



Instituto Geográfico Nacional
CENTRO NACIONAL DE INFORMACIÓN GEOGRÁFICA

Aportaciones sobre la evaluación normalizada de la calidad de metadatos



Antonio F. Rodríguez

Paloma Abad, Alejandra Sánchez, Laura Alemany

www.ign.es



JIIDE 2019



Cáceres, 23/25 octubre 2019



- 
- 
1. Utilidad de los metadatos
 2. Calidad de metadatos
 3. Artículo UJA y UNIZAR
 4. Un estudio sobre el CODSI
- Metadatos de datos
 5. Conclusiones

- Objetivos
 1. Encontrar recursos, saber que existen
 2. Compararlos y seleccionar el que más nos conviene
 3. Utilizarlos correctamente
 4. Para gestionarlos (organizarlos y mantenerlos) en una IDE
- Por lo que resultan claves en la gestión de una IDE
- Muy útiles para generar los indicadores de seguimiento INSPIRE
- Pero ¿los usamos los productores? ¿los utilizan los usuarios? ¿les resultan útiles?



- Punto de partida
 - «*A method for checking the quality of geographic metadata based on ISO 19157*» de M. A. Ureña y F.J. Ariza (U. de Jaén) y J. Nogueras y J. Lacasta (U. de Zaragoza)
 - IJGIS, Vol. 33, 2018 Issue 1
- Propone un método de aplicación de ISO 19157:2013 Calidad a los metadatos
 - *Sobre Q de metadatos de IG hay reglas y directrices INSPIRE, ANZLIC, FGDC...*
 - *No hay un marco general y completo para evaluar la CdM*

- DQ_Completeness
 - DQ_CompletenessCommission
 - DQ_CompletenessOmission
- DQ_LogicalConsistency
 - DQ_ConceptualConsistency
 - DQ_DomainConsistency
 - DQ_FormatConsistency
 - DQ_TopologicalConsistency
- DQ_TemporalAccuracy
 - DQ_TemporalConsistency
 - DQ_TemporalValidity
- DQ_ThematicAccuracy
 - DQ_ThematicClassificationCorrectness
 - DQ_NonQuantitativeAttributeCorrectness
- DQ_PositionalCorrectness
- DQ_QualityOfFreeText

- No tiene en cuenta las Normas de Ejecución NSPIRE
- Algunos elementos
 - Inspección completa
 - Muestreo ISO 2859-2, **un lote**, AQL (Límite de Calidad Aceptable) = 4,16 % reg. erróneos
95, 84 % reg. correctos
- Métodos de medida
- Aplicación a los 4824 registros del catálogo de la IDEE (2016)
 - 3640 de CD hoy: 6324 CD y series
 - 1184 de servicios hoy: 4067 servicios10 391 en total (CODSI ha mejorado)

- Resultados:

CRITERIOS	PASAN	NO PASAN
26	9	17

- DQ_PositionalCorrectness

- Se verifica si la Bbox de los metadatos contiene el punto devuelto por Geonames al buscar las localizaciones textuales en los metadatos

7,25 % No pasa

- DQ_QualityOfFreeText (*abstract, purpose, lineage*)

- Calidad global

- Evaluada por un panel de expertos, de 1 a 5, sobre una muestra de 80/4824
- No definido, mala calidad o regular (42 %, 75 %, 73 %)

No pasa

- Índice de legibilidad (Fernández-Huerta, 1959)

- Función (Tasa palabras/frase, Tasa sílabas/palabra)
- Inspección completa (fallan 49 %, 41 %, 50 %)

No pasa

- DQ_Completeness
 - DQ_CompletenessCommission
 - DQ_CompletenessOmission
- DQ_LogicalConsistency
 - DQ_ConceptualConsistency
 - DQ_DomainConsistency
 - DQ_FormatConsistency
 - DQ_TopologicalConsistency
- DQ_TemporalAccuracy
 - DQ_TemporalConsistency
 - DQ_TemporalValidity
- DQ_ThematicAccuracy
 - DQ_ThematicClassificationCorrectness
 - DQ_NonQuantitativeAttributeCorrectness
- DQ_PositionalCorrectness
- DQ_QualityOfFreeText

- Análisis de las U. de Zaragoza y Jaén
 - El planteamiento es brillante
 - Algunos métodos admiten discusión
 - Pero concluyen que CdM es menor de lo esperado/deseado



- Catálogo Oficial de Datos y Servicios INSPIRE
 - Reglamento de Metadatos incluye 19 elementos para CDE
 - 207 CDE y 2 series (agosto 2019)
- 99,76 % pasa el test de CdM del validador INSPIRE
 - Cumplen las Normas de Ejecución INSPIRE
- 70 % pasa los test de interoperabilidad
 - Metadatos (CDE) → Serv. de visualización y descarga → Conformidad de datos, serv. y metadatos

- Hasta ahora se ha hecho un gran esfuerzo
 - Aspectos sintácticos y de conformidad (ISO 19115, INSPIRE)
 - Cumplimiento de normas ISO, Reglamentos europeos
- Pero la utilidad para usuario y productor creemos que depende también de
 - Aspectos semánticos

Elemento de CdM	INS	Verificación INSPIRE
DQ_CompletenessCommission	Sí	Que no haya elementos de cardinalidad sobreabundantes
DQ_CompletenessOmission	Sí	Que no falten elementos obligatorios o condicionales
DQ_ConceptualConsistency (<i>schema</i>)	Sí	Que se valida el XML Schema INSPIRE y el ISO 19115
DQ_FormatConsistency	Sí	Se verifica conformidad con ISO 19139
DQ_DomainConsistency	Parc	Se verifican listas controladas y tipos de datos
DQ_Domain Consistency (Extent)	Parc	Se verifica que esté en Europa y se hace un GetMap
DQ_ThematicClassificationCorrectness	Parc	Que sean datos/servicios, que estén las <i>keyword</i> recomendadas
DQ_NonQuantitativeAttributesCorrectness	Parc	E-mail, address, role, organization, responsible, title

Elementos de CdM no incluidos en el test INSPIRE

Elemento de CdM	INS	Verificación INSPIRE
DQ_ConceptualConsistency (scale)	No	Que no hay contradicciones entre la resolución y otros elementos
DQ_ConceptualConsistency (nº objects/dim)	No	Que no haya contradicciones
DQ_TemporalConsistency	No	Que las fechas están ordenadas (ceación, publicación, revisión...)
DQ_TopologicalConsistency	No	Consistencia Extent con otros elementos
DQ_TopologicalConsistency (<i>parent id</i>)	No	Que los parent ID apuntan a la serie
DQ_TemporalValidity	No	Que UpdateFrequency es coherente con updates y fecha actual
DQ_Positional Correctness (NUEVO)	No	Comprobación de Bbox con un punto de Geonames (literales)
DQ_QualityOfFreeText (NUEVO)	No	Calidad de <i>abstract</i> , <i>purpose</i> y <i>lineage</i> , por panel y fórmula

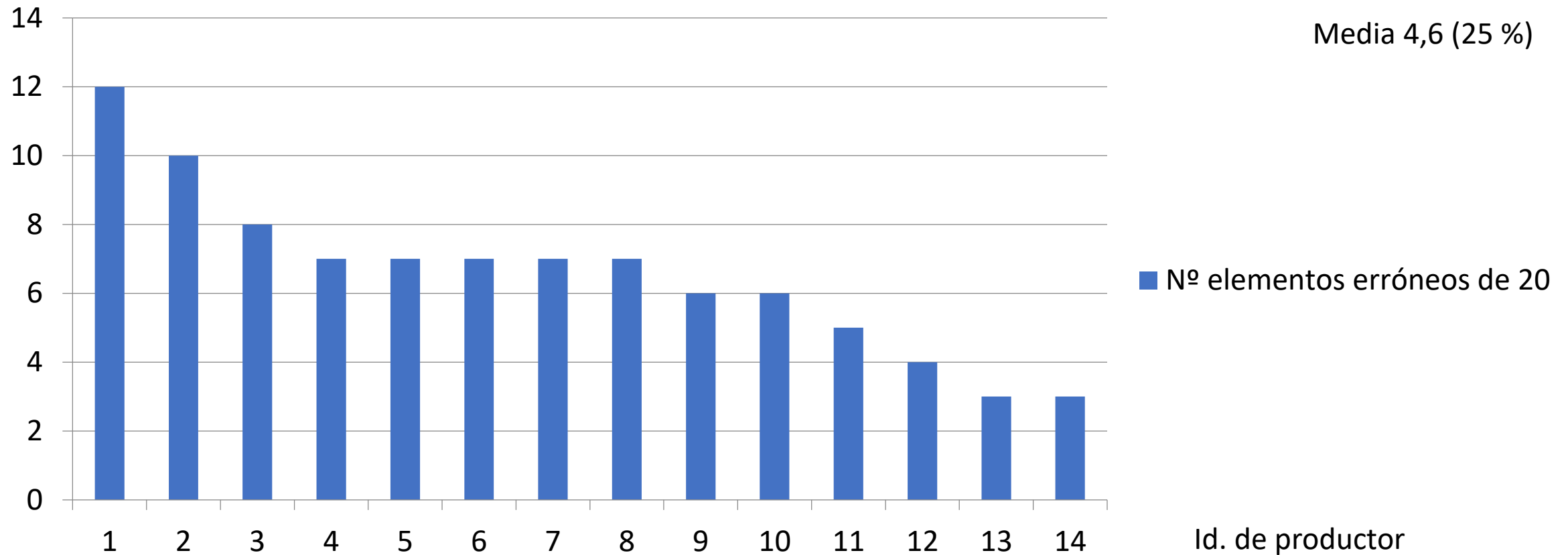
8 no

- La mayoría de los elementos de CdM no verificados por el test INSPIRE en el CODSI → Calidad de contenido
 - DQ_QuantitativeAttributesCorrectness
 - DQ_NonQuantitativeAttributesCorrectness
- Se han tomado 20 elementos y una muestra de 14 registros el CODSI A I y II
 - 1 registro por productor (14 productores de 31)
- Criterio: que los metadatos sean útiles para buscar, seleccionar, utilizar y gestionar recursos por usuarios y productores
 - He aplicado mis propios criterios semánticos

- Título que identifique
- Estandarizar
 - Resumen estandarizado: al menos Tema+ extensión + escala
 - Bbox
- Siempre:
 - Resolución
 - Frecuencia de actualización
 - Propósito
 - Linaje
 - Autor
- Fecha
 - Cuando los datos describen realidad
- E-mail
 - Institucional
 - Que responda
- Licencia
- Restricciones de uso
- Cuidado con el idioma
- Textos sin erratas
- Revisión semántica general

Nº de elementos erróneos (de 20) en un productor

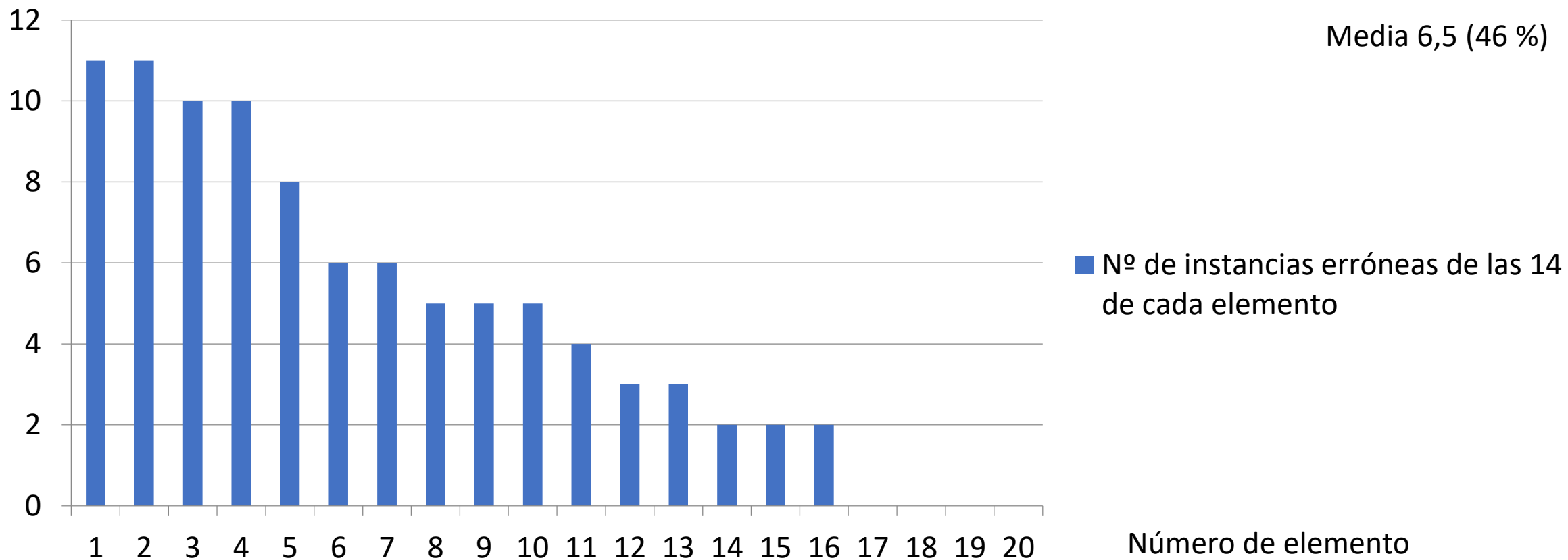
Nº elementos erróneos de 20





Nº de instancias erróneas de cada elemento

Nº de instancias erróneas de las 14 de cada elemento



- **11 Resumen** (tema, ámbito, resolución, err.)
- **11 Org. responsable** (author)
- **10 Validez temporal** (updateFrequency)
- **10 Condiciones de acceso** (licencia)
- **8 e-mail** (sin respuesta, personal)
- **6 idioma del recurso** (spa, glg, eus, cat)
- **6 Restricciones de acceso** (no limitations)
- **5 Título** (genérico, errores)
- **5 Propósito** (falta, erróneo)
- **5 Linaje** (falta, sin pasos)
- **4 Consistencia temporal** (fechas en orden)
- **3 Resolución** (falta, ppp)
- **3 Corrección posicional** (Bbox errónea)
- **2 Extensión** (poco ajustada)
- **2 idioma de los metadatos** (error)
- **2 Clasificación temática** (errónea)
- **0 Tipo de recurso** (serie, CDE)
- **0 Consistencia topológica**
- **0 Id. del recurso**
- **0 Fecha de los metadatos**

- En una muestra de 25 resúmenes
 - Hay 28 errores detectados por un procesador de texto
 - Media 1,2 errores/resumen
 - El 82 % de los resúmenes con algún error
- Además, poco legibles, desordenados y desenfocados

- Conclusiones

- Revisa tus metadatos
 - Normas ISO, Reglamento, *Guidelines*
- Mantenlos al día
- Utilízalos

- Problemas pendientes

- Multilingüismo
- Consistencia metadatos-capabilities-descargas
- Especialización (documentalistas)

- Medidas

- GTT Metadatos y catálogo

- Recomendación
- Formación
- Verificación

- Red de expertos

- Documentalistas
- Cartotecarios
- AGE + CC. AA. (~25)



¡Muchas gracias!

Antonio F. Rodríguez
afrodriguez@fomento.es