

JIDEE 2015, 6 de Noviembre de 2015, Sevilla

LA GESTIÓN Y LA TECNOLOGÍA DE LOS IDENTIFICADORES PERSISTENTES

Francisco J López-Pellicer, Jesús Barrera, Julián González
F. Javier Zarazaga-Soria, Emilio Romero, Paloma Abad

Objetivos de la charla

- ¿Qué es un PID?
- PID en INSPIRE
- HTTP URI como PID en INSPIRE
- Esquema Nacional de Interoperabilidad
- Impacto en HTTP URI de INSPIRE
- Caso de estudio
- Conclusión: lo mínimo para tener éxito

¿Qué es un PID?

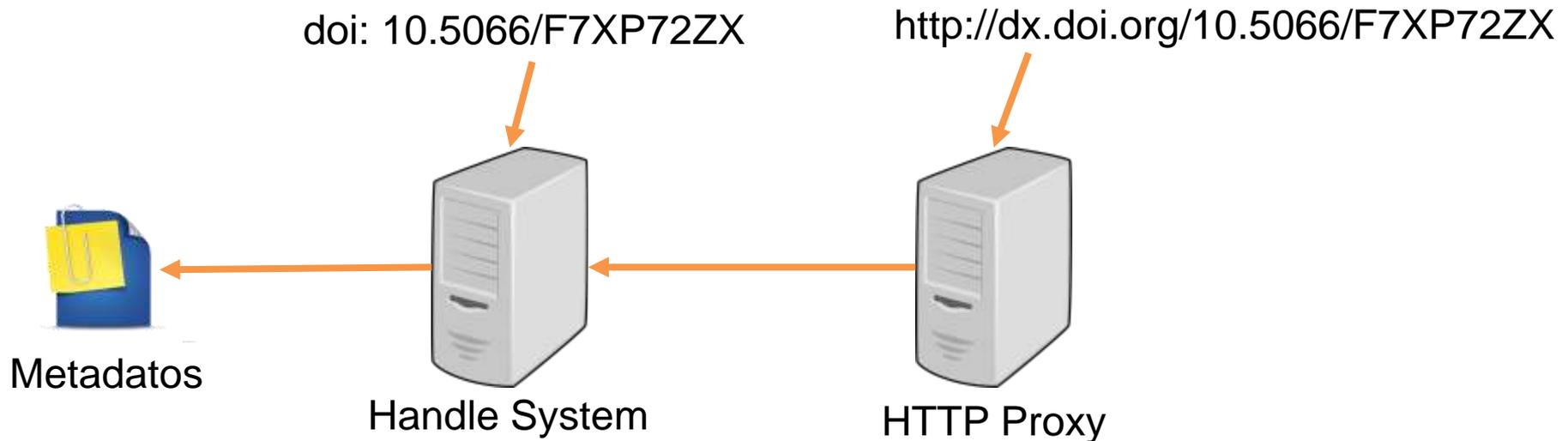
- Identificador en un Sistema de Información
- Propiedades básicas
 - Consistencia referencial → Semántica
 - Conformidad con un esquema → Sintáctica
 - Persistente → Útil
- Ejemplos
 - DNI → Personas
 - ISBN → Publicaciones

¿Qué es un PID?

- Propiedades extra (según necesidad)
 - Externo → Sistemas de Sistemas
 - Accionable → Tecnología de resolución
 - Globalmente único → Creación distribuida
 - Persistente a largo plazo → Archivo
 - Autorizado → Verificable
 - Trazable → Confiable

¿Qué es un PID?

- ¿Y es útil para una IDE?
 - USGS : Las colecciones de datos tienen DOI



QUIZ: ¿Qué PID tenemos ya en un IDE que son accionables?
 ENTONCES: ¿Y por qué USGS usa DOI?

¿Qué es un PID?

- URI
 - http (+protocolo) → servidorId/localId
 - PURL y similares → `http://purl.org/[id]`
 - Proxy con urn → LSID → `http://lsid.tdwg.org/[lsid]`
 - Proxy con handle → DOI → `http://dx.doi.org/[doi]`
 - urn (puede existir un *resolver*)
 - urn:lsid → biodiversidad
 - urn:uuid → representación uuid
 - otros (urn:isbn, urn:lex:eu)
 - otros (mailto:, tag:, ...)
- Handle (+protocolo) → autoridadId/localId
 - Doi → objetos digitales
- UUID
- Otros

Los PID solo funcionan si una comunidad da sentido y asegura su consistencia

PID en INSPIRE

- Identificar de forma única elementos relevantes
- ¿Cuáles son los de mayor impacto?
- Tipo
 - Objetos espaciales
 - Series de conjuntos de datos espaciales
 - Conjuntos de datos espaciales
 - Espacios de nombres (*namespaces*)
- Tema
 - Anexo I y II → Reglamento 1089/2010

PID en INSPIRE

- Propiedades básicas
 - Consistencia referencial → Por definición
 - Conformidad con un esquema → Modelo de datos
 - Persistente → Misma vida que el objeto

PID en INSPIRE

- Propiedades extra
 - Externo → Espíritu de la norma
 - Accionable → Arquitectura SOA INSPIRE
 - Globalmente único → Depende de buenas prácticas
 - Persistente a largo plazo → No
 - Autorizado → Depende de buenas prácticas
 - Trazable → Limitado

HTTP URI como PID en INSPIRE

- Se impulsa desde INSPIRE
 - Externo → Sí
 - Accionable → Web API (REST)
 - Globalmente único → Requiere que todos los PID sean accionables
 - Persistente a largo plazo → Riesgo de no persistencia a corto plazo
 - Autorizado → ¿Qué pasa si el objeto espacial se publica en una URI distinta?
 - Trazable → Depende de buenas prácticas

HTTP URI como PID en INSPIRE

- Ejemplos de problemas con HTTP URI
 - **Publicación descontrolada de datos**
 - No autorizado, multiples identificadores equivalentes
 - **Asignar HTTP URI sin publicar en Web**
 - No accionable, no globalmente único y no hay consistencia referencial
 - **Cambio de autoridad**
 - No accionable, cambio de identificador o inconsistencia al publicar
 - **Cambio de dominio**
 - No accionable, cambio de identificador o inconsistencia al publicar

HTTP URI como PID en INSPIRE

- Necesitamos un coordinador
- Funciones
 - Registro de *namespaces* a creadores
 - Funcionalidad mínima necesaria
 - Regla 1 namespace = 1 tipo de 1 creador
 - Pueden haber *namespaces* equivalentes
 - Servicio de resolución de URI
 - HTTP URI → localización actual del objeto
 - Asegura que la URI se accionable siempre (via Web API o Servicios INSPIRE)
 - Servicio de verificación de URI
 - HTTP URI → namespace ¿es válido? ¿equivalentes?
 - Resuelve problemas de cambio de propietario

Esquema Nacional de Interoperabilidad

- Criterios y recomendaciones AA.PP sobre interoperabilidad
 - Organizativos
 - Semánticos
 - Técnicos
 - Temporales
- Se desarrolla mediante NTIs
 - ¿Condicionan la decisión de usar HTTP URI?

Impacto en HTTP URI de INSPIRE

- Gobernanza

- No hay una NTI que defina el marco para las políticas de PID
 - Hace difícil crear un marco común en España
 - Algunas NTI definen PID propios (documento electrónico, expediente electrónico)
- Podemos basarnos en NTI que definen políticas

- Financiación

- Experiencia otras NTI con PID pueden dar orientación

Impacto en HTTP URI de INSPIRE

- Arquitectura
 - Condicionados por la NTI de Reutilización de recursos de información
 - Define un patrón de HTTP URI
 - ¿Se puede ignorar la NTI de RISP?
 - ENI no regula gestión de los nombres de dominios público
 - No hay un registro, no hay una política unificada
 - Hace que los PID HTTP sean frágiles

Caso de estudio

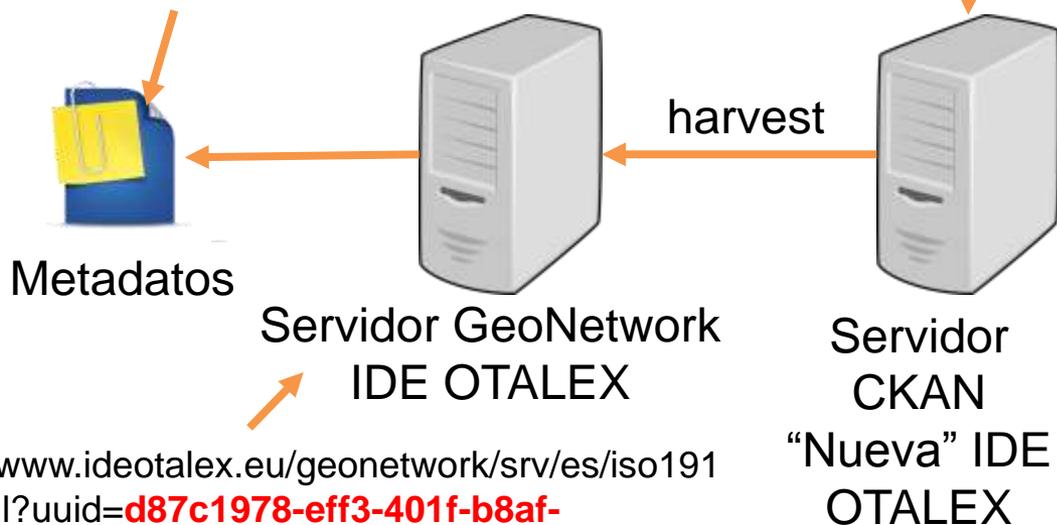
No reinventar la rueda
PID en la AA.PP para documento electrónico es <DIR3, año, ID local>

- IDE OTALEX → RISP (UPM – UNIZAR)

<http://www.ideotalex.eu/catalogo/A11016339/d87c1978-eff3-401f-b8af-9d87ecfa8430>

DIR3 – id del suceso

Creador: JUNTA DE EXTREMADURA. Agencia Extremeña de la Vivienda, el Urbanismo y el Territorio
Desapareció en 2007!!!



datos.gob.es

harvest

harvest

Metadatos

Servidor GeoNetwork
IDE OTALEX

Servidor
CKAN
"Nueva" IDE
OTALEX

Convenio: Podemos construir cualquier URI con cualquier formato pero siempre que contenga ese DIR3 y esa UUID sabremos a qué nos referimos

<http://www.ideotalex.eu/geonetwork/srv/es/iso19139.xml?uuid=d87c1978-eff3-401f-b8af-9d87ecfa8430>

Conclusión: Lo mínimo para tener éxito

- Si tenemos ...
 - PID del creador (posiblemente lo mejor usar DIR3)
 - ID local del recurso
 - UUID (aunque yo no los usaría sin contexto!!!)
 - contexto / ID local único
 - Sabemos cuando cambia su propietario (e.g. establecimiento de estructuras, cese, cambio de competencias) → Registro de espacios nombres
- ... la adaptación a cualquier esquema HTTP URI de PID sería sostenible a largo plazo
 - INSPIRE, RISP, ...

Gracias por su atención

Francisco J Lopez-Pellicer

fjlopez@unizar.es

IAAA, Universidad de Zaragoza