



DH@Madrid Summer School 2016

Cómo crear un perfil de aplicación de datos para el análisis poético

Tecnologías digitales aplicadas al estudio de la poesía

Mariana Curado Malta

mariana@iscap.ipp.pt



European
Research
Council

Madrid, 27 junio al 1 de julio de 2016



Mariana Curado Malta

- Profesora e investigadora en el Politécnico de Oporto (Portugal **P.PORTO**)
- Doctorada en Web Semántica en la Universidad de Minho (Portugal)
- Thesis: “Contributo metodológico para o desenvolvimento de perfis de aplicação de metadatos no contexto da Web Semantica”



CENTRO **ALGORITMI**



European
Research
Council

Plan del *workshop*

- Definiciones □ MAP, schema de metadados
- Contexto
- Como desarrollar un MAP □ método (pasos)
- Manos a la obra ...



Perfil de Aplicación de metadatos (MAP)

Conjunto de **elementos** de schemas de metadatos, de **restricciones** (reglas sobre los datos) y de **guías** de aplicación para un contexto específico de aplicación



Schemas de metadatos

- Schemas de metadatos o vocabularios RDF
- Un schema de metadatos es un conjunto de términos para describir cosas
- Los términos pueden ser:
 - Clases
 - Propriedades (object or data properties)



Perfil de Aplicación de metadatos (MAP)

- Conjunto de elementos de schemas de metadatos, de restricciones (reglas sobre los datos) y de guías de aplicación para un contexto específico de aplicación
- Elementos de schemas □ de diferentes schemas de metadatos - es un “mix & match” [3]



¿Objetivo del MAP?

Interoperabilidad

Comunidades comunicando
automáticamente = “sin permiso”



Sin muros

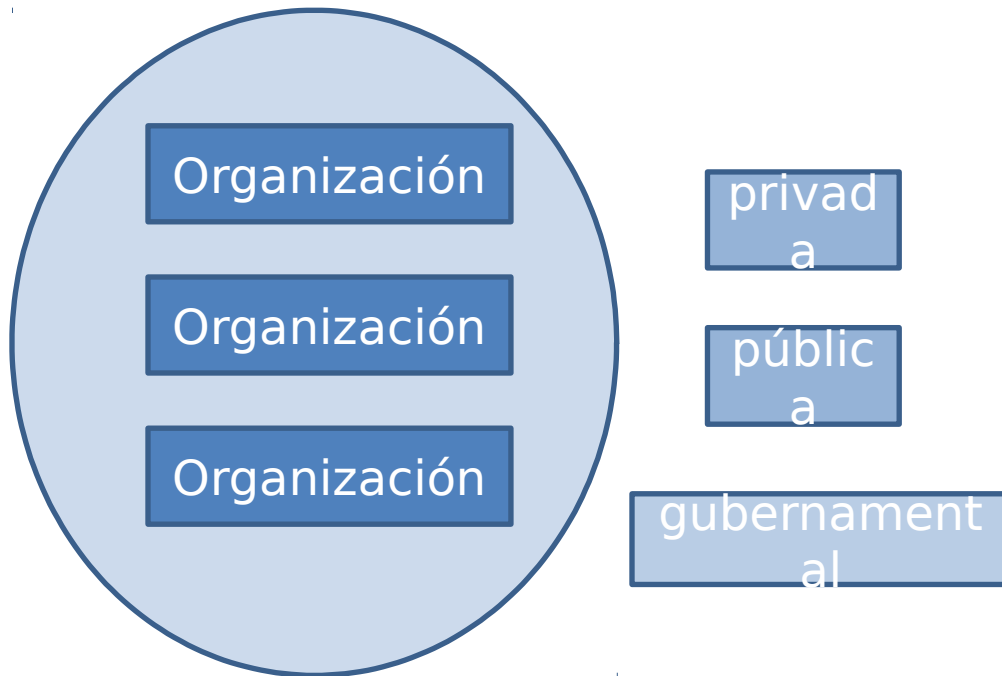
Dinámica

Utilizaciones no esperadas de los
datos

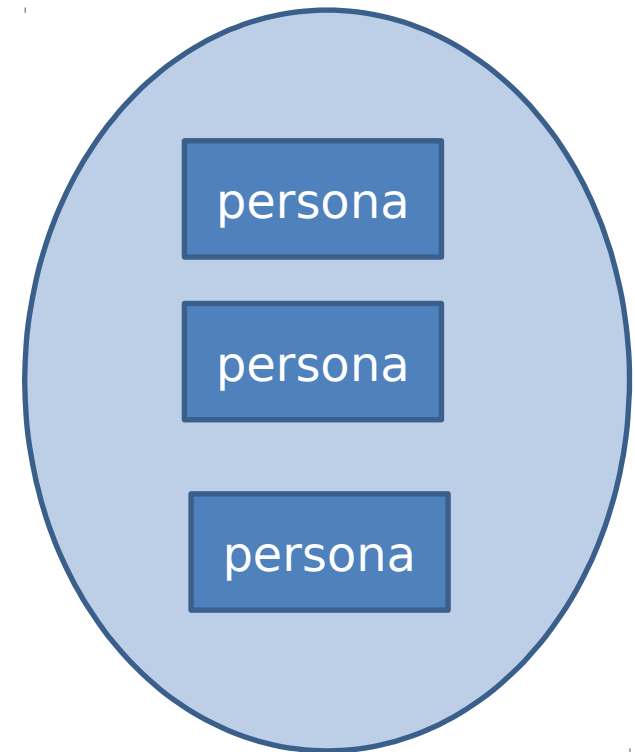


¿Comunidades?

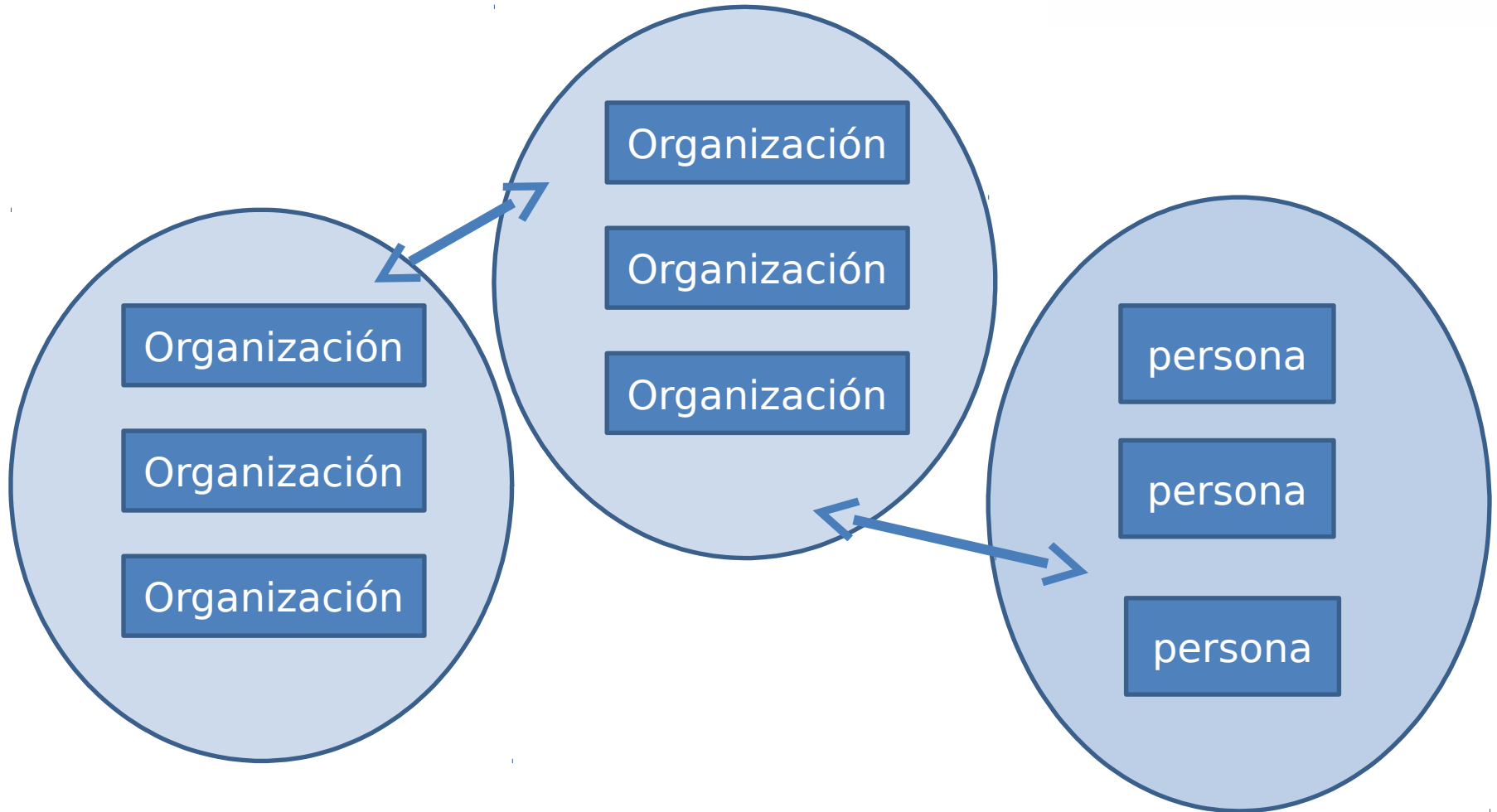
Comunidade de organizações



Grupos (organizados)



¿Comunidades?



MAP: ¿cómo hacerlo?

- Un método es importante
- Ayuda a mantener el foco
- A no pasar por alto los pasos
- A mejorar los resultados
- Y a reducir al mínimo los problemas de mantenimiento



MAP: pasos

1. Requisitos Funcionales

2. Modelo de Dominio

3. Description Set Profile

□ Objetivo de las próximas 2 horas:
aprender a hacer!



Manos a la obra...



Desarrollar un MAP

Paso 1

Definir los Requisitos Funcionales



Paso 1

Requisitos

Funcionales

- Las funcionalidades del sistema (*¿qué es el sistema?*)
- Exemplo, listar los poemas por:
 - Incipit
 - Idioma
 - Rima
 - *¿qué más puede ser interesante?*



Paso 1

Requisitos

Funcionales (RF)

¿Cómo obtener los requisitos?

1. Utilizar recursos existentes (Bases de datos, websites, documentación)
2. Hacer entrevistas, observar personas trabajando ...
3. Hacer casos de uso (*Use Cases*) [4]

□ Sacar los RF de los resultados



Paso 1 - RF

FRBR

- Functional requirements for Bibliographic Records (FRBR) [5]
- Un modelo conceptual para representar recursos bibliográficos



Paso 1 - RF

otros recursos importantes

- TEI – Text Encoding Initiative [6]
- Datos enlazados en la Biblioteca Nacional de España [7] □ ¿cuál es el modelo de datos?



Paso 2 - MD

Exercício Prático

- Conjunto de recursos en el fichero Recursos.xls
- ¿Qué deseamos hacer com esos recursos? ¿Compartir algo? ¿Hay databases que son iguales? ¿Dónde? ¿Quién es nuestra comunidad?



Desarrollar un MAP

Paso 2

Definir el Modelo de Dominio



Paso 2

Modelo de Dominio (MD)

- También llamado de Modelo de Datos (Data Model)
- Modelar la realidad que deseamos capturar
- Para servir los Requisitos Funcionales



Paso 2

Modelo de Dominio (MD)

- Modelar es:
 - identificar las cosas
 - Identificar las propiedades de las cosas
 - Identificar las relaciones entre las cosas
 - Y aúnidentificar las restricciones sobre las propiedades



Paso 2 - MD

- **Clase.** *Un constructo que representa cosas en el mundo real o/y mundo informacional*, ej.: “obra literaria”, “poema”, “manuscrito” o un concepto como “bibliografía”.
- **Relación.** *Una relación entre dos clases; como la relación entre el libro y un poema que contiene* (i.e. el obra literaria contiene poema), o la relación entre una obra literaria y una bibliografía (i.e. obra literaria tiene bibliografía).
(En RDF las relaciones se representan como *object properties*)
- **Propriedad.** *Una característica de una clase* en una dimensión particular, ej.: “nombre”, “fecha de nacimiento”, “número de estrofas”, “tipo de rima”.



Paso 2 - MD

Técnicas para representar MD

- Permiten representar información de una forma padronizada
- Permiten colocar las classes, relaciones y propiedades en diagramas padronizados (con una notación gráfica bien definida)
- Más sobre modelación en [8]



Paso 2 - MD

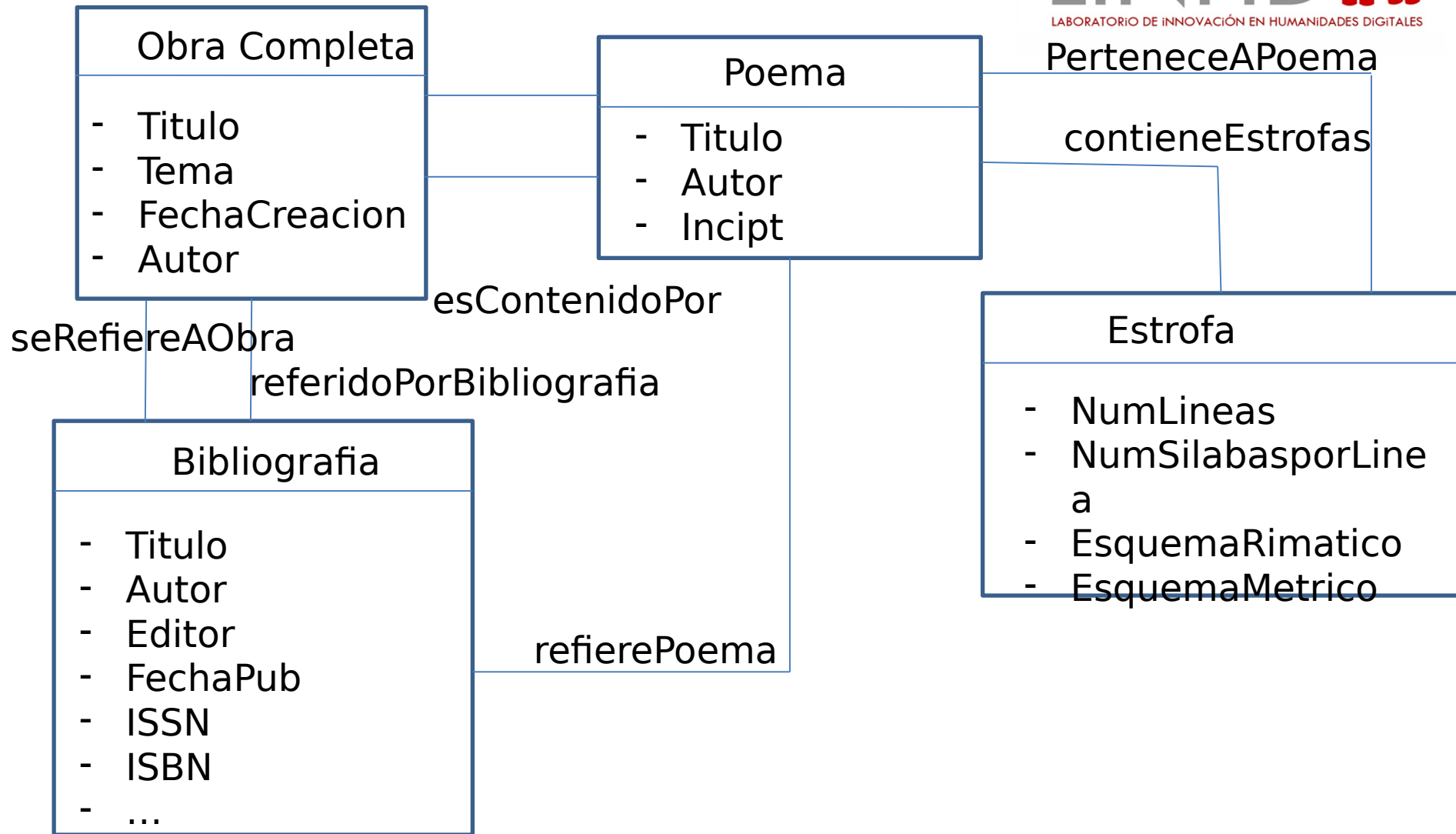
Técnicas para representar MD

- Ejemplos de técnicas más comunes:
 - UML – Unified Modeling Language [9]
 - ER – Entity Relationship model [10] (o Modelo relacional)
 - ORM – Object Role Modeling [11]
 - Y aún Grafo RDF (porque no?)
 - O cualquier outro tipo de técnica que el equipo de trabajo comprenda



Paso 2 - MD

Ejemplo contienePoemas



Desarrollar un MAP

Paso 3

Definir el Description Set Profile



Paso 3

Description Set

Profile (DSP)

- Exploración del entorno
- Alineamiento de vocabulários
- Matriz de restricciones
- Teste de la matriz de restricciones
- Codificación DSP



Paso 3 - DSP

Exploración del entorno

- ¿ Que es?
- Conocer el entorno □ conocer los schemas/vocabularios que sirven el dominio
- Para aplicar en el alineación de vocabularios



Paso 3a

Alineación de vocabs

| MD Classe | MD Propriedade | Schema de metadatos | Schema de metadatos Nombre de la propriedad |
|----------------|-------------------|---------------------|---|
| Obra Literaria | Título | dcterms | title |
| | ID | ??? | |
| | URI | Bibo | URI |
| | Incipit | dcterms | ??? |
| | Poema | dcterms | hasPart |
| Poema | Título | dcterms | title |
| | ID | | isPartOf |
| | Incipit | dcterms | ??? |
| | Verse | RemeTeca | ??? |
| | Estrofa | RemeTeca | |
| | Folio | | |



Paso 3a - Alineación de vocabs - Buenas Prácticas

- Utilizar lo que existe (mayor interoperabilidad con otras comunidades)
- Utilizar schemas standard
- Empezar por los schemas mas generalistas y por los mas utilizados
- Cuando hay términos que son parecidos pero no exactamente lo que queremos, crear *sub clases* o *sub propiedades*



Paso 3a - Alineación de vocabs - Buenas Practicas

- Si no existe ningunos terminos en ningun schema que pueda describir lo que necesitamos □ crear!
- Crear en un ambiente muy estable que sea persistente (ex: purl.org, W3C)
- Diseminar el vocabulario □ registrar en un lugar relevante (ex: LOV [12])



Paso 3a - Alineación de vocabs - utilizar lo que existe

- Vocabularios con uso general: DCMI [13], RDFS [14]
- Para nomear cosas: FOAF [15], Vcard [16]
- Para direcciones: Vcard
- Para citas bibliográficas: Biba [17]
- Para describir poesía: ReMeTca [18]



Paso 3a - Alineación de vocabs - ventajas de reutilizar

- Mejora la interoperabilidad
 - Ej.: dcterms:created es una fecha. Puedo usar el standard `2016-06-30^xsd:date`. Si uso un formato cualquier ex: “30 Junio 2016” es mas trabajoso el tratamiento automatico de las maquinas en la comunidad de metadados
- Credibilidad □ porque muestra profesionalismo
- Mas fácil y mas barato □ porque no es necesario crear nada!



Paso 3a - Alineación de vocabs - Vocabularios controlados

- Nombre técnico: Vocabulary Encoding Scheme (VES)
- Un conjunto de terminos, como una lista
- Controla errores, mejora interoperabilidad
- Ex general: DCMI type Vocabulary [19]
- Ex poesia:
<http://vocabularios.caicyt.gov.ar/pmc>



Paso 3a - Alineación de vocabs - Control de sintaxis

- Nombre técnico: Syntax Encoding Scheme (SES)
- Un conjunto de caracteres y de reglas que describen como el termino tiene de ser escrito
- Controla errores, mejora interoperabilidad
- Ex *XML Schema Built-in Datatypes* [20]: definir como se escribe la fecha *xsd:date* define “YYYY-MM-DD”



Paso 3b

Matriz de restricciones

- Abrir el fichero MR&Test.xls



Paso 3c

Test de la Matriz de Rest.

- Probar la matriz con recursos existentes
- Hacer el test con recursos de orígenes diferentes
- ¿Es posible describir los recursos con lo que hemos definido?
- ¿No? ¿Por qué? □ reiterar....



Paso 3c

Test de la Matriz de Rest.

| Test in Lab | |
|------------------------|-------------------------------|
| Resource A | |
| Label | Value |
| Instance of Class | Class X |
| Property A1 | |
| Property A2 | |
| as many rows as needed | |
| Resource B | |
| Label | Value |
| Instance of Class | Class Y |
| Property B1 | |
| Property B2 | |
| as many rows as needed | |
| Resource n | (as many resources as needed) |
| Label | Value |
| Instance of Class | Class Z |
| Property n1 | |
| Property n2 | |
| as many rows as needed | |



Paso 4

DSP

- Codificar todo lo que esta en la matriz de restricciones
- Utilizar el lenguaje de marcación DSP [21]



Manos a la obra...

Ejercicio Practico!



más imágenes divertidas en risasinmas.com



Referencias

- [1] <http://dublincore.org/documents/profile-guidelines/>
- [2] <http://dublincore.org/documents/singapore-framework/>
- [3] <http://www.ariadne.ac.uk/issue25/app-profiles>
- [4] Schneider, G., & Winters, J. P. (2001). *Applying use cases: a practical guide* (Second Edi). Boston,: Addison-Wesley.
- [5] Carlyle, A. (2006). *Understanding FRBR as a conceptual model: FRBR and the bibliographic universe*. *Library resources & technical services*, 50(4), 264.
- [6] <http://www.tei-c.org/release/doc/tei-p5-doc/en/Guidelines.pdf>
- [7] <http://datos.bne.es>
- [8] Halpin, T., & Morgan, T. (2010). *Information modeling and relational databases*. Morgan Kaufmann.
- [9] <http://agilemodeling.com/artifacts/classDiagram.htm>
- [10] Chen, P. P. S. (1976). *The entity-relationship model—toward a unified view of data*. *ACM Transactions on Database Systems (TODS)*, 1(1), 9-36
- [11] <http://www.orm.net/>
- [12] Linked Open Vocabularies: <http://lov.okfn.org/dataset/lov/>



Referencias

- [13] <http://dublincore.org/documents/dcmi-terms/>
- [14] <https://www.w3.org/TR/rdf-schema/>
- [15] <http://xmlns.com/foaf/spec/>
- [16] <https://www.w3.org/TR/vcard-rdf/>
- [17] <http://lov.okfn.org/dataset/lov/vocabs/bibo>
- [18] <http://lov.okfn.org/dataset/lov/vocabs/remetca>
- [19] <http://purl.org/dc/dcmitype/>
- [20] <https://www.w3.org/TR/rdf11-concepts/#xsd-datatypes>
- [21] <http://dublincore.org/documents/dc-dsp/>

