



Online, 2021 October 4-15

The role of Dublin Core in the publication of Open Geospatial Data

Javier Nogueras-Iso

Advanced Information Systems Lab (IAAA)

Aragon Institute of Engineering Research (I3A)

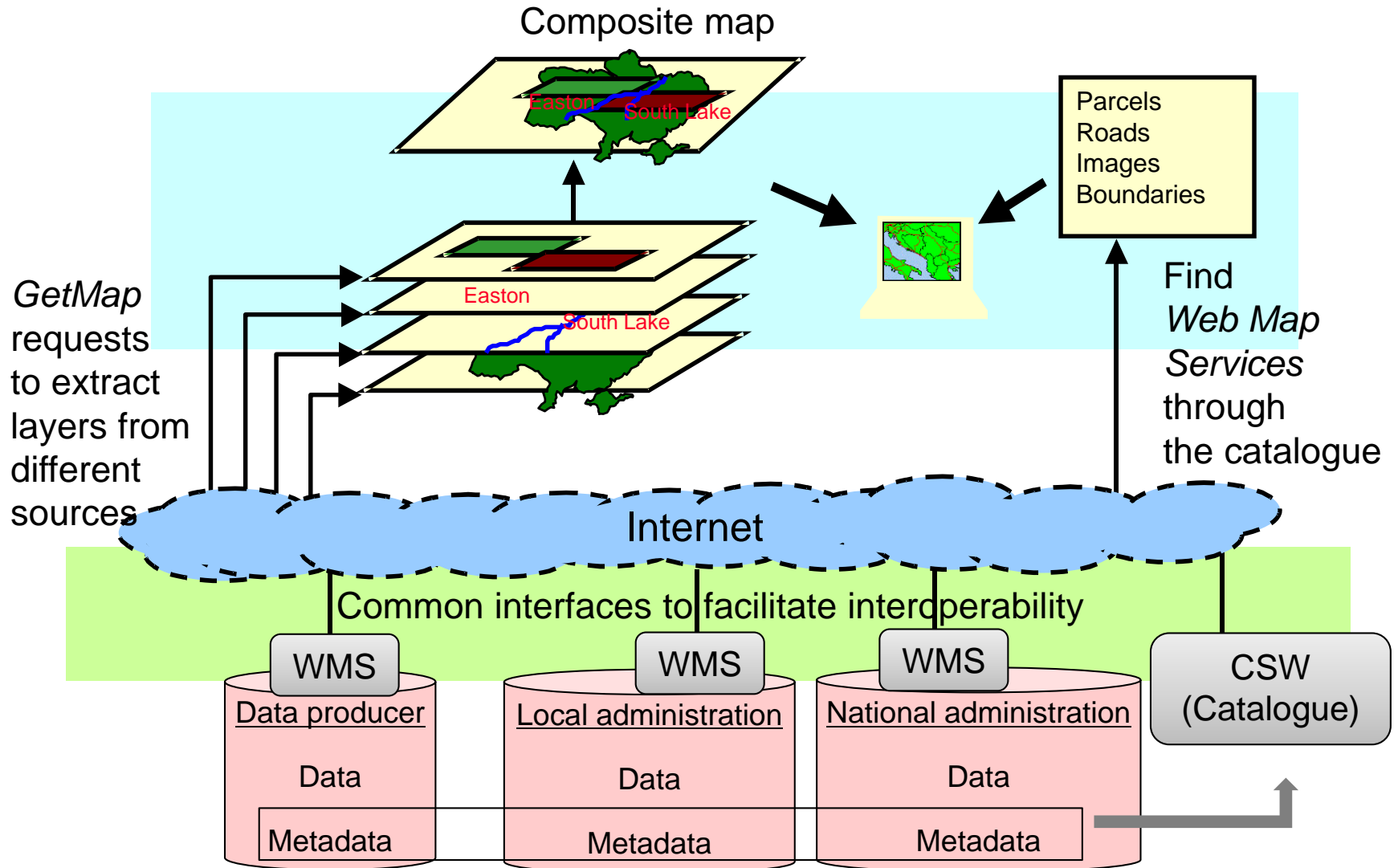
Universidad de Zaragoza, Spain

October 14, 2021

```
<meta name="HTTP-EQUIV"
  content="text/html;charset=iso-8859-1" >
  <meta name="IAAA" coords="7,9,167,32"
  content="http://iaaa.index.html/1/hi.html" shape="RECT" >
  <meta name="IAAA" coords="7,9,167,32"
  content="http://iaaa.index.html/1/hi.html" shape="RECT" >
  <meta name="Section" content="Diseño_Gráfico" >
  <meta name="Description" content="Some enterprises
  re-engineering, let's take over." >
  <meta name="Description" content="Some enterprises
  re-engineering, let's take over." >
  <meta name="Description" content="Diseño_Gráfico" >
  <meta name="IAAA" coords="7,9,167,32"
  content="http://iaaa.index.html/1/hi.html" shape="RECT" >
  <meta name="HTTP-EQUIV"
  content="text/html;charset=iso-8859-1" >
```


Ability to combine data on real time from different sources

- **Metadata: the glue for the different components of an SDI**



Problem: Complexity of geographic metadata standards

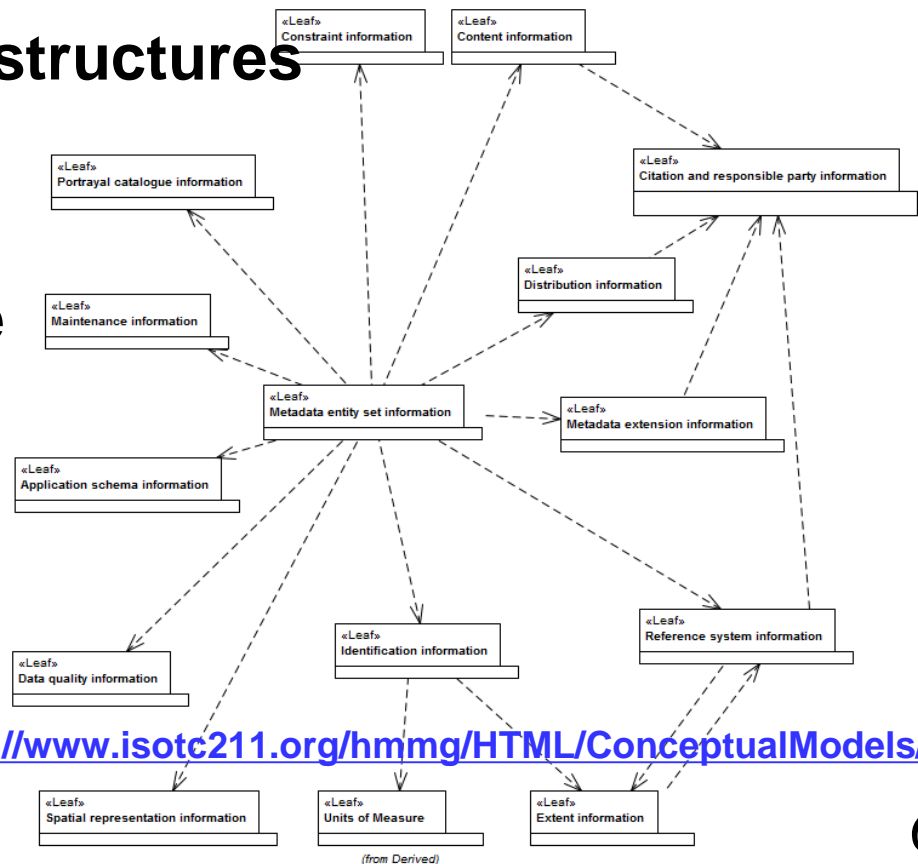
- **Alternatives**

- FGDC: Content Standard for Digital Geospatial Metadata
- ISO/TC 211: ISO 19115 “Geographic Information – Metadata”

- **More than 400 elements, organized in hierarchical structures**

- **Need for specialized metadata edition tools**

- **Need for human expertise**

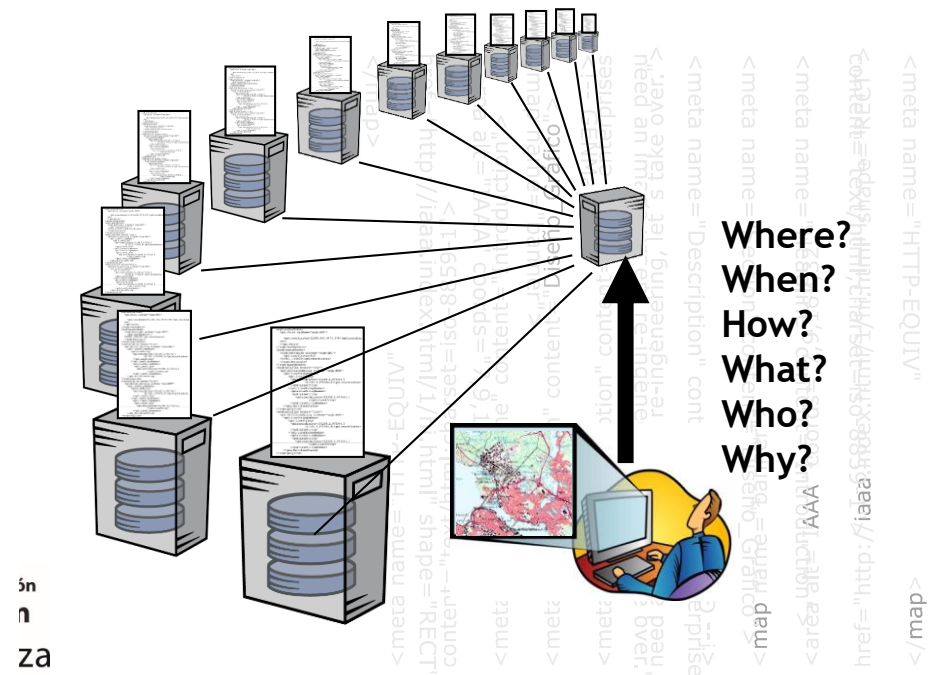


Interoperability between different geographic metadata standards

- The Catalog Services for the Web (CSW) protocol binding of the OGC Catalog Services specification is based on Dublin Core (2003-2007)**
 - Dublin Core for queryable properties
 - Dublin Core for returnable properties in result records

Table 53 — Mapping of Dublin Core names to XML element names

Dublin Core element name	OGC queryable term	XML element name
title	Title	dc:title
creator		dc:creator
subject	Subject	dc:subject
description	Abstract	dct:abstract
publisher		dc:publisher
contributor		dc:contributor
date	Modified	dct:modified
type	Type	dc:type
format	Format	dc:format
identifier	Identifier	dc:identifier
source	Source	dc:source
language		dc:language
relation	Association	dc:relation
coverage	BoundingBox	ows:BoundingBox
rights		dc:rights



Interoperability across different application domains

- **CSDGM – Dublin Core mapping**

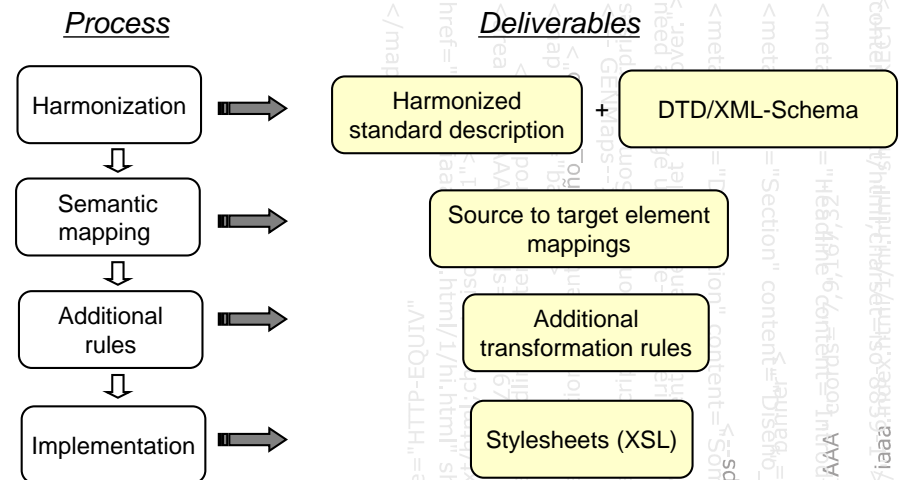
- Developed by United States Geological Survey (2000)

- **ISO 19115 – Dublin Core mapping**

- European Territorial Management Information Infrastructure (ETeMII) European project (1999-2001)
- CEN/ISSS Metadata for Multimedia Information – Dublin Core Workshop project: “improving discovery of geographic information in cross-domain searching” (2002-2003)

DUBLIN-CORE			ISO 19115	
Nr	Element	refinement	Core	Element (longname)
1	TITLE	<none>	Y	MD_Metadata.identificationInfo > MD_DataIdentification.citation > CI_Citation.title
1	TITLE	alternative	N	MD_Metadata.identificationInfo > MD_DataIdentification.citation > CI_Citation.alternateTitle
2	CREATOR		Y	MD_Metadata.identificationInfo > MD_DataIdentification.pointOfContact > CI_ResponsibleParty (when role="originator")

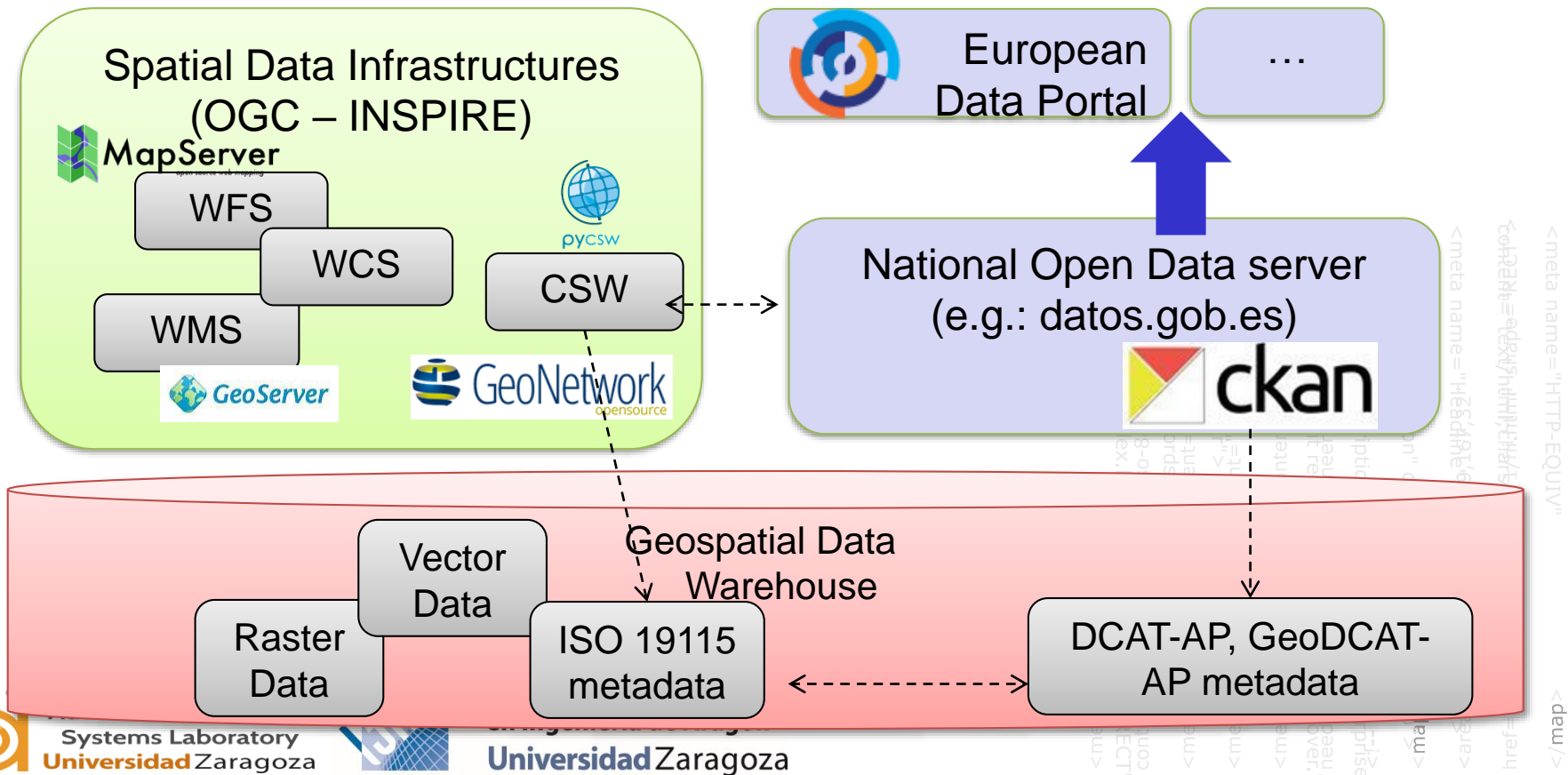
CWA 14857, mapping



CWA 14856, crosswalk development

3. Dublin Core in the Open Geospatial Data era

- With the launch of Open Data initiatives (2010 onwards), Open Data portals are a big competitor to SDI download services (geospatial data are open government data)



Metadata approaches directly based on Dublin Core

- **DCAT: Data Catalogue vocabulary proposed by W3C for the description of Open Data (2012-)**
 - Implemented in well-known software packages such as CKAN
- **DCAT-AP: European Application Profile of DCAT for public sector datasets (2013-)**
- **GeoDCAT-AP: extension of DCAT-AP for geographic information (2015-)**
 - Designed to assure compliance with European INSPIRE directive for establishing a spatial information infrastructure in Europe (and ISO 19115)

```
<meta name="HTTP-EQUIV"
content="text/html; charset=iso-8859-1" shape="RECT"
<area coords="7,9,167,32"
<meta name="Section" content="Diseno_Grupo"
<meta name="Description" content="
<meta name="GENMaps"
<meta name="Diseno_Grupo"
<map
<area
href="http://iaei"
</map>
```

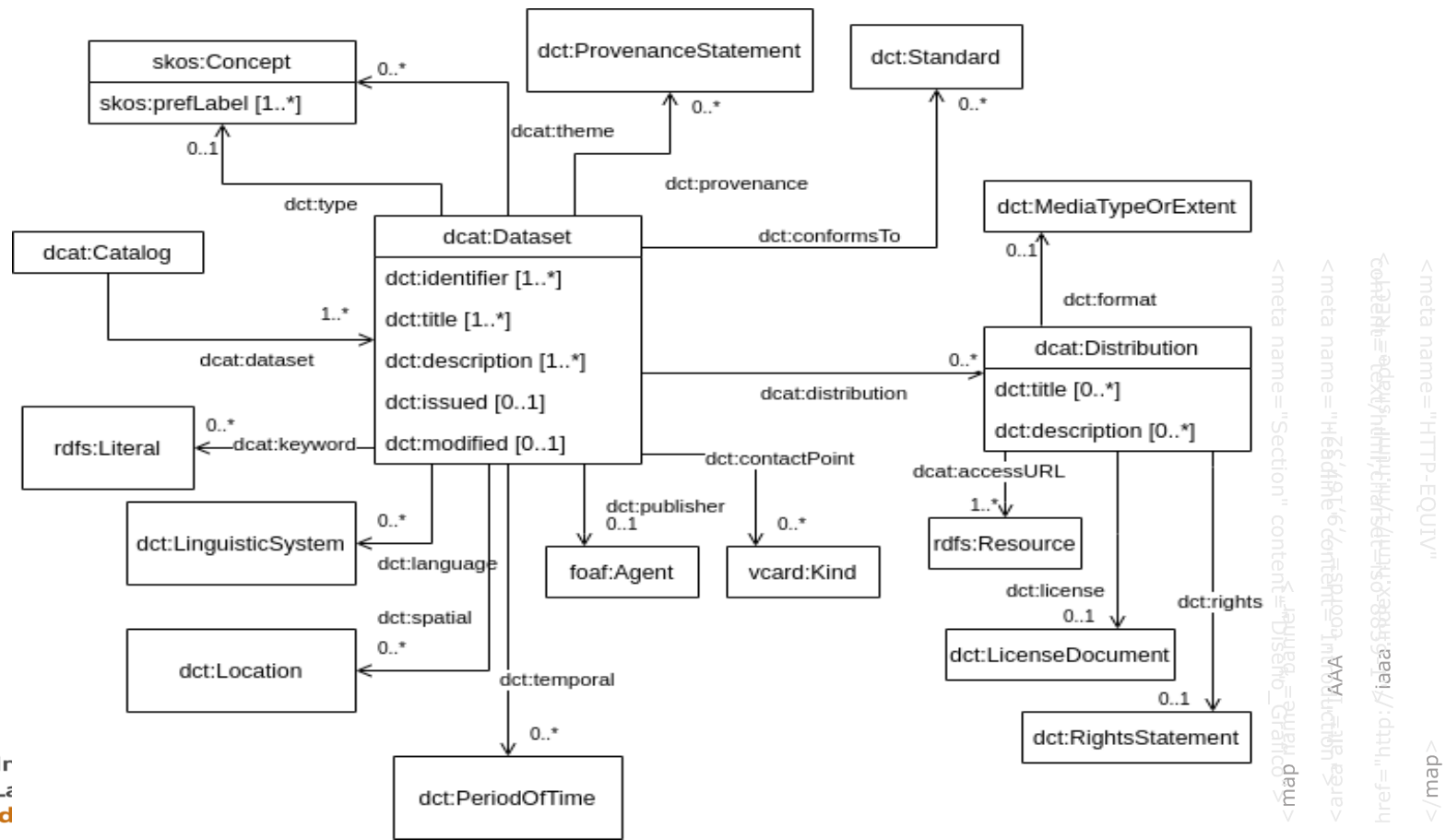
Example of application of GeoDCAT-AP



- **TRAFAIR: Understanding traffic flows to improve air quality**
 - European project co-financed by the Connecting Europe Facility of the European Union (2018 – 2020)
 - Goal: Design and develop the necessary infrastructure to estimate the pollution level on urban scale (6 European cities, different size)
 - Sub-goals:
 - Provide real-time monitoring of air pollution
 - Develop an air quality forecasting service based on the weather forecasts and the urban traffic flow
 - Publish monitoring and forecasting air quality and traffic data as open data
 - Develop applications for end-users and public administrations

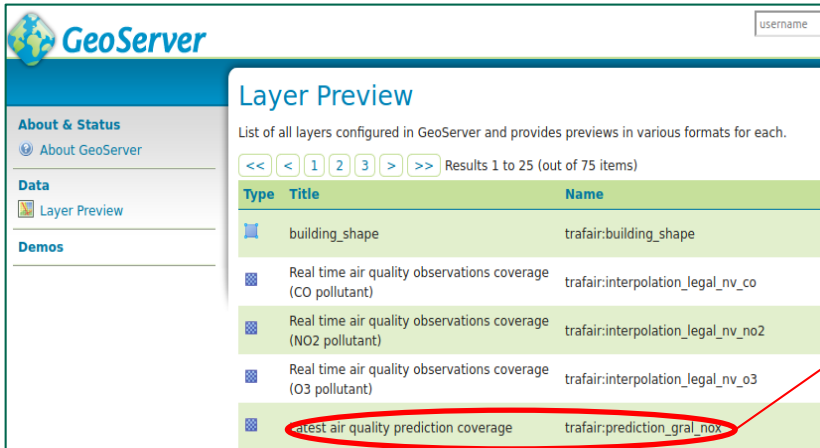
Light adoption of GeoDCAT-AP

- Core properties with direct binding to ISO 19115
- Properties compatible with DCAT-AP, elements editable with CKAN software



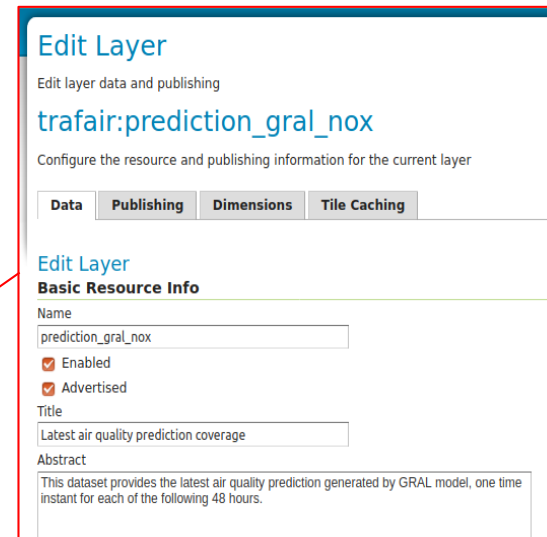
Automate as much as possible the generation of metadata without human interaction

- Metadata is harvested from OGC services through its *GetCapabilities*



The screenshot shows the GeoServer 'Layer Preview' interface. A table lists various layers, with the entry 'Latest air quality prediction coverage' (trafa:prediction_gral_nox) circled in red. A green arrow points from this entry to the CKAN interface below.

Type	Title	Name
building_shape		trafa:building_shape
Real time air quality observations coverage (CO pollutant)		trafa:interpolation_legal_nv_co
Real time air quality observations coverage (NO2 pollutant)		trafa:interpolation_legal_nv_no2
Real time air quality observations coverage (O3 pollutant)		trafa:interpolation_legal_nv_o3
Latest air quality prediction coverage		trafa:prediction_gral_nox



The screenshot shows the 'Edit Layer' interface for 'trafa:prediction_gral_nox'. It displays configuration options for 'Basic Resource Info', including Name, Title, and Abstract. A red arrow points from this interface to the CKAN metadata below.

Edit Layer
trafa:prediction_gral_nox

Basic Resource Info

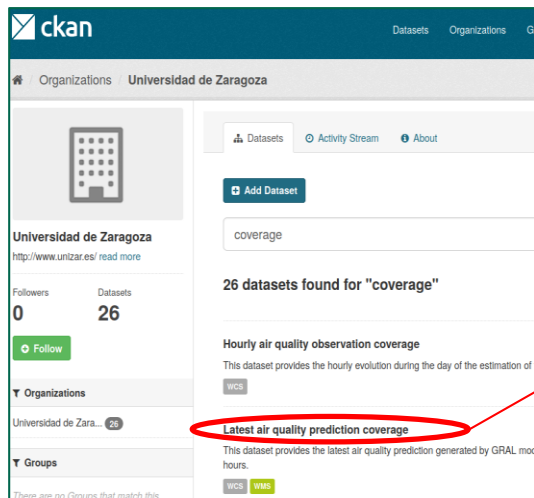
Name: prediction_gral_nox

Enabled

Advertised

Title: Latest air quality prediction coverage

Abstract: This dataset provides the latest air quality prediction generated by GRAL model, one time instant for each of the following 48 hours.



The screenshot shows the CKAN 'Datasets' interface for 'Universidad de Zaragoza'. A search for 'coverage' has returned 26 datasets. The entry 'Latest air quality prediction coverage' is circled in red. A red arrow points from this entry to the metadata XML below.

26 datasets found for "coverage"

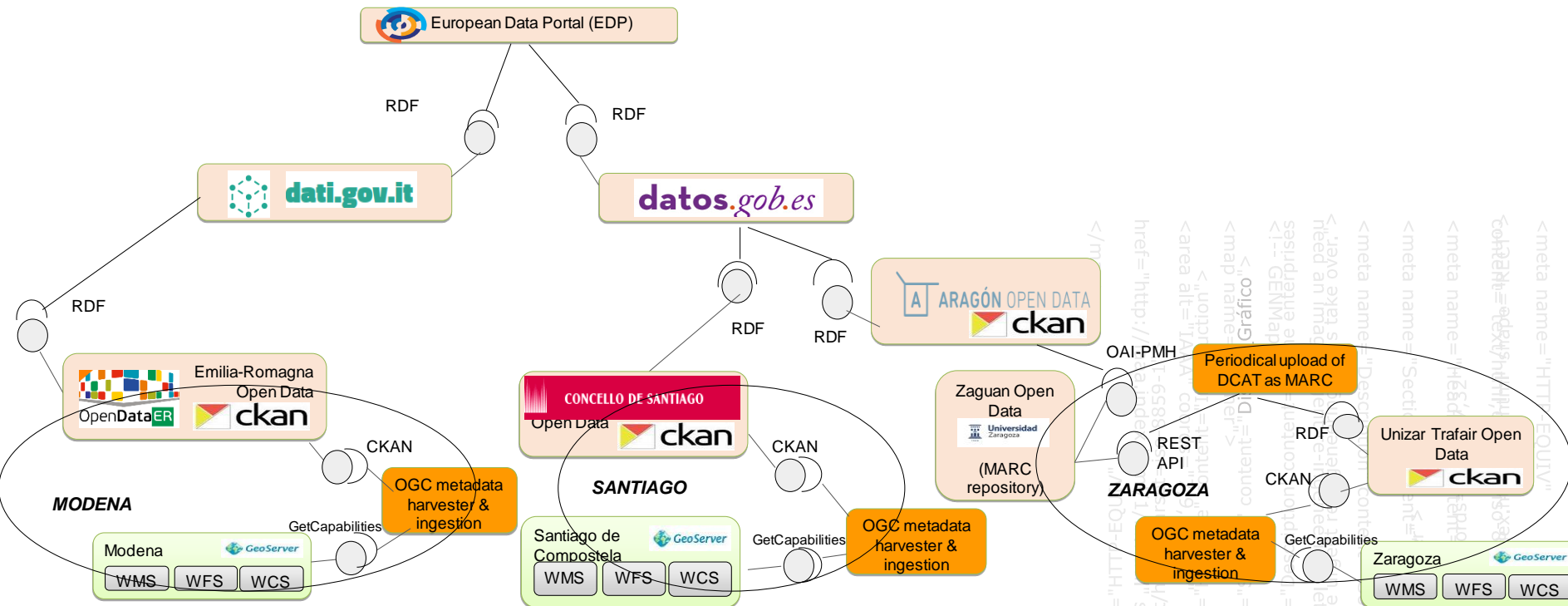
Hourly air quality observation coverage

Latest air quality prediction coverage

```
<dc:Dataset rdf:about="http://atila.unizar.es:3394/dataset/dbb3799e-dc88-4e8f-995b-7d8"
<dc:title>Latest air quality prediction coverage</dc:title>
<dc:description>This dataset provides the latest air quality prediction generated by
<dc:type>http://inspire.ec.europa.eu/metadata-codelist/ResourceType/series</dc:type>
<dc:theme rdf:resource="http://inspire.ec.europa.eu/metadata-codelist/TopicCategory/
<dc:temporal>
  <dc:PeriodOfTime rdf:nodeID="N4229e255693543359f8bdad8d23747c4">
    <schema:startDate rdf:datatype="http://www.w3.org/2001/XMLSchema#dateTime">2019-07
    <schema:endDate rdf:datatype="http://www.w3.org/2001/XMLSchema#dateTime">2020-10-3
  </dc:PeriodOfTime>
</dc:temporal>
<dc:theme rdf:resource="http://inspire.ec.europa.eu/theme/ac"/>
<dc:keyword>TRAFair</dc:keyword>
<dc:keyword>prediction_gral_nox</dc:keyword>
```

Deployment of Open Data in Zaragoza, Santiago and Modena

- **A solution based on Open Source software packages**
 - GeoServer: management of spatial data
 - CKAN: platform for deploying Open Data portals



4. The challenge of quality

- **Standardization efforts include test suites to assure**
 - completeness (commission or omission of metadata elements)
 - consistency (compliance with metadata format and structure/domain of metadata elements)
- **However, less attention is paid to accuracy (“accurate description of resources using factual and correct information”)**
 - Should a catalog maintain metadata records incorrectly classified (*boundaries map* classified as containing *land use data*)?
 - Should a resource locator link to an unreachable and defunct web site?
- **Why not so accurate?**
 - Geographic metadata has been a mandate for SDIs
 - , but not essential for final users that visualize/render these digital assets through online map clients, Open Street Map, Google Maps, ...

```
<meta name="HTTP-EQUIV" data-bbox="970 560 985 790"/>  
</map />  
</area alt="content" data-bbox="915 560 930 990"/>  
<meta name="Description" data-bbox="855 560 870 990"/>  
<meta name="Section" data-bbox="885 560 900 990"/>  
<meta name="Description" data-bbox="805 560 820 990"/>  
<meta name="Some entities" data-bbox="795 560 810 990"/>  
<img alt="Urgent rendering, let's take care" data-bbox="835 560 850 990"/>  
<meta name="Description" data-bbox="745 560 760 990"/>  
<meta name="Section" data-bbox="735 560 750 990"/>  
<meta name="Some entities" data-bbox="725 560 740 990"/>  
<img alt="Urgent rendering, let's take care" data-bbox="745 560 760 990"/>  
<meta name="Description" data-bbox="705 560 720 990"/>  
<meta name="Section" data-bbox="695 560 710 990"/>  
<meta name="Some entities" data-bbox="685 560 700 990"/>  
<img alt="Urgent rendering, let's take care" data-bbox="705 560 720 990"/>
```


Some approaches to verify the quality of metadata

- ## European Data Portal's Metadata Quality Assessment Methodology

Dimension	Indicator
Findability	Keywords available (Dataset/keyword)
	Category available (Dataset/theme)
	Spatial information available (Dataset/spatial)
	Temporal information available (Dataset/temporal)
Accessibility	Most frequent AccessURL status code=200 (Distribution/accessURL)
	DownloadURL available (Distribution/downloadURL)
	Most frequent DownloadURL status code=200 (Distribution/downloadURL)
Interoperability	Format available (Distribution/format)
	Media type available (Distribution/mediaType)
	Format/ media type from vocabulary (Distribution/format or Distribution/mediaType)
	Non-proprietary (Distribution/format or Distribution/mediaType)
	Machine readable (Distribution/format or Distribution/mediaType)
	DCAT-AP compliance (all entities and properties)
Reusability	License available (Distribution/license)
	License from vocabulary (Distribution/license)
	Access rights available (Dataset/accessRights)
	Access rights from vocabulary (Dataset/accessRights)
	Contact point available (Dataset/contactPoint)
	Publisher available (Dataset/publisher)
Contextuality	Rights available (Distribution/rights)
	File size available (Distribution/byteSize)
	Issued date available (Dataset/issued or Distribution/issued)
	Modified date available (Dataset/modified or Distribution/modified)

Some approaches to verify the quality of metadata

- **Adaptation of ISO 19157 “Geographic Information – Data Quality” to Open Data metadata**
 - Accuracy and correctness of temporal, positional, and non-quantitative attribute information are covered

Quality category	Quality element
DQ_Completeness	DQ_Completeness Commission
	DQ_CompletenessOmission
DQ_LogicalConsistency	DQ_ConceptualConsistency
	DQ_DomainConsistency
	DQ_FormatConsistency
	DQ_TopologicalConsistency
DQ_TemporalQuality	DQ_TemporalConsistency
	DQ_TemporalValidity
DQ_ThematicAccuracy	DQ_ThematicClassificationCorrectness
	DQ_NonQuantitativeAttribute Correctness
	DQ_PositionalCorrectness
	DQ_QualityOfFreeText

datos.gob.es (2019)

Sample based quality control

Pass Fail

E	●	●	DQ_TheClasCorDatThet_CR
	●	●	DQ_TheNQADatRef_CR
	●	●	DQ_TheNQADatCon_CR
F	●	●	DQ_TheNQADisAcc_CR
	●	●	DQ_TheNQADisLic_CR
	●	●	DQ_QFTDatTitO_CR
	●	●	DQ_QFTDatDesO_CR

References

- F.J. Zarazaga-Soria, J. Nogueras-Iso, M. Ford, 2003. Guidance material for mapping between Dublin Core and ISO in the Geographic Information domain. CWA 14856, CEN/ISSS Workshop– metadata for multimedia information–Dublin Core.
- J. Nogueras-Iso, F.J. Zarazaga-Soria, J. Lacasta, R. Béjar, P.R. Muro-Medrano, 2004. Metadata Standard Interoperability: Application in the Geographic Information Domain. *Computers, Environment and Urban Systems*, 28: 611-634.
- D. Nebert, A. Whiteside, P. Vretanos (ed), 2007. OpenGIS Catalogue Services Specification, versión 2.02. Open Geospatial Consortium Inc.
- European Commission, 2016. GeoDCAT Application profile for data portals in Europe, GeoDCAT-AP v1.0.1, <https://joinup.ec.europa.eu/release/geodcat-ap/101>
- J. Nogueras-Iso, H. Ochoa-ortiz, M.A. Jañez, J.R.R. Viqueira, L. Po, R. Trillo-Lado, 2020. Automatic publication of Open Data from OGC services: the use case of TRAFAIR project. The Twelfth International Conference on Advanced Geographic Information Systems, Applications, and Services, GEOProcessing 2020, Valencia (Spain), 21-25 November 2020.
- Publications Office of the European Union, 2020. Metadata Quality Assessment Methodology. How EDP measures the quality of Harvested Metadata. <https://www.europeandataportal.eu/mqa/methodology>
- J. Nogueras-Iso, J. Lacasta, M.A. Ureña-Cámara, F.J. Ariza-López, 2021. Quality of Metadata in Open Data Portals. *IEEE Access*, 9: 60364-60382.

