

¿Tenemos en cuenta la Calidad a la hora de documentar los datos?

Paula Díaz, Joan Masó

Análisis aplicado a los
metadatos distribuidos por
la IDEC



GEOVIQUA

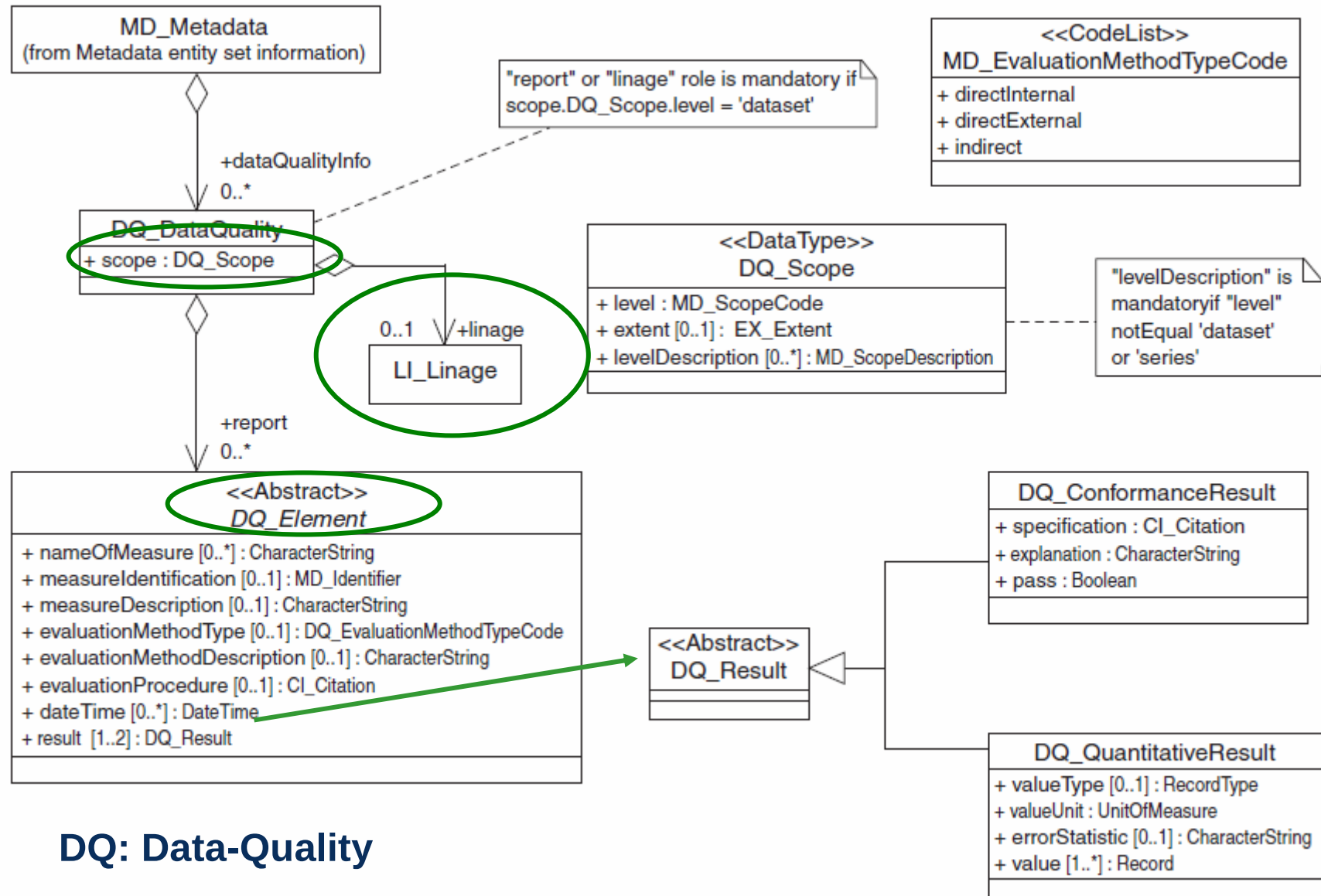
QUALity aware Visualisation for the Global Earth
Observation system of systems

- Introducción y contextualización del estudio
- Metodología y materiales
- Análisis de los elementos de calidad en los metadatos (IDEC)
- Conclusiones

- Análisis previos:
 - Díaz P., Masó J., Zabala, A. (2009). 'Análisis comparativo de los metadatos distribuidos por la IDEC y la IDECLM como ejemplos de IDE autonómicas'. *Jornadas Técnicas de la IDE de España, JIDEE*. Murcia.
 - Díaz P., Masó J., Guimet J. (2010) 'Comparative Quality Assessment of Metadata. Two Regional SDI case studies'. *INSPIRE Conference*. Krakow.
- Análisis de la calidad descrita en los metadatos distribuidos por la IDEC.
- GeoViQua: *QUAlity aware VIsualisation for the Global Earth Observation system of systems*. Financiado por la EC
 - GEOSS Common Infrastructure (GCI)
 - Calidad de los datos: Búsqueda, Visualización y Descarga de datos.

- Documentos de MD de la IDEC de 2010 en XML: 14616
- Extracción masiva de los elementos de Calidad de los metadatos para su análisis (ISO 19115).
 - *Report* (clases o elementos de calidad)
 - *Linage*
 - *Scope*
- ¿Cómo? Con una aplicación de MiraMon (GestBD) para la extracción del Xpath de los documentos de XML

Esquema de la información de Calidad ISO 19115:2003



DQ: Data-Quality

- **GESTBD**

- For /L %%g in (0,1,15) do C:\Miramon\GestBD_24 MD_L.dbf "UUID_2010\00029829-eadf-11de-b435-b36c953320f5.XML" * FileID "/MD_Metadata/fileIdentifier/gco:CharacterString" V * statement
"/MD_Metadata/dataQualityInfo/DQ_DataQuality/lineage/LI_Lineage/statement[0]/gco:CharacterString" V * description
"/MD_Metadata/dataQualityInfo/DQ_DataQuality/lineage/LI_Lineage/processStep/LI_ProcessStep[%%g]/description/gco:CharacterString" V * dateTime
"/MD_Metadata/dataQualityInfo/DQ_DataQuality/lineage/LI_Lineage/processStep/LI_ProcessStep[%%g]/dateTime/gco:DateTime" V * rationale
"/MD_Metadata/dataQualityInfo/DQ_DataQuality/lineage/LI_Lineage/processStep/LI_ProcessStep[%%g]/rationale[0]/gco:CharacterString" V * processor
"/MD_Metadata/dataQualityInfo/DQ_DataQuality/lineage/LI_Lineage/processStep/LI_ProcessStep[%%g]/processor[0]/CI_ResponsibleParty/*" N * rationale
"/MD_Metadata/dataQualityInfo/DQ_DataQuality/lineage/LI_Lineage/processStep/LI_ProcessStep[%%g]/rationale[0]/gco:CharacterString" V * description
"/MD_Metadata/dataQualityInfo/DQ_DataQuality/lineage/LI_Lineage/source/LI_Source[%%g]/description[0]/gco:CharacterString" V * denominator
"/MD_Metadata/dataQualityInfo/DQ_DataQuality/lineage/LI_Lineage/source/LI_Source[%%g]/scaleDenominator[0]/MD_RepresentativeFraction/denominator[0]/gco:Integer" V * sourceCitation
"/MD_Metadata/dataQualityInfo/DQ_DataQuality/lineage/LI_Lineage/source/LI_Source[%%g]/sourceCitation[0]/CI_Citation/title/gco:CharacterString" V * CI_Date
"/MD_Metadata/dataQualityInfo/DQ_DataQuality/lineage/LI_Lineage/source/LI_Source[%%g]/sourceCitation[0]/CI_Citation/date/CI_Date/date/*" N * Extentdesc
"/MD_Metadata/dataQualityInfo/DQ_DataQuality/lineage/LI_Lineage/source/LI_Source[%%g]/sourceExtent[0]/EX_Extent/description/*" N * GeoElement
"/MD_Metadata/dataQualityInfo/DQ_DataQuality/lineage/LI_Lineage/source/LI_Source[%%g]/sourceExtent[0]/EX_Extent/geographicElement/*" N * temporalElement
"/MD_Metadata/dataQualityInfo/DQ_DataQuality/lineage/LI_Lineage/source/LI_Source[%%g]/sourceExtent[0]/EX_Extent[0]/temporalElement/*" N * verticalElement
"/MD_Metadata/dataQualityInfo/DQ_DataQuality/lineage/LI_Lineage/source/LI_Source[%%g]/sourceExtent[0]/EX_Extent[0]/verticalElement/*" N /SAC /MUT

- **GESTBD**

- For /L %%g in (0,1,15) do C:\Miramon\GestBD_24 MD_L.dof "UUID_2010\00029829-eadf-11de-b435-b36c953320f5.XML" * FileID "/MD_Metadata/fieldIdentifier/gco:CharacterString" V * statement
"/MD_Metadata/dataQualityInfo/DQ_DataQuality/lineage/LI_Lineage/statement[0]/gco:CharacterString" V * description
"/MD_Metadata/dataQualityInfo/DQ_DataQuality/lineage/LI_Lineage/processStep/LI_ProcessStep[%%g]/description/gco:CharacterString" V * dateTime
"/MD_Metadata/dataQualityInfo/DQ_DataQuality/lineage/LI_Lineage/processStep/LI_ProcessStep[%%g]/dateTime/gco:DateTime" V * rationale
"/MD_Metadata/dataQualityInfo/DQ_DataQuality/lineage/LI_Lineage/processStep/LI_ProcessStep[%%g]/rationale[0]/gco:CharacterString" V * processor
"/MD_Metadata/dataQualityInfo/DQ_DataQuality/lineage/LI_Lineage/processStep/LI_ProcessStep[%%g]/processor[0]/CI_ResponsibleParty/*" N * rationale
"/MD_Metadata/dataQualityInfo/DQ_DataQuality/lineage/LI_Lineage/processStep/LI_ProcessStep[%%g]/rationale[0]/gco:CharacterString" V * description
"/MD_Metadata/dataQualityInfo/DQ_DataQuality/lineage/LI_Lineage/source/LI_Source[%%g]/description[0]/gco:CharacterString" V * denominator
"/MD_Metadata/dataQualityInfo/DQ_DataQuality/lineage/LI_Lineage/source/LI_Source[%%g]/scaleDenominator[0]/MD_RepresentativeFraction/denominator[0]/gco:Integer" V * sourceCitation
"/MD_Metadata/dataQualityInfo/DQ_DataQuality/lineage/LI_Lineage/source/LI_Source[%%g]/sourceCitation[0]/CI_Citation/title/gco:CharacterString" V * CI_Date
"/MD_Metadata/dataQualityInfo/DQ_DataQuality/lineage/LI_Lineage/source/LI_Source[%%g]/sourceCitation[0]/CI_Citation/date/CI_Date/date/*" N * Extentdesc
"/MD_Metadata/dataQualityInfo/DQ_DataQuality/lineage/LI_Lineage/source/LI_Source[%%g]/sourceExtent[0]/EX_Extent/description/*" N * GeoElement
"/MD_Metadata/dataQualityInfo/DQ_DataQuality/lineage/LI_Lineage/source/LI_Source[%%g]/sourceExtent[0]/EX_Extent/geographicElement/*" N * temporalElement
"/MD_Metadata/dataQualityInfo/DQ_DataQuality/lineage/LI_Lineage/source/LI_Source[%%g]/sourceExtent[0]/EX_Extent[0]/temporalElement/*" N * verticalElement
"/MD_Metadata/dataQualityInfo/DQ_DataQuality/lineage/LI_Lineage/source/LI_Source[%%g]/sourceExtent[0]/EX_Extent[0]/verticalElement/*" N /SAC /MUT

- **GESTBD**

- For /L %%g in (0,1,15) do C:\Miramon\GestBD_24 MD_L.dof "UUID_2010\00029829-eadf-11de-b435-b36c953320f5.XML" * FileID "/MD_Metadata/fieldIdentifier/gco:CharacterString" V * statement
"/MD_Metadata/dataQualityInfo/DQ_DataQuality/lineage/LI_Lineage/statement[0]/gco:CharacterString" V * description
"/MD_Metadata/dataQualityInfo/DQ_DataQuality/lineage/LI_Lineage/processStep/LI_ProcessStep[%%g]/description/gco:CharacterString" V * dateTime
"/MD_Metadata/dataQualityInfo/DQ_DataQuality/lineage/LI_Lineage/processStep/LI_ProcessStep[%%g]/dateTime/gco:DateTime" V * rationale
"/MD_Metadata/dataQualityInfo/DQ_DataQuality/lineage/LI_Lineage/processStep/LI_ProcessStep[%%g]/rationale[0]/gco:CharacterString" V * processor
"/MD_Metadata/dataQualityInfo/DQ_DataQuality/lineage/LI_Lineage/processStep/LI_ProcessStep[%%g]/processor[0]/CI_ResponsibleParty/*" N * rationale
"/MD_Metadata/dataQualityInfo/DQ_DataQuality/lineage/LI_Lineage/processStep/LI_ProcessStep[%%g]/rationale[0]/gco:CharacterString" V * description
"/MD_Metadata/dataQualityInfo/DQ_DataQuality/lineage/LI_Lineage/source/LI_Source[%%g]/description[0]/gco:CharacterString" V * denominator
"/MD_Metadata/dataQualityInfo/DQ_DataQuality/lineage/LI_Lineage/source/LI_Source[%%g]/scaleDenominator[0]/MD_RepresentativeFraction/denominator[0]/gco:Integer" V * sourceCitation
"/MD_Metadata/dataQualityInfo/DQ_DataQuality/lineage/LI_Lineage/source/LI_Source[%%g]/sourceCitation[0]/CI_Citation/title/gco:CharacterString" V * CI_Date
"/MD_Metadata/dataQualityInfo/DQ_DataQuality/lineage/LI_Lineage/source/LI_Source[%%g]/sourceCitation[0]/CI_Citation/date/CI_Date/date/*" N * Extentdesc
"/MD_Metadata/dataQualityInfo/DQ_DataQuality/lineage/LI_Lineage/source/LI_Source[%%g]/sourceExtent[0]/EX_Extent/description/*" N * GeoElement
"/MD_Metadata/dataQualityInfo/DQ_DataQuality/lineage/LI_Lineage/source/LI_Source[%%g]/sourceExtent[0]/EX_Extent/geographicElement/*" N * temporalElement
"/MD_Metadata/dataQualityInfo/DQ_DataQuality/lineage/LI_Lineage/source/LI_Source[%%g]/sourceExtent[0]/EX_Extent[0]/temporalElement/*" N * verticalElement
"/MD_Metadata/dataQualityInfo/DQ_DataQuality/lineage/LI_Lineage/source/LI_Source[%%g]/sourceExtent[0]/EX_Extent[0]/verticalElement/*" N /SAC /MUT

- Obtenemos 3 tablas DBF que contienen la información seleccionada
 - Facilidad de análisis
 - Información clasificada por UUID

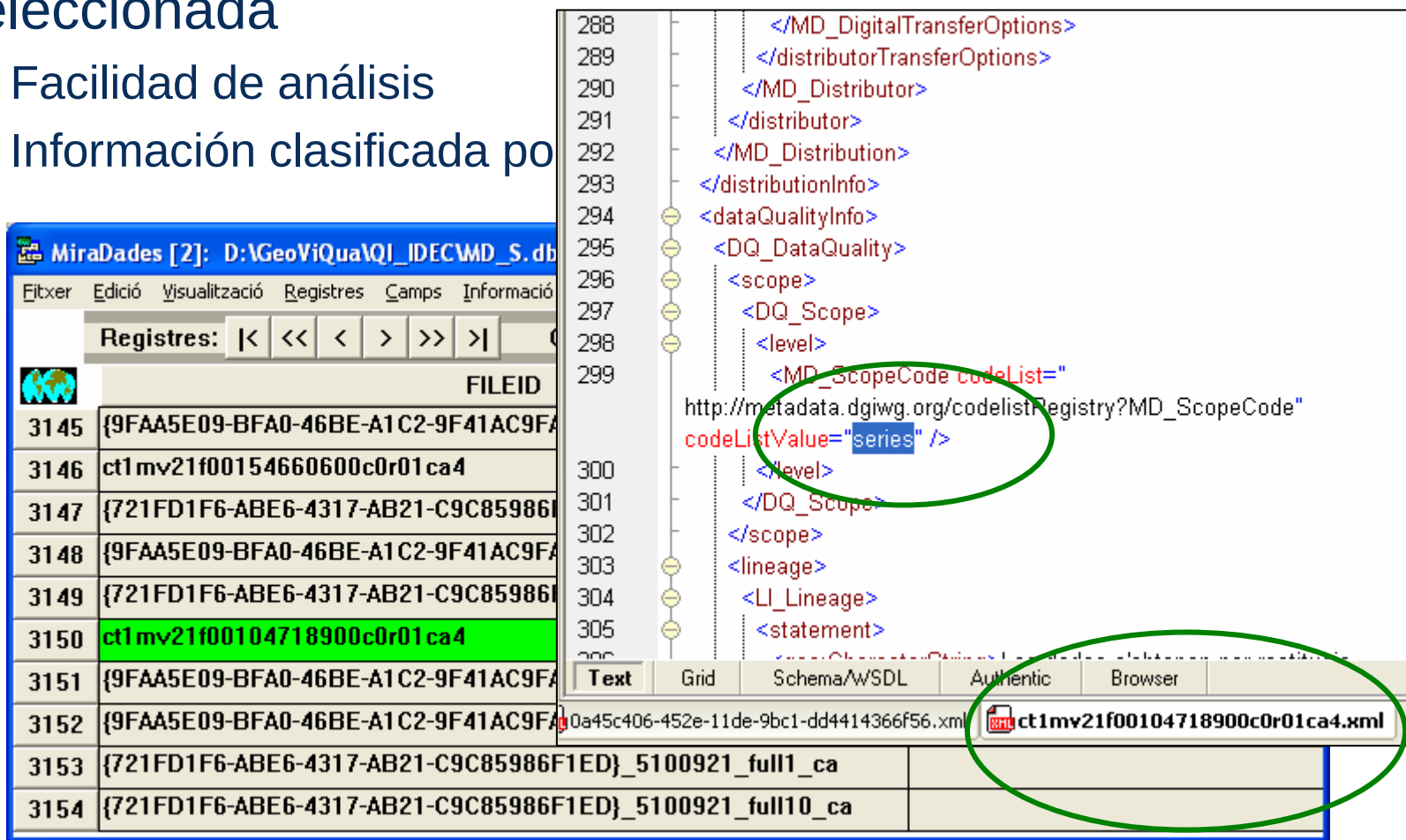
MiraDades [2]: D:\GeoViQua\QI_IDEC\WD_S.dbf

Fitxer Edició Visualització Registres Camps Informació Eines Ajuda

Registres: |< << < > >> >| Camps: |< << < > >> >|

	FILEID	SCOPECODE1
3145	{9FAA5E09-BFA0-46BE-A1C2-9F41AC9FAB7B}_31200004_full16_ca	series
3146	ct1mv21f00154660600c0r01ca4	series
3147	{721FD1F6-ABE6-4317-AB21-C9C85986F1ED}_100132_ca	dataset
3148	{9FAA5E09-BFA0-46BE-A1C2-9F41AC9FAB7B}_31200004_full17_ca	series
3149	{721FD1F6-ABE6-4317-AB21-C9C85986F1ED}_100214_ca	dataset
3150	ct1mv21f00104718900c0r01ca4	series
3151	{9FAA5E09-BFA0-46BE-A1C2-9F41AC9FAB7B}_31200004_full18_ca	series
3152	{9FAA5E09-BFA0-46BE-A1C2-9F41AC9FAB7B}_11400007_full6_ca	dataset
3153	{721FD1F6-ABE6-4317-AB21-C9C85986F1ED}_5100921_full1_ca	
3154	{721FD1F6-ABE6-4317-AB21-C9C85986F1ED}_5100921_full10_ca	

- Obtenemos 3 tablas DBF que contienen la información seleccionada
 - Facilidad de análisis
 - Información clasificada por



The screenshot shows the MiraDades [2] application interface. The main window displays a table of records with columns 'FILEID' and a long alphanumeric string. Record 3150 is highlighted in green. To the right, an XML view of the selected record is shown, with a green circle highlighting the `<MD_ScopeCode codeList="http://metadata.dgiwg.org/codelistRegistry?MD_ScopeCode" codeListValue="series" />` element. Another green circle highlights the file path `ct1mv21f00104718900c0r01ca4.xml` at the bottom of the XML view.

FILEID	Record
3145	{9FAA5E09-BFA0-46BE-A1C2-9F41AC9FA}
3146	ct1mv21f00154660600c0r01ca4
3147	{721FD1F6-ABE6-4317-AB21-C9C85986F1ED}
3148	{9FAA5E09-BFA0-46BE-A1C2-9F41AC9FA}
3149	{721FD1F6-ABE6-4317-AB21-C9C85986F1ED}
3150	ct1mv21f00104718900c0r01ca4
3151	{9FAA5E09-BFA0-46BE-A1C2-9F41AC9FA}
3152	{9FAA5E09-BFA0-46BE-A1C2-9F41AC9FA}
3153	{721FD1F6-ABE6-4317-AB21-C9C85986F1ED}_5100921_full1_ca
3154	{721FD1F6-ABE6-4317-AB21-C9C85986F1ED}_5100921_full10_ca

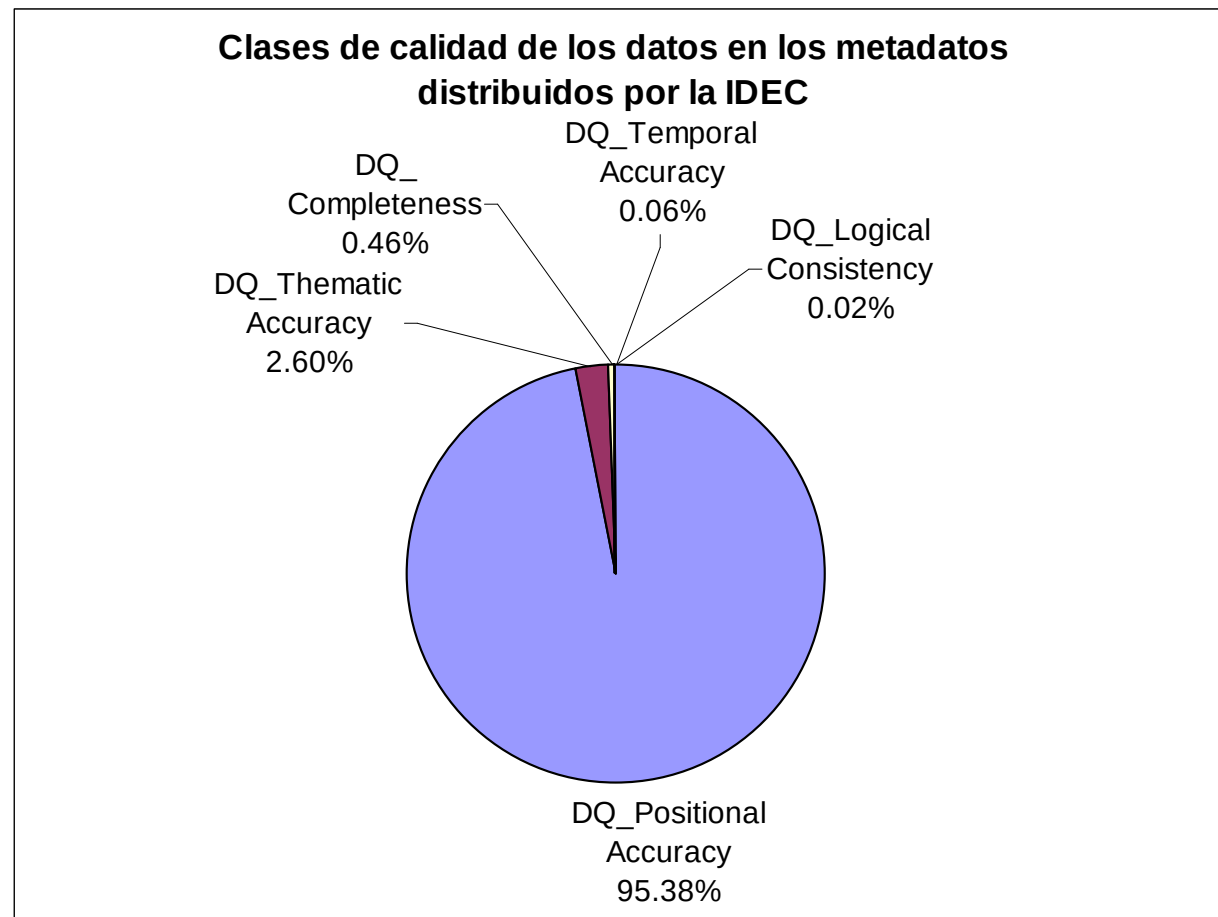
Análisis de los elementos de calidad en los metadatos (IDEC)

- Clasificación inicial de los metadatos
 - No describen la calidad del dato
 - La descripción es incorrecta
 - Cuantitativos
 - Cualitativos
 - Sigue una especificación



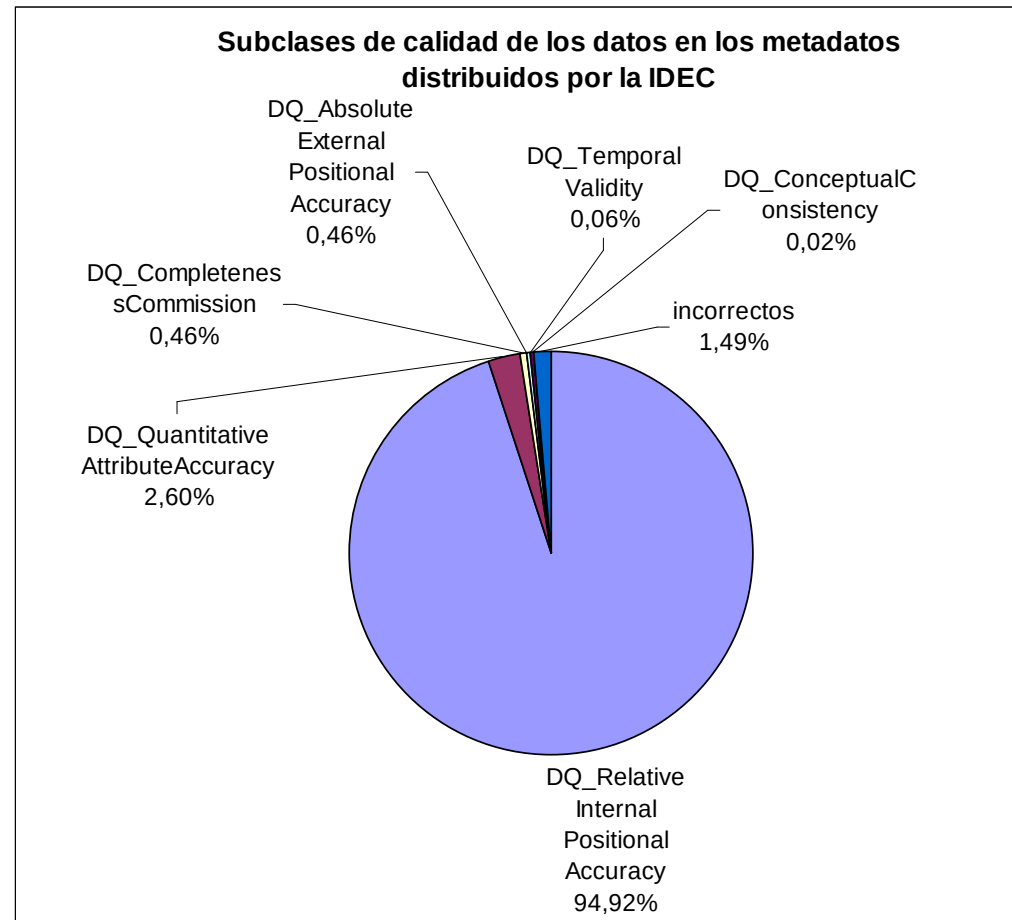
Análisis de los elementos de calidad en los metadatos (IDEC)

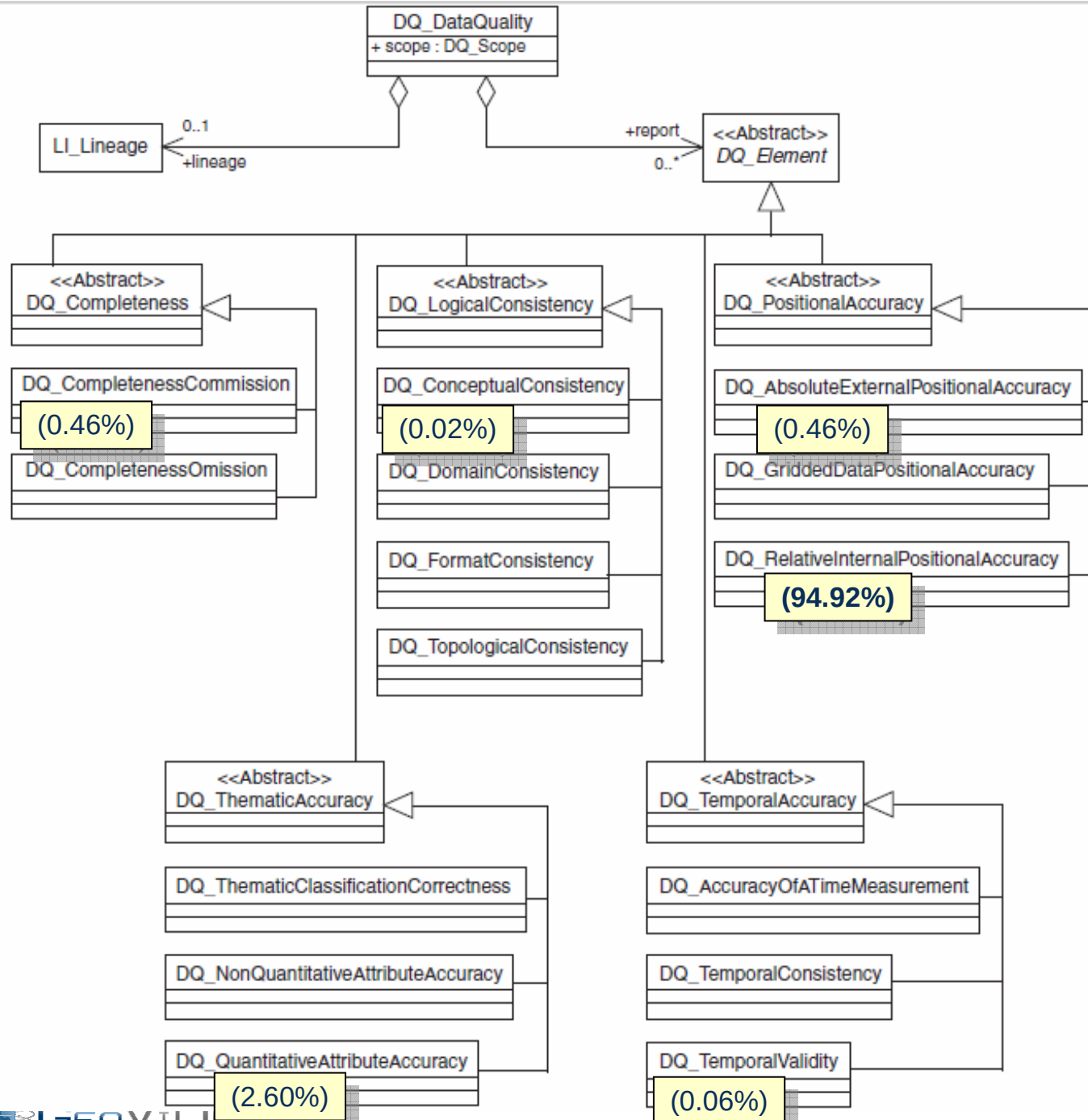
- Clasificación de la información de calidad según **clases**



Análisis de los elementos de calidad en los metadatos (IDEC)

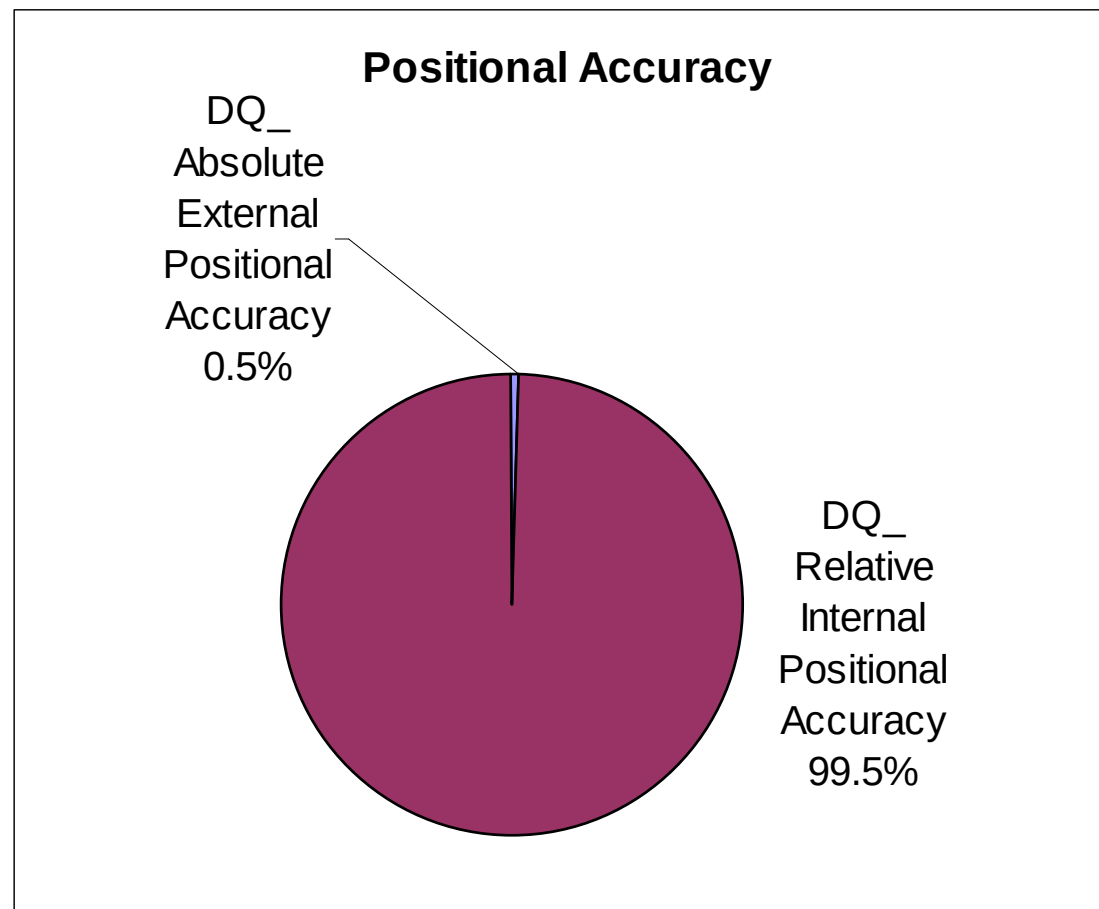
- Clasificación de la información de calidad según **subclases**





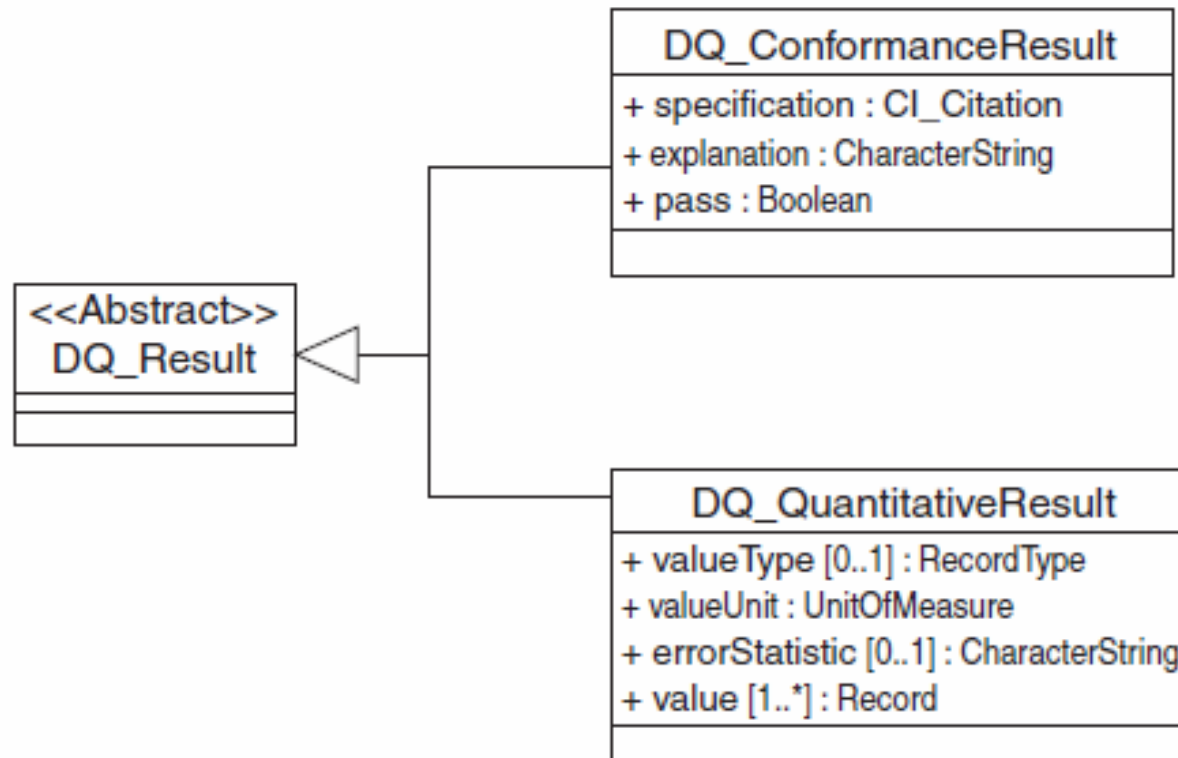
Análisis de los elementos de calidad en los metadatos (IDEC)

- La **clase** más representada: Exactitud posicional (96.35%)



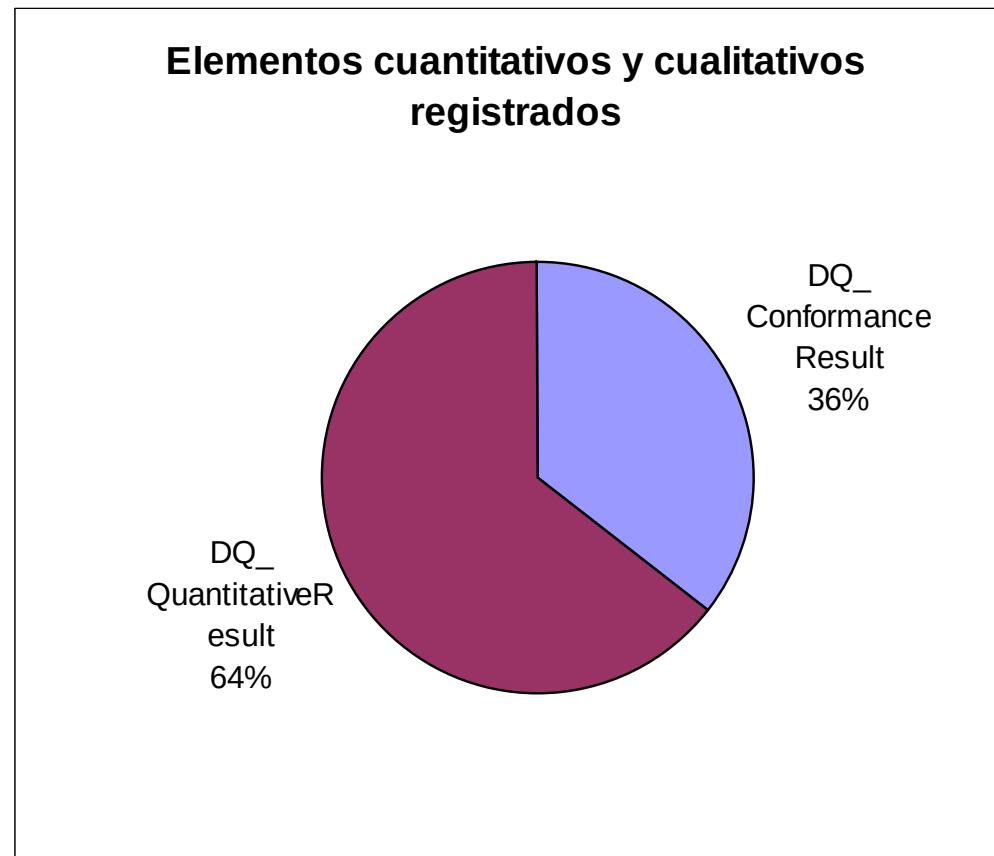
Análisis de los elementos de calidad en los metadatos (IDEC)

- Resultados Cuantitativos vs. Cualitativos

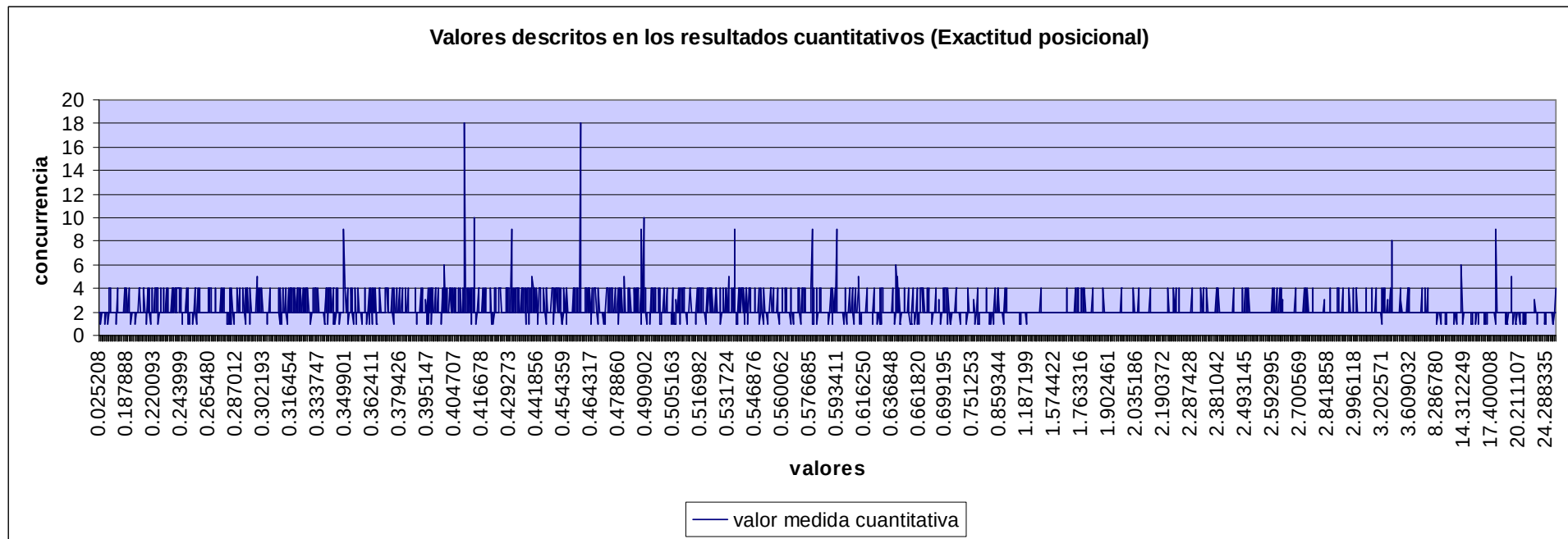


Análisis de los elementos de calidad en los metadatos (IDEC)

- Resultados Cuantitativos vs. Cualitativos
 - Los resultados cuantitativos se registran en un 64% frente al 36% de los cualitativos

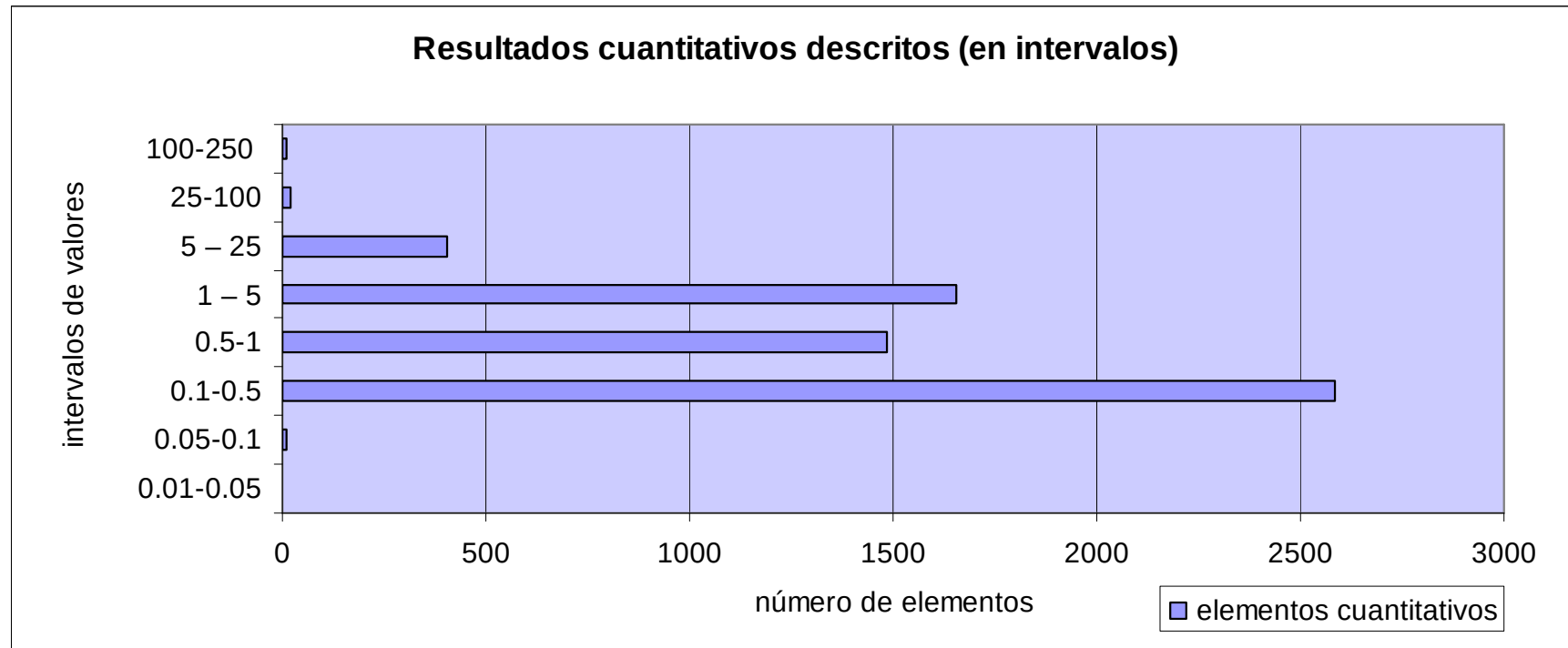


- Los datos **cuantitativos**:
 - Los más representados: Exactitud posicional (96.35%)



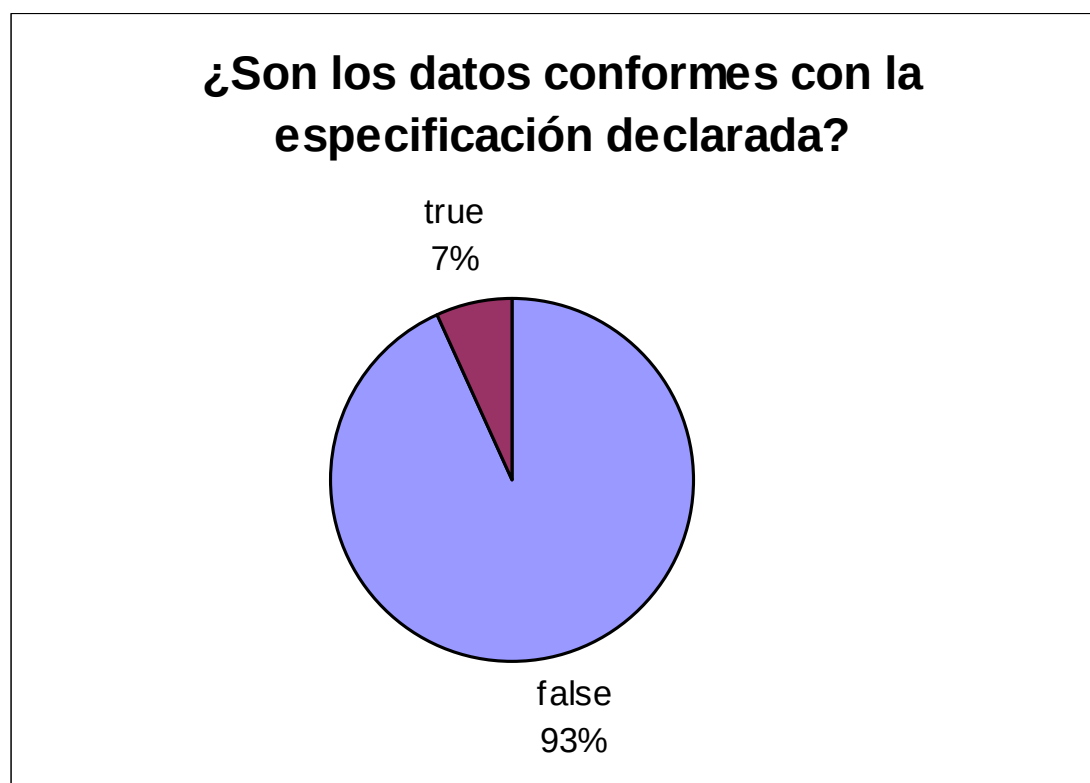
Análisis de los elementos de calidad en los metadatos (IDEC)

- Los datos **cuantitativos**:
 - Los más representados: Exactitud posicional (96.35%)



Análisis de los elementos de calidad en los metadatos (IDEC)

- Datos Cualitativos. Conformidad de los resultados
 - Los resultados cualitativos no superan, en su mayoría, la especificación declarada



Análisis de los elementos de calidad en los metadatos (IDEC)

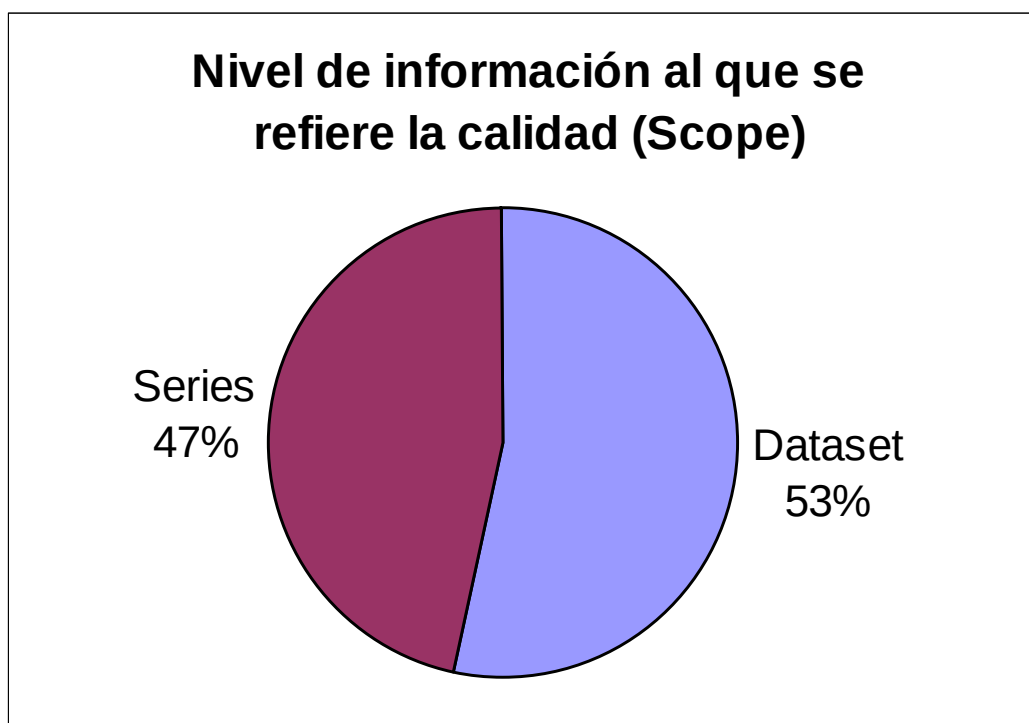
- Niveles descritos en la especificación ISO 19115: 2003

B.5.25 MD_ScopeCode <<CodeList>>

	Name	Domain code	Definition
1.	MD_ScopeCode	ScopeCd	class of information to which the referencing entity applies
2.	attribute	001	information applies to the attribute class
3.	attributeType	002	information applies to the characteristic of a feature
4.	collectionHardware	003	information applies to the collection hardware class
5.	collectionSession	004	information applies to the collection session
6.	dataset	005	information applies to the dataset
7.	series	006	information applies to the series
8.	nonGeographicDataset	007	information applies to non-geographic data
9.	dimensionGroup	008	information applies to a dimension group
10.	feature	009	information applies to a feature
11.	featureType	010	information applies to a feature type
12.	propertyType	011	information applies to a property type
13.	fieldSession	012	information applies to a field session
14.	software	013	information applies to a computer program or routine
15.	service	014	information applies to a capability which a service provider entity makes available to a service user entity through a set of interfaces that define a behaviour, such as a use case
16.	model	015	information applies to a copy or imitation of an existing or hypothetical object
17.	tile	016	information applies to a tile, a spatial subset of geographic data

Análisis de los elementos de calidad en los metadatos (IDEC)

- Nivel de información de la calidad descrita



- La **calidad** de los datos se describe en un 22% de los metadatos aproximadamente
- Los elementos más descritos corresponden a la exactitud **posicional** (95% aprox. de la info. descrita).
- La calidad **temática** se describe mínimamente (3% aprox.)
- Errores en la descripción de la calidad (1.5% aprox.)
 - Unidades de la medida del error no descrita (99%)
 - Afirman no superar la especificación de conformidad (93%)
 - Medida incompleta o mal descrita (1.5%)
 - Medidas de clase incorrecta (e.g. DQ_TemporalAccuracy: “Restitucio Fotogrametrica Numerica a escala 1:25.000”)
 - Descripción de calidad No del dato (e.g. Concentración de dioxinas)

- Es necesario incentivar a los productores a informar sobre las medidas de calidad que acompañan a los datos
 - La calidad de los datos es importante para determinar la idoneidad de un conjunto de datos a la aplicación concreta, diferente de la originalmente pensada.
- Los errores son debidos a falta de información
 - Son necesarias herramientas que faciliten la tarea de informar de la calidad
 - Tutoriales
 - Sistematización del registro de la calidad
- El proyecto GeoViQua ayudará a acercar la información de la calidad al usuario
 - Visualización
 - Búsqueda y descarga



UAB
Universitat Autònoma
de Barcelona

**¡Gracias!
Gràcies!
Eskerrik asko!
Graças!
Obrigado!**



GEOVIQUA

QUALity aware Visualisation for the Global Earth
Observation system of systems