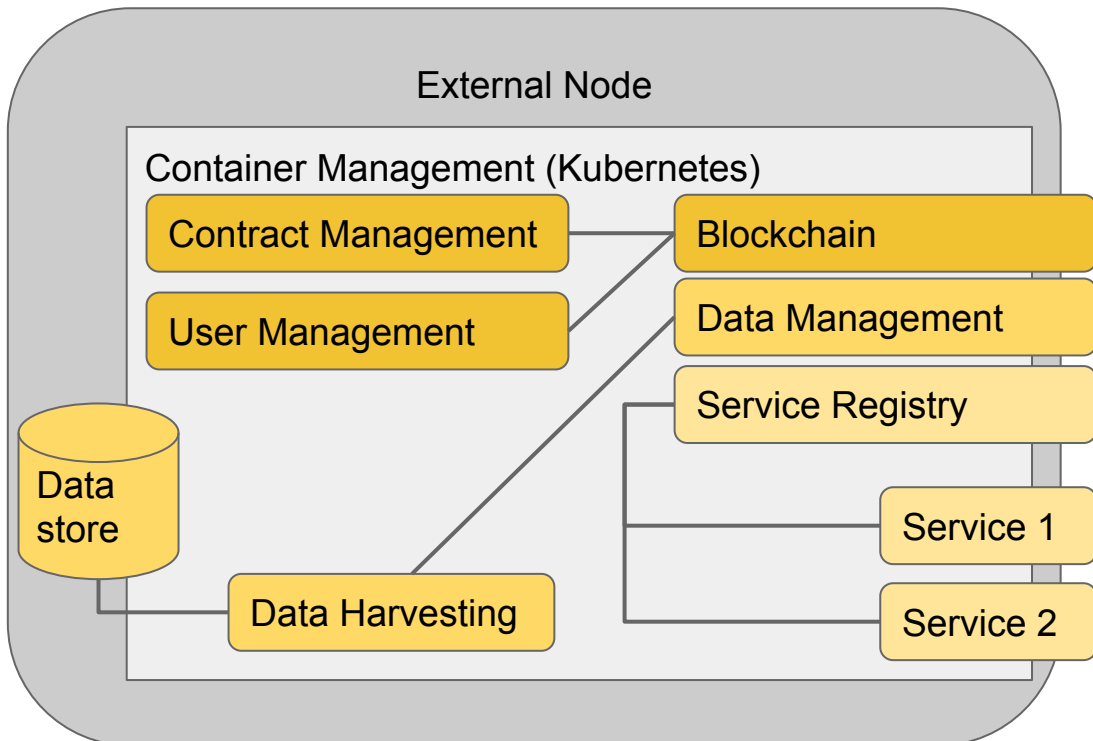
- Data is the new gold
 - Business and society benefit from an agile data economy
 - Data providers, processors, consumers
- Centralization is not an option
 - Privacy / IP Issues
 - Vast amounts of data
 - Great variety of independent data workflows
 - Organizations have data not originally intended for selling

Architecture Overview

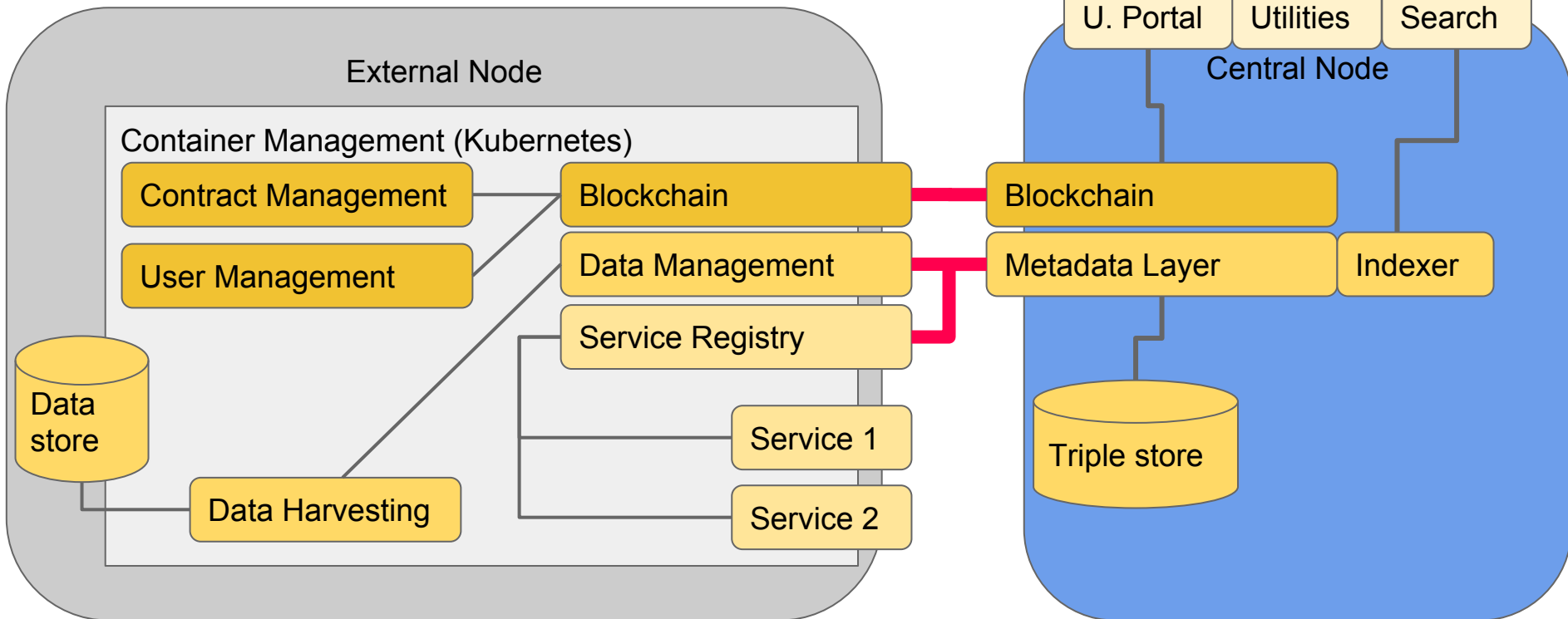
Overall architecture



Overall architecture



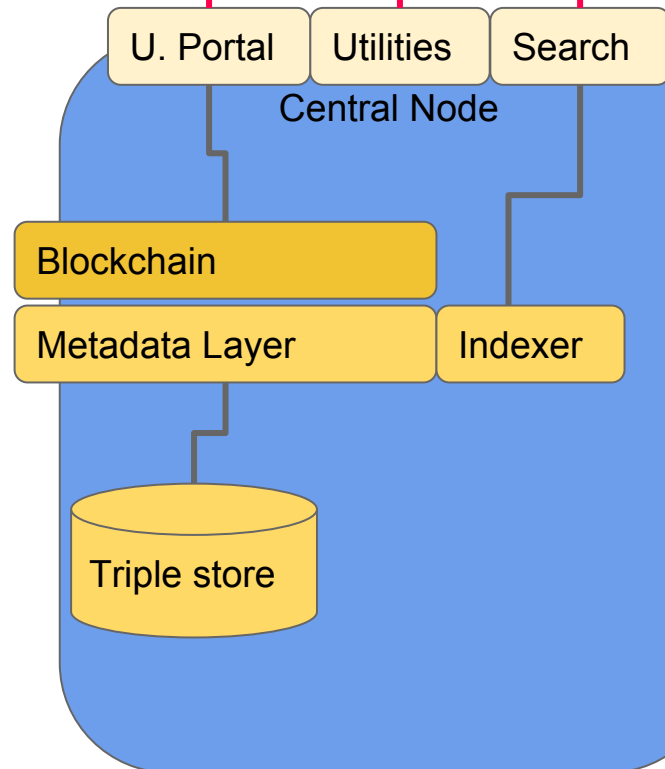
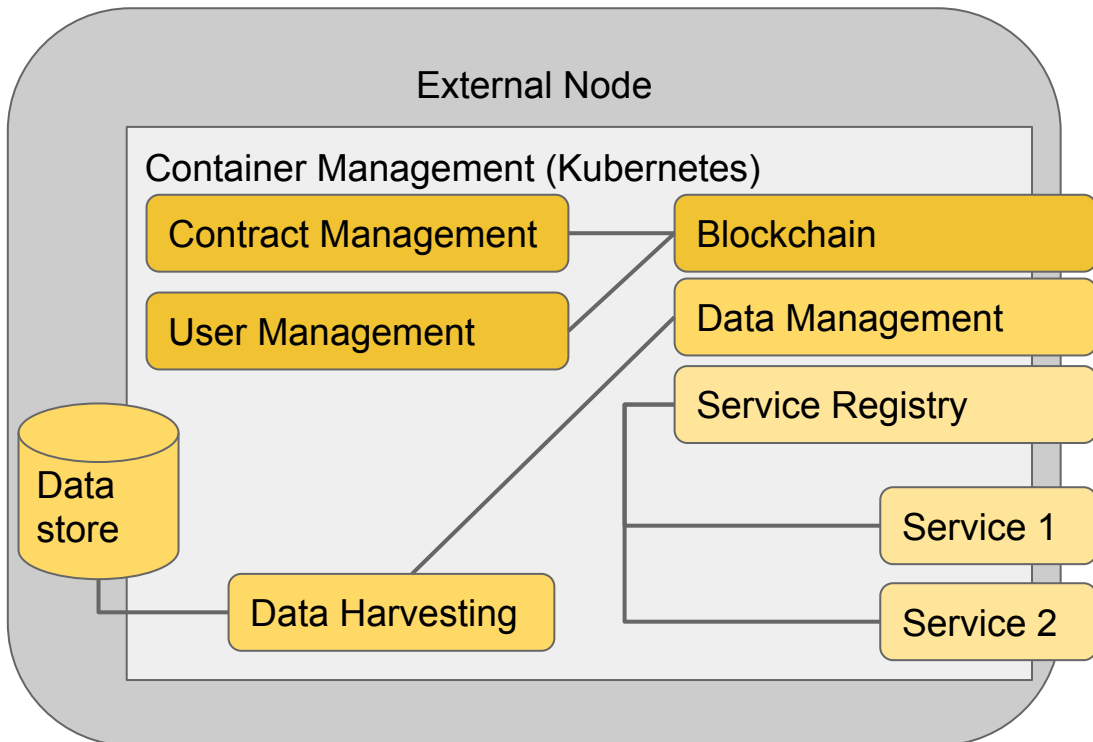
Overall architecture



Overall architecture



Consumer



A catalogue of decentralized repositories

A catalogue of decentralized repositories

Organizations

- Have their own metadata formats
- They use their own controlled vocabularies (if any)
- They get to keep their data and associated processes
- Can have from one to 10^8 datasets

A catalogue of decentralized repositories

Organizations

- Have their own metadata formats
- They use their own controlled vocabularies (if any)
- They get to keep their data and associated processes
- Can have from one to 10^8 datasets



The DMA

- Provides a unified search interface
- Links, enriches and contextualizes metadata
- Provides pointers to the actual data / services
- Stays up-to-date

A catalogue of decentralized repositories

Organizations

- Have their own metadata formats
- They use their own controlled vocabularies (if any)
- Use XML/JSON/YAML to describe their assets



The DMA

- Extends DCAT
- Has a set of controlled vocabularies
- Stores in RDF for semantic processing
- Uses Blockchain

A catalogue of decentralized repositories

Organizations

- Have their own metadata formats
- They use their own controlled vocabularies (if any)
- Use XML/JSON/YAML to describe their assets

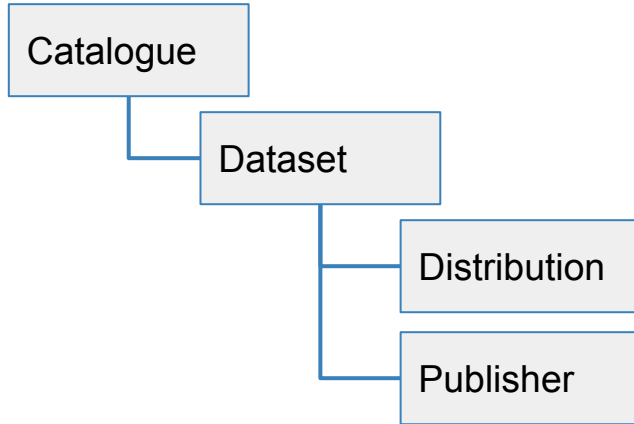


The DMA

- Extends DCAT
- Has a set of controlled vocabularies
- Stores in RDF for semantic processing
- Uses Blockchain

(From arbitrary metadata To DMA metadata)

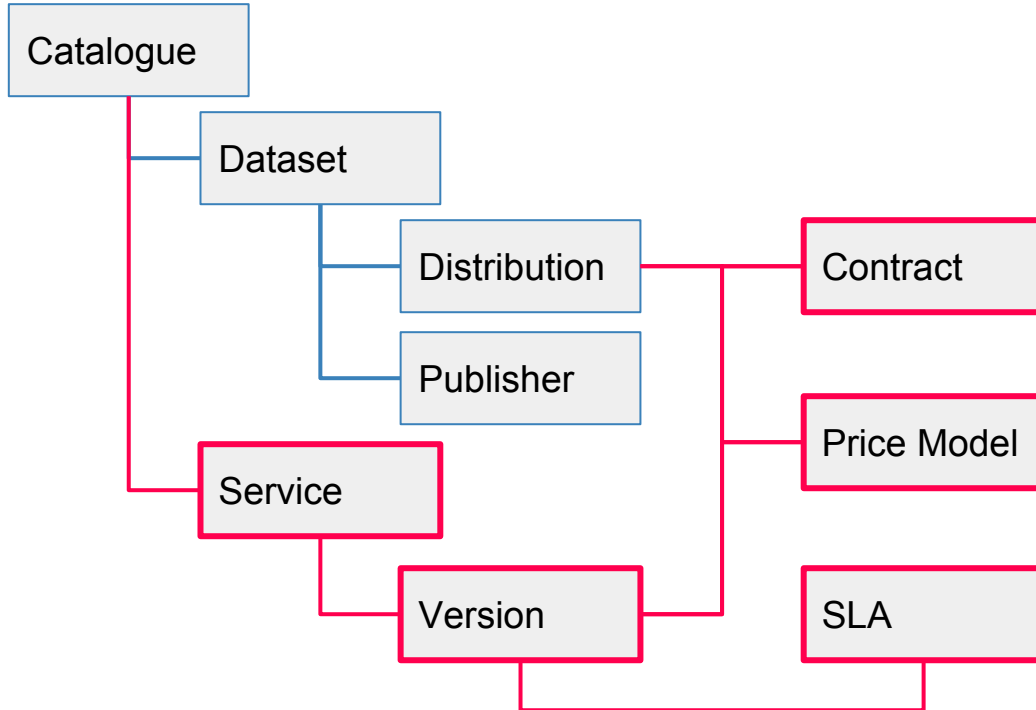
From arbitrary metadata To DMA metadata



The DMA

- Extends DCAT

From arbitrary metadata To DMA metadata



The DMA

- Extends DCAT
- Uses Blockchain (the IDs of all these resources, and the depicted relations are stored in a distributed ledger)



From arbitrary metadata To DMA metadata

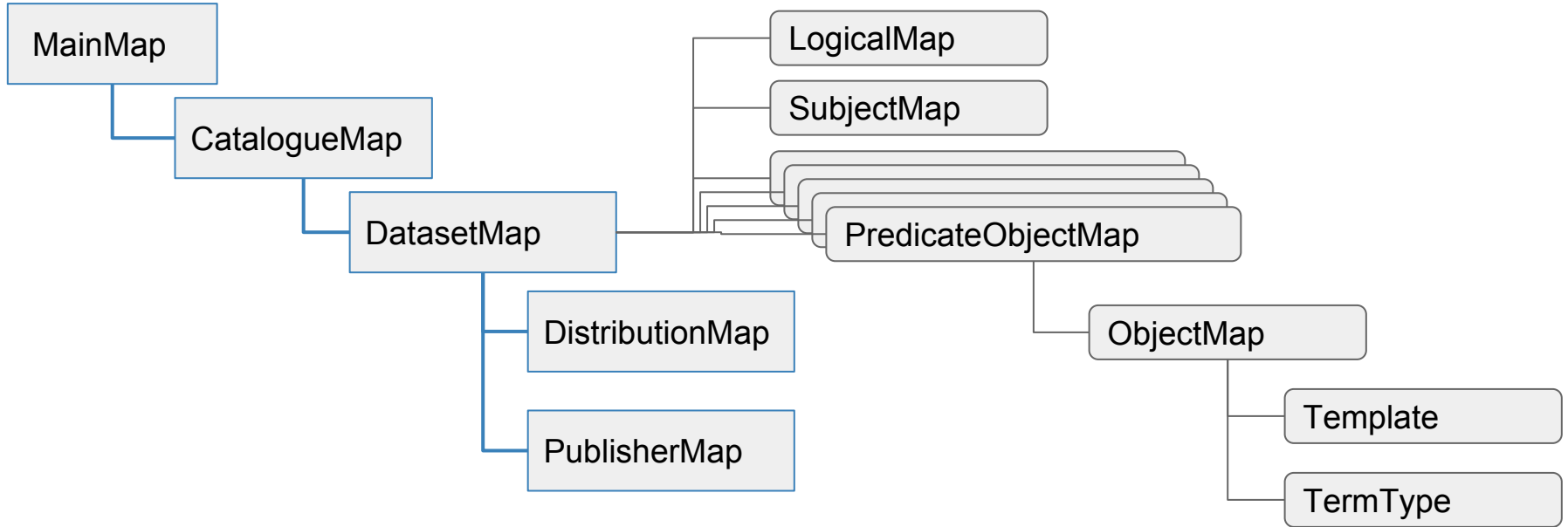
```
{
  "help": "https://www.europeandataportal.eu/data/en/api/3/action/package_show",
  "success": true,
  "result": {
    "owner_org": "be748ad2-b244-48f7-815b-d2d06164046c",
    "maintainer": null,
    "private": false,
    "maintainer_email": null,
    "num_tags": 7,
    "/id": "99c1a383-c222-4f5b-918c-02ed9ba0ff66",
    "metadata_created": "2017-03-03T18:55:31.346475",
    "metadata_modified": "2017-03-03T19:12:43.637759",
    "author": null,
    "author_email": null,
    "temporal": [
      {
        "start_date": "2001-03-01T00:00:00",
        "end_date": "2001-09-30T00:00:00"
      }
    ],
    "state": "active",
    "version": null,
    "license_id": "cc-by",
    "type": "dataset",
    "resources": [
      {
        "cache_last_updated": null,
        "package_id": "99c1a383-c222-4f5b-918c-02ed9ba0ff66",
        "webstore_last_updated": null,
        "id": "44f4182f-3c20-41be-ae3-MYID5",
        "size": null,
        "state": "active",
        "hash": "",
        "description": "A list of all courses and their events at the Vienna University of Economics and Business in SS01",
        "format": "CSV",
        "mimetype_inner": null,
        "url_type": null,

```

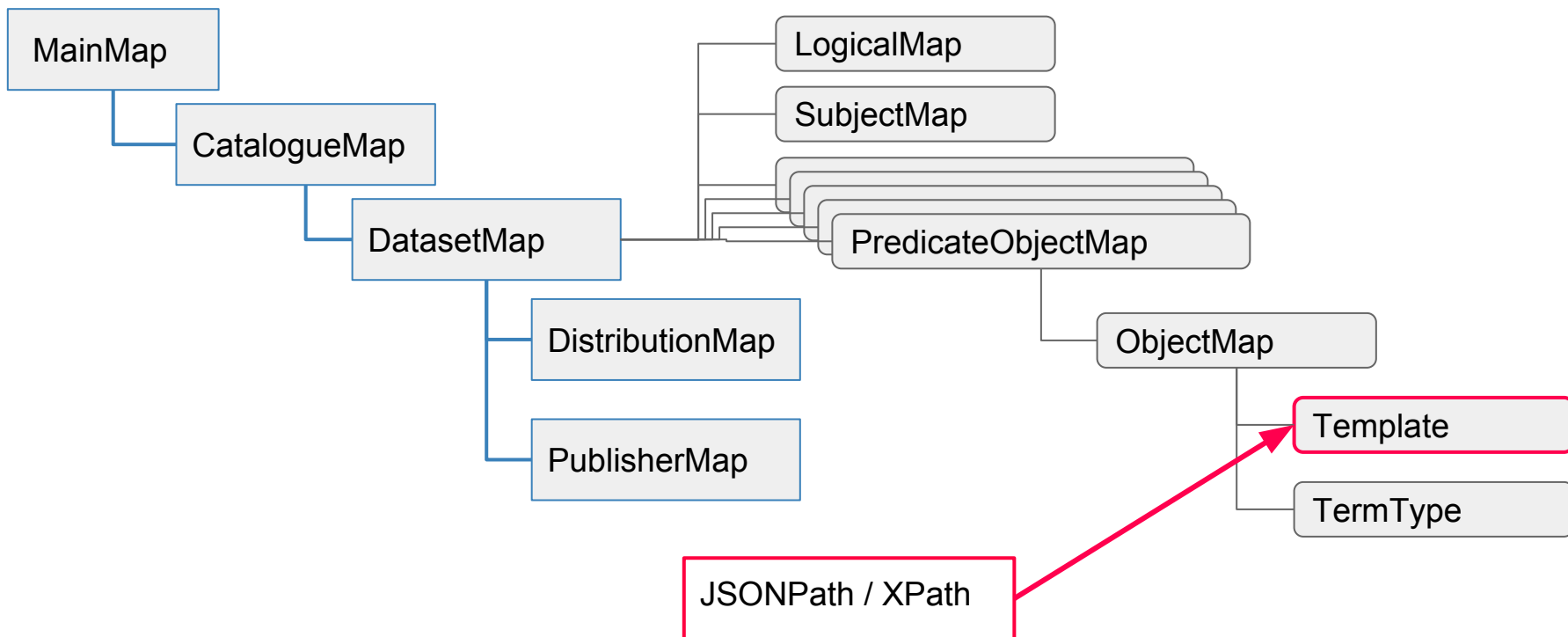
```
<http://external.node/catalog_1> a <http://www.w3.org/ns/dcat#Catalog>;
  <http://purl.org/dc/terms/identifier> "be056863-d93b-45f8-8f7-6c4ff069c861";
  <http://purl.org/dc/terms/title> "Catalogue of the DMA external node 1";
  <http://purl.org/dc/terms/description> "External node 1 is owned by organization X";
  <http://purl.org/dc/terms/language> http://id.loc.gov/vocabulary/iso639-1/en;
  <http://purl.org/dc/terms/publisher> _:bnode-b_dc-publisher;
  <http://purl.org/dc/terms/rights> "public";
  <http://www.w3.org/ns/dcat#dataset> <http://external.node/repository/8167cc12-1d40-484e-bf4b-2e3a1f5ba09c>.
  _:bnode-b_dc-publisher a <http://xmlns.com/foaf/0.1/Organization>;
  <http://purl.org/dc/terms/identifier> "6ffcc0cf-ba45-4f73-bb77-ff3799332938".
<http://external.node/repository/8167cc12-1d40-484e-bf4b-2e3a1f5ba09c> a <http://www.w3.org/ns/dcat#Dataset>;
  <http://purl.org/dc/terms/identifier> "8167cc12-1d40-484e-bf4b-2e3a1f5ba09c";
  <http://purl.org/dc/terms/title> "A list of all courses and their events at the Vienna University of Economics and Business in SS01";
  <http://purl.org/dc/terms/description> "A list of all courses and their events at the Vienna University of Economics and Business in SS01";
  <http://www.w3.org/ns/dcat#keyword> "https://dma.poolparty.biz/dmav/64";
  <http://www.w3.org/ns/dcat#theme> "http://eurovoc.europa.eu/c_5ea6e5c4";
  <http://datamarket.at/2017/07/dmav/core#User_Generated_Tags> "European";
  <http://purl.org/dc/terms/language> "http://id.loc.gov/vocabulary/iso639-1/en";
  <http://purl.org/dc/terms/accessRights> "https://dma.poolparty.biz/dmav/12";
  <http://datamarket.at/2017/07/dmav/core#extUID> "44f4182f-3c20-41be-ae3-MYID5";
  <http://purl.org/dc/terms/publisher> _:bnode-b_dcat-dataset_dc-publisher;
  <http://www.w3.org/ns/dcat#distribution>
<http://external.node/repository/8167cc12-1d40-484e-bf4b-2e3a1f5ba09c/b1a2b63e-12e0-46c9-8d24-5a11c50c872e>;
  <http://www.w3.org/ns/dcat#contactPoint> _:bnode-b_dcat-dataset_dc-contactPoint.
  _:bnode-b_dcat-dataset_dc-publisher a <http://xmlns.com/foaf/0.1/Organization>;
  <http://xmlns.com/foaf/0.1/name> "Open Data Portal@WU";
  <http://xmlns.com/foaf/0.1/mbox> "mailto:contact@data.wu.ac.at".

<http://external.node/repository/8167cc12-1d40-484e-bf4b-2e3a1f5ba09c/b1a2b63e-12e0-46c9-8d24-5a11c50c872e> a
<http://www.w3.org/ns/dcat#Distribution>;
  <http://purl.org/dc/terms/identifier> "8167cc12-1d40-484e-bf4b-2e3a1f5ba09c/b1a2b63e-12e0-46c9-8d24-5a11c50c872e";
  <http://purl.org/dc/terms/title> "Default";
  <http://purl.org/dc/terms/description> "Default Distribution";
  <http://purl.org/dc/terms/license> <http://none.com/cc-by>;
  <http://purl.org/dc/terms/format> "CSV";
  <http://datamarket.at/2017/07/dmav/core#PriceModel> "1d57bd57-7367-4bdb-ac43-1362a458ce60";
  <http://www.w3.org/ns/dcat#accessURL>
<http://data.wu.ac.at/portal/dataset/169e2d7c-41f6-493b-b229-88fac2a0321a/resource/5150029d-d4c9-472f-8d8a-4e28f456ae41/download/allcourse
sandevents01s.csv>.
```

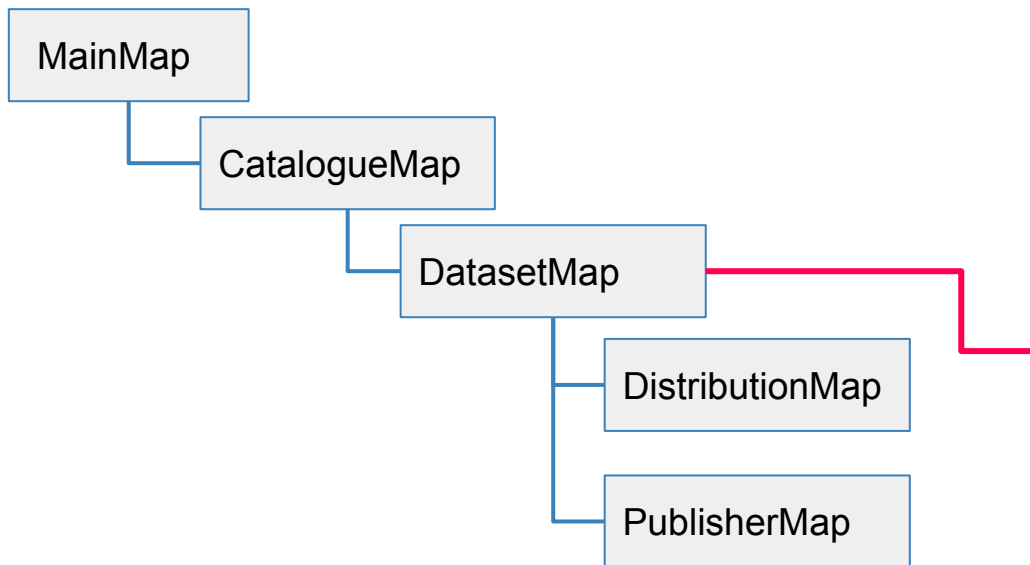
From arbitrary metadata To DMA metadata



From arbitrary metadata To DMA metadata



From arbitrary metadata To DMA metadata



```
<ex:dcatdatasetMap>
  rml:logicalSource <ex:LogSource>;
  rr:subjectMap [ rr:template "DATASET_ID_PLACEHOLDER";
                 rr:termType rr:IRI;
                 rr:class dcat:Dataset; ];

  rr:predicateObjectMap
  [ rr:predicate dc:identifier;
    rr:objectMap
    [ rr:template "[DATASET_ID_PLACEHOLDER]";
      rr:termType rr:IRI;
    ]; ];

  rr:predicateObjectMap
  [ rr:predicate dc:title;
    rr:objectMap
    [ rr:template "{$.[\"result\"].[\"resources\"]}[0].[\"description\"]";
      rr:termType rr:Literal;
    ]; ];

  rr:predicateObjectMap
  [ rr:predicate dc:description;
    rr:objectMap
    [ rr:template "{$.[\"result\"].[\"resources\"]}[0].[\"description\"]";
      rr:termType rr:Literal;
    ]; ];
```










From arbitrary metadata To DMA metadata



Schema Mapper

Step 1 **Step 2** Step 3

Step 2: Match Fields
Some help text to explain this

Title *	↔	<input type="text" value="Type or select"/>	
Description *	↔	<input type="text" value="Type or select"/>	
Publisher *	↔	<input type="text" value="Type or select"/>	
Tags *	↔	<input type="text" value="Type or select"/>	
User Tags	↔	<input type="text" value="Type or select"/>	
Language	↔	<input type="text" value="Type or select"/>	
Contact *	↔	<input type="text" value="Type or select"/>	
Theme *	↔	<input type="text" value="Type or select"/>	
Access Rights *	↔	<input type="text" value="Type or select"/>	

[< Previous](#) [Next >](#)

From arbitrary metadata To DMA metadata



Schema Mapper

Step 1 **Step 2** Step 3

Step 2: Match Fields

Some help text to explain this

Title *	↔	Type or select	⚙
Description *	↔	Type or select	⚙
Publisher *	↔	Type or select	⚙
Tags *	↔	Type or select	⚙
User Tags	↔	Type or select	⚙
Language	↔	Type or select	⚙
Contact *	↔	Type or select	⚙
Theme *	↔	Type or select	⚙
Access Rights *	↔	Type or select	⚙

[← Previous](#) [Next →](#)



Path selection

Currently selected path: "result" -> "resources" -> "0"

Values for the given path: No values found, please add items until a leaf-element gets detected

[Add by direct selection](#) [Add by filter](#)

Some Select a sub-item by clicking on it or use the button of the bottom to add all sub-items.

"cache_last_updated"
"package_id" (!)
"webstore_last_updated"
"id" (!)
"size"
"state" (!)
"hash" (!)
"description" (!)
"format" (!)
"mimetype_inner"
"url_type"
"translation"
"access_url"
"mimetype"
"cache_url"

From arbitrary metadata To DMA metadata

Metadata fields that are enriched:

- Description (dct:description)
- Title (dct:title)
- User-generated tags (dcat:keyword)

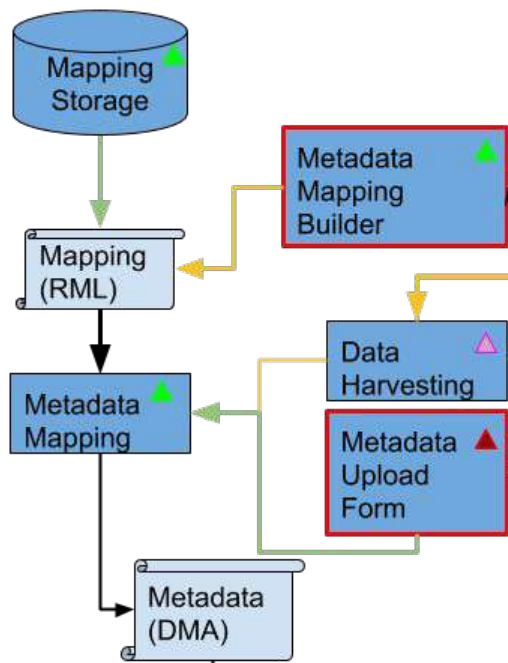
```
dct:description "The Q1 report of ..."
```

PP

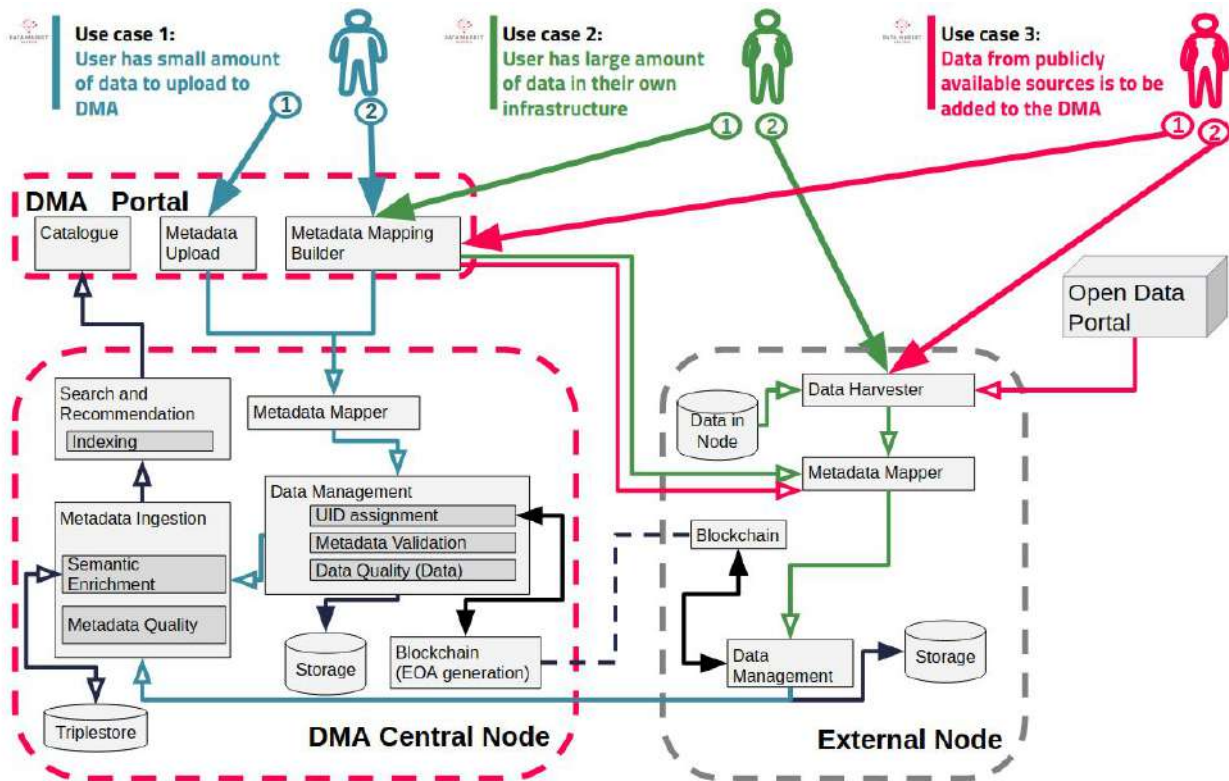
```
dct:description <_:descr> .  
  
<_:descr>  
nif:isString "The Q1 report of ..." .  
  
<_:descr#char=7,12> a nif:RFC5147String  
;  
nif:anchorOf "report"^^xsd:string ;  
nif:beginIndex "7"^^xsd:int ;  
nif:endIndex "12"^^xsd:int ;  
nif:annotation  
<http://eurovoc.europa.eu/2891>;  
nif:referenceContext <_:descr> .
```

A catalogue of decentralized repositories II.

A catalogue of decentralized repositories



A catalogue of decentralized repositories



Thanks!

victor.mireles@semantic-web.com