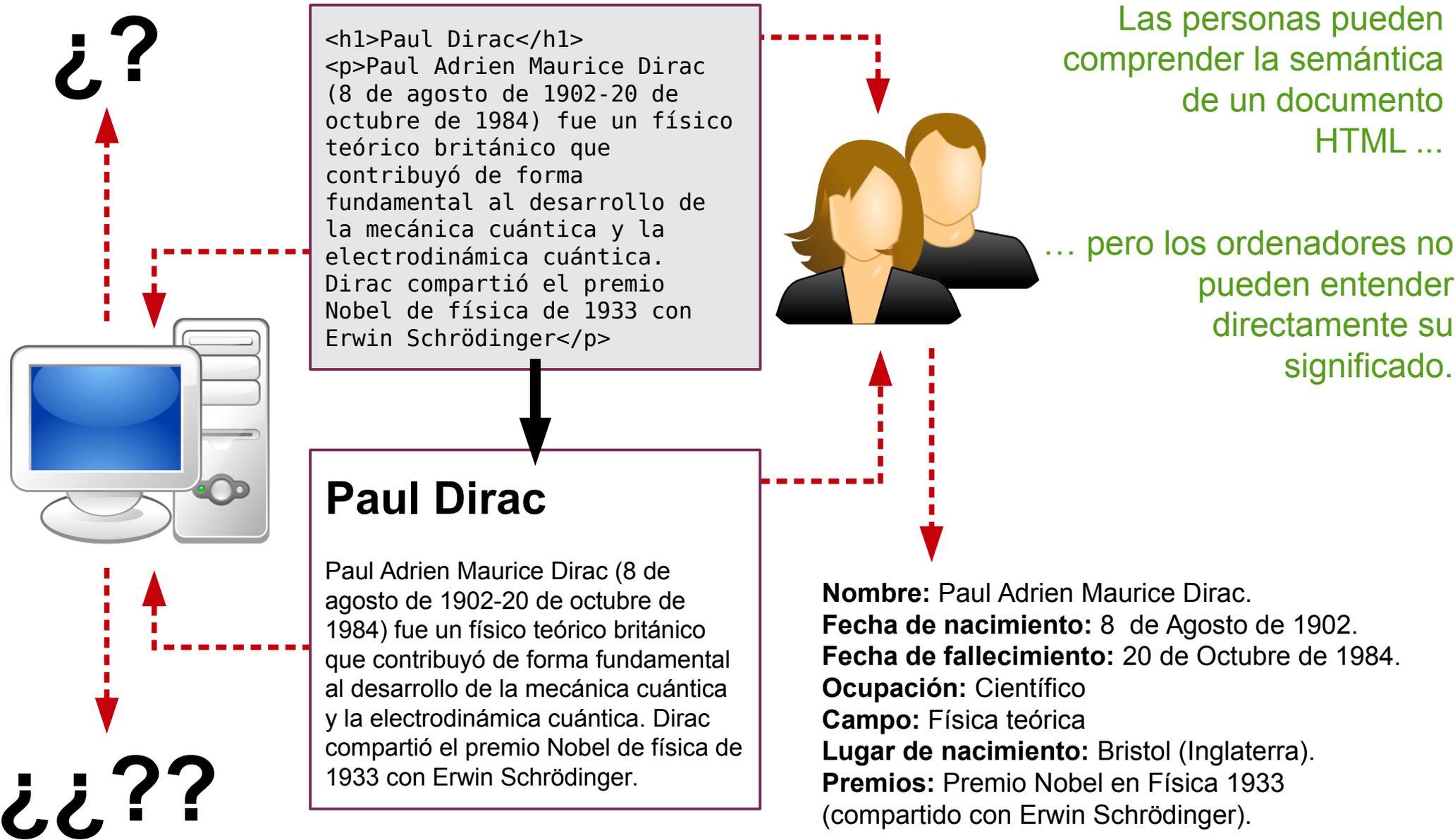
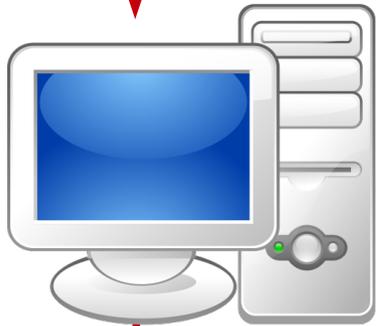


Semántica y máquinas



Semántica y máquinas

```
<person>
  <name>
    <firstname>Paul</firstname>
    <middlename>Adrien</middlename>
    <middlename>Maurice</middlename>
    <surname>Dirac</surname>
  </name>
  <birthDate>1902-08-08</birthDate>
  <deathDate>1984-10-20</deathDate>
  <birthPlace>
    <city>Bristol</city>
    <country>England</country>
  </birthPlace>
  <occupation>Scientist</ocupation>
  <field>Physics (theoretical)</field>
  <awards>
    <award>
      <awardName>Nobel price in Physics</awardName>
      <year>1933</year>
      <sharedWith>Erwin Schrödinger</sharedWith>
    </award>
  </awards>
</person>
```



Nombre: Paul Adrien Maurice Dirac.
Fecha de nacimiento: 8 de Agosto de 1902.
Fecha de fallecimiento: 20 de Octubre de 1984.
Ocupación: Científico
Campo: Física teórica
Lugar de nacimiento: Bristol (Inglaterra).
Premios: Premio Nobel en Física 1933 (compartido con Erwin Schrödinger).

Las programas necesitan recibir la información en un **formato** que siga un **modelo de datos**.

De este modo las aplicaciones informáticas pueden extraer **datos estructurados** (datos conformes a un esquema fijo) y ejecutar procesos que operen utilizando la semántica de los datos.

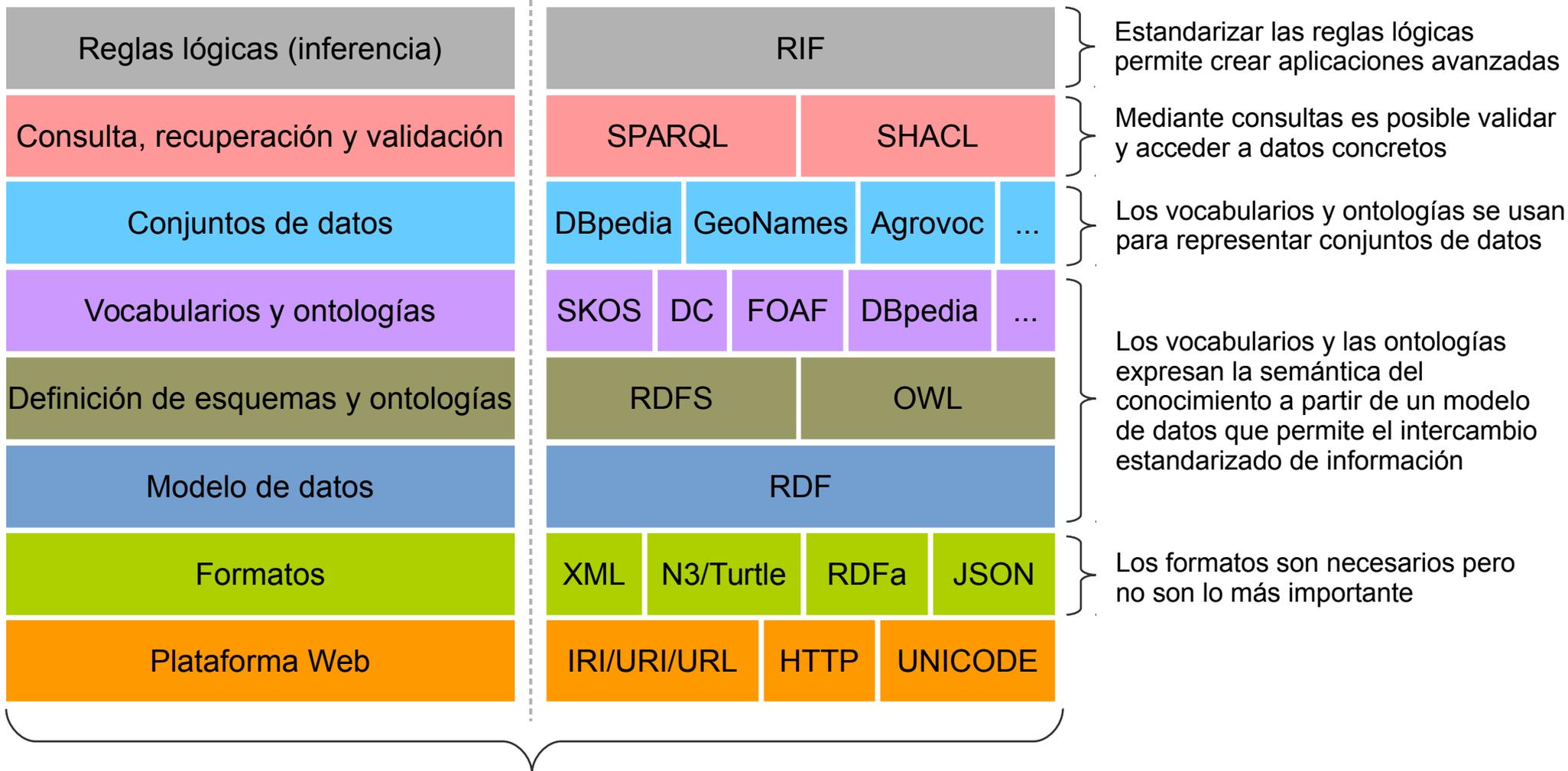
Visión general de la Web Semántica

- La **Web Semántica no es diferente de la Web convencional**. Se trata de un enfoque concreto para publicar, acceder y utilizar datos estructurados.
- La Web Semántica se basa en **descripciones de recursos** en vez de páginas.
- Un recurso es **cualquier objeto que pueda ser descrito**: libros, personas, conceptos, lugares, imágenes, videos, etc).
- Es posible definir **diferentes tipos de enlaces** entre los recursos para describir la naturaleza de la relación entre los mismos.
- Los **vocabularios y las ontologías** son colecciones de “términos” para describir los atributos y las relaciones entre recursos.
- Los vocabularios y las ontologías son públicos y sus **terminos tienen una semántica específica**: p.e. el término “title” de Dublin Core (dcterms:title) se refiere al título de un recurso.
- Deben utilizarse **estándares abiertos** para definir vocabularios y ontologías.
- Las **referencias URI** permiten identificar tanto los recursos como los términos de los vocabularios y ontologías.

Arquitectura de la Web Semántica

Conceptos y Abstracciones

Especificaciones y soluciones



La Web Semántica está basada en la Web Convencional

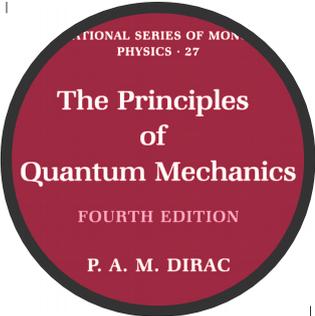
Modelo de datos RDF



- RDF es un modelo de datos que estructura la información sobre recursos en sentencias con la forma de **sujeto-predicado-objeto** (triples).
- El objeto de una sentencia puede ser otro recurso (**propiedad de objeto**) o datos alfanuméricos (**propiedad de datos**).
- Las sentencias RDF pueden agruparse en **grafos** interconectados con nodos (resources) y arcos (propiedades). Ambos, nodos y arcos, se identifican mediante **URIs**. El grafo en su totalidad también tiene una URI asociada. Un **conjunto de datos** puede contener uno o más grafos.
- **RDF Schema** y **OWL** complementan a RDF y permite crear taxonomías de clases y propiedades para definir vocabularios y ontologías.

Ejemplo de grafo RDF

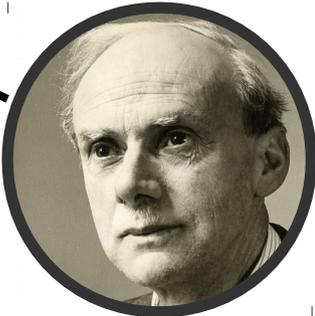
bne:bimo0000786818



dcterms:title

“Principios de Mecánica Cuántica”@es

dcterms:creator

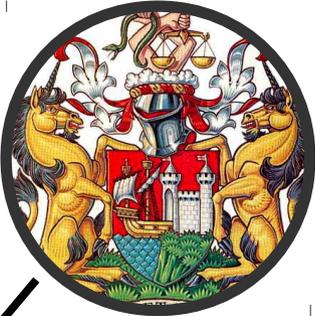


schema:name

“Paul Dirac”@en

viaf:17350683

dbo:birthPlace



schema:name

“Bristol”@en

geonames:2654675

dcterms: <http://purl.org/dc/terms/>
dbo: <http://dbpedia.org/ontology/>
viaf: <http://viaf.org/viaf/>

schema: <http://schema.org/>
geonames: <http://sws.geonames.org/>
bne: <http://datos.bne.es/resource/>

Serialización RDF

Las sentencias de los grafos RDF se codifican (serialización) utilizando algún formato legible por máquina para que los sistemas informáticos puedan leer y procesar los datos automáticamente.

RDF/XML	<pre><rdf:Description rdf:about="http://www.wikidata.org/entity/Q83186"> <dcterms:creator rdf:resource="https://viaf.org/viaf/96994048"/> <dcterms:title>Romeo and Juliet</dcterms:title> </rdf:Description></pre>	Permite utilizar las tecnologías XML (XSLT, XPath, XQuery, etc.)
Turtle	<pre><http://www.wikidata.org/entity/Q83186> dcterms:creator <https://viaf.org/viaf/96994048"> ; dcterms:title "Romeo and Juliet"@en .</pre>	Compacto y ligero, perfecto para las máquinas.
JSON-LD	<pre>@id:"http://www.wikidata.org/entity/Q83186", "dcterms:creator": "https://viaf.org/viaf/96994048", "dcterms:title": {"@language": "en", "@value": "Romeo and Juliet"}</pre>	Diseñado para su consumo por Javascript
RDFa	<pre><div resource="http://www.wikidata.org/entity/Q83186"> <h1 property="dcterms:title">Romeo and Juliet</h1> <p>Author: William Shakespeare</p> </div></pre>	Permite insertar sentencias RDF en el código de documentos HTML

SPARQL

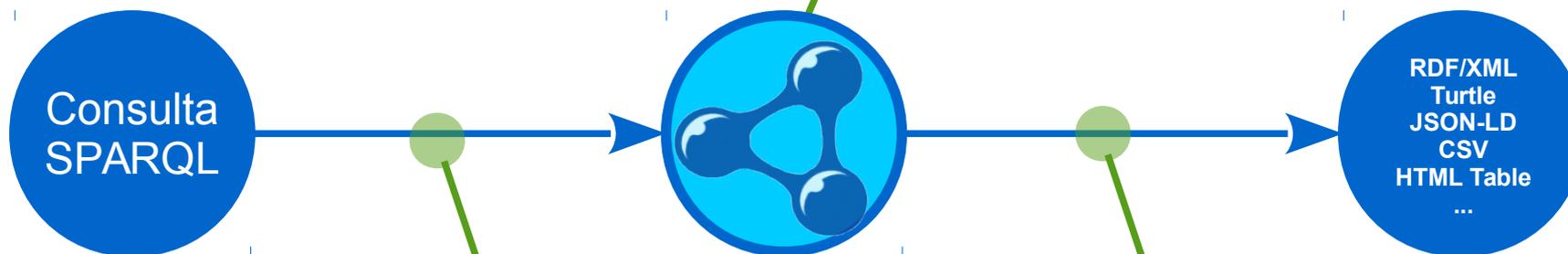
Las sentencias RDF de un conjunto de datos pueden almacenarse en un fichero de texto plano para su descarga masiva. SPARQL permite recuperar sentencias específicas que cumplan ciertas condiciones. **SPARQL es para RDF lo que SQL es para el Modelo Relacional.**

```
select ?lugarNacimiento
WHERE {
  ?persona schema:name "Paul Dirac" .
  ?persona dbo:birthPlace ?lugarNacimiento .
}
```



¿Donde nació Paul Dirac?

Un SPARQL Endpoint es un servicio web RESTful para realizar consultas SPARQL



Las consultas SPARQL recuperan sentencias almacenadas en una base de datos RDF (“Triplestore”).

Los resultados se obtienen en un formato específico.

Linked Open Data

- Se utilizan **URIs dereferenciables** para identificar y acceder a las descripciones de recursos. Para ello una URI debe poder ser utilizada como una URL que además de identificar el recurso permite acceder al mismo.
- El protocolo HTTP debe utilizarse junto con técnicas de **negociación de contenido** para obtener los datos en el formato solicitado por el cliente web, como por ejemplo un navegador web (formato HTML), código Javascript (JSON-LD) o un cliente RDF (RDF/XML, N3/Turtle, etc).
- Puesto que los recursos se identifican con URIs dereferenciables, es posible definir **enlaces a recursos externos** desde un conjunto de datos. Esto permite descubrir nuevos enlaces siguiendo dichos enlaces.
- Los conjuntos de datos deben estar disponibles para su **descarga** en diversos formatos RDF y a través de un SPARQL Endpoint para la recuperación selectiva.
- Los conjuntos de datos deben tener una **licencia de uso** para su reutilización abierta.

W3C Data Activity: Construyendo la Web de Datos

- Hay una **demanda creciente respecto a la disponibilidad de conjuntos de datos** (“big data” o “small data”), de diversas temáticas, fuentes y formatos.
- Muchos conjuntos de datos existentes se publican sin coordinación o colaboración entre editores de datos para aplicar **vocabularios y técnicas de modelado comunes**.
- El W3C es consciente de esta realidad y desarrolla iniciativas para superar ciertos problemas relativos a la edición y publicación de datos y para facilitar **integración y procesamiento de datos** a nivel de toda la Web.
- Esto es posible proporcionando **formatos estándar para el intercambio de datos, modelos, herramientas y orientación**, incluso con tecnologías diferentes de RDF.
- **Las personas y las organizaciones** deberían ser capaces de **compartir datos** en la medida de lo posible, usando herramientas existentes y prácticas de trabajo de forma que permita a otros obtener y agregar valor a los datos y adaptarlos a sus necesidades.
- El W3C ayudará al desarrollo de un ecosistema de datos mayor, asegurando la **interoperabilidad** y facilitando el **desarrollo de aplicaciones** teniendo en cuenta a las diferentes comunidades (usuarios, desarrolladores, editores, etc)

Tipos de grupos del W3C

- **Grupo de Trabajo (Working Group - WG):**
Edita determinados documentos tales como recomendaciones, informes de seguimiento técnico, casos de uso, requisitos de conformidad, etc...
- **Grupo de Interés (Interest Groups - IG):**
Reune a personas interesadas en evaluar posibles tecnologías y políticas web. Se trata de un foro para el intercambio de ideas.
- **Grupo de Coordinación (Coordination Groups - CG):**
Gestiona las interrelaciones y facilita la comunicación con otros grupos dentro o fuera del W3C.

Data Activity: Grupos del W3C

- **Permissions & Obligations Expression WG**

Desarrollo de recomendaciones para representar los permisos de acceso y obligaciones de los usuarios para el uso de contenidos digitales en general y conjuntos de datos en particular.

- **Spatial Data on the Web WG**

Formalización de estándares y buenas prácticas sobre información geo-espacial.

- **RDF Data Shapes WG**

Creación de un lenguaje que complementa a SPARQL para la definición de restricciones estructurales de los grafos RDF.

- **Data on the Web Best Practices WG**

Definición de un conjunto de buenas prácticas para la aplicación de múltiples tecnologías y desarrollo de vocabularios necesarios para el desarrollo de un ecosistema de datos en la web.

- **Semantic Web IG**

Foro para asistir a los desarrolladores y usuarios de tecnologías de la Web Semántica como RDF, SPARQL, SKOS, OWL y otros vocabularios y aplicaciones.

- **Semantic Web Health Care and Life Sciences IG**

Desarrollo y apoyo al uso de las tecnologías de la Web Semántica en el ámbito de la sanidad, ciencias de la vida, investigación clínica y la medicina traslacional.

Web = Tecnologías + Metodologías

Tecnologías

Metodologías

HTML CSS

PHP Javascript

Java WordPress

Apache

Experiencia de usuario **Pautas de Accesibilidad**

Arquitectura de la Información **Usabilidad**

Diseño responsivo

RDF Fuseki

Turtle DBpedia

Jena XML OWL

Calidad y uso de los datos

Linked Buenas Open Data Prácticas

Diseño de ontologías

Data on the Web Best Practices WG: Actividades



Datos abiertos y datos cerrados

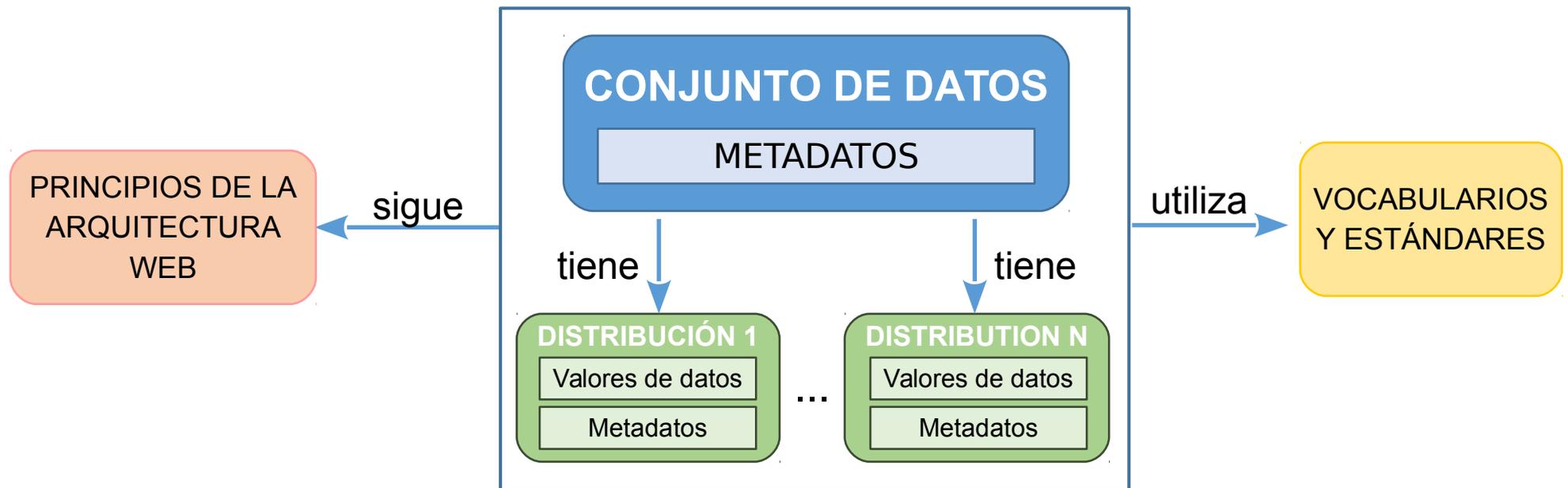
La *Open Knowledge Foundation* define **Datos abiertos** como “datos que pueden ser libremente usados, modificados y compartidos por cualquiera para cualquier propósito”. (<http://opendefinition.org>)

... pero las **tecnologías web** son **aplicables** a datos que no son abiertos, o en escenarios que combinan **datos abiertos y cerrados**.

Escenarios de datos cerrados

- **Datos tras un firewall**: Se trata de datos cerrados creados dentro de una organización, bloqueados por un firewall u otro sistema de control de acceso; pueden enlazar con datos abiertos alojados en cualquier parte y representados mediante estándares web.
- **Política de los editores**: Los datos pueden ser cerrados debido a determinadas políticas seguidas por los editores y suministradores de los datos (p.e. datos sensibles para la competencia entre empresas) o datos cuyo acceso está limitado legalmente por parte de las administraciones públicas (p.e. datos sobre seguridad nacional, datos sanitarios, etc)
- **Ciclo de vida de los datos**: Hay periodos entre la generación de los datos y su publicación como datos abiertos en los que no están disponibles por motivos de análisis, validación o ventaja estratégica (p.e. desarrollo industrial, investigación científica, etc). En estos casos se consideran como datos cerrados.
- **Datos no disponibles mediante HTTP**: Los datos publicados **únicamente** mediante protocolos diferentes del protocolo HTTP (p.e. FTP) tienen una interoperabilidad muy limitada y desde el punto de vista de los datos abiertos no se encuentran en la web y deben considerarse como datos cerrados.

Publicación de Datos en la Web



- **Conjunto de datos:** Colección de datos, publicados o recopilados por un agente y disponible para su acceso y descarga en uno o más formatos.
- **Distribución:** Los conjuntos de datos se publican en diferentes distribuciones para facilitar su uso y consumo. Ejemplos: Un fichero CSV, un SPARQL Endpoint, una API, una fuente RSS.
- **Metadatos:** Información sobre la estructura, descripción, acceso, calidad, procedencia, licencia y uso de conjuntos de datos y distribuciones. Contribuyen a la confianza y reutilización de los datos.
- **Identificación:** Cualquier recurso (todo un conjunto de datos o un elemento específico) debería publicarse con una URI estable para que puedan ser referenciados y enlazados.
- **Vocabularios y estándares:** El uso de vocabularios de datos y estándares promueve la interoperabilidad entre conjuntos de datos.

Datos en la Web: Retos (1/2)

- **Metadatos**: Los datos no podrán ser encontrados o reutilizados por alguien diferente del editor si no se han proporcionado suficientes metadatos.
- **Licencia**: En función del tipo de licencia adoptado por el editor, existirán ciertas restricciones para compartir y reutilizar los datos.
- **Procedencia**: Sin conocer la procedencia, los consumidores no tienen forma de confiar en el origen, integridad y credibilidad de los datos.
- **Calidad**: La calidad de un conjunto de datos puede tener un gran impacto en la calidad de las aplicaciones que hacen uso de los mismos.
- **Versiones**: Un conjunto de datos publicados en la web pueden cambiar a lo largo del tiempo. Por este motivo deben crearse nuevas versiones del conjunto de datos.
- **Identificadores**: El uso de URIs para descubrir, usar y citar recursos en la web depende del uso de URIs que los identifique y puedan utilizarse para dereferenciarlos en Internet.
- **Formatos**: El mejor y más flexible mecanismo de acceso no tiene sentido a menos que se suministren los datos en formatos que permitan su uso y reutilización.

Datos en la Web: Retos (2/2)

- **Vocabularios**: Los vocabularios definen los conceptos y propiedades para describir y representar un área de interés. Agrupan los términos que pueden aplicarse para clasificar y describir recursos, caracterizar relaciones y definir restricciones de uso.
- **Acceso**: Es preciso proporcionar formas de acceso sencillo y estándares a los datos en la web para que tanto las personas como las máquinas puedan compartirlos.
- **Preservación**: Por diversos motivos los editores de los datos precisan borrarlos. Eliminar sin más un conjunto de datos es una mala práctica.
- **Retroalimentación**: Los editores quieren que los datos que publican se ajusten a las necesidades de los usuarios. La retroalimentación por parte de estos ayuda a mejorar los datos publicados y fomenta la publicación de nuevos datos.
- **Enriquecimiento**: Los datos pueden mejorarse o perfeccionarse en su origen o una vez procesados utilizando diversos recursos de información.
- **Reedición**: La reutilización de los datos es otro modo de publicación. Es posible combinar conjuntos de datos, crear aplicaciones web para visualizar o utilizar los datos para generar nuevos conjuntos de datos, traducirlos, transformarlos, etc.

Beneficiós de las buenas prácticas



Reuse:

Mayores posibilidades de reutilizar un conjunto de datos por diferentes grupos de usuarios.



Trust:

Mayor confianza en los conjuntos de datos por parte de los usuarios y consumidores.



Comprehension:

Mejor comprensión de los metadatos, estructura y significado de los datos y naturaleza del conjunto de datos.



Access :

Las personas y las máquinas pueden acceder a datos actualizados de múltiples formas.



Linkability:

Es posible definir enlaces entre recursos de datos (conjuntos de datos o datos individuales).



Interoperability:

Mayor facilidad para alcanzar acuerdos entre desarrolladores, editores y consumidores.



Discoverability:

Las máquinas podrán descubrir automáticamente conjuntos de datos o datos concretos dentro de estos.



Processability:

Las máquinas podrán procesar y manipular automáticamente el contenido de los conjuntos de datos.

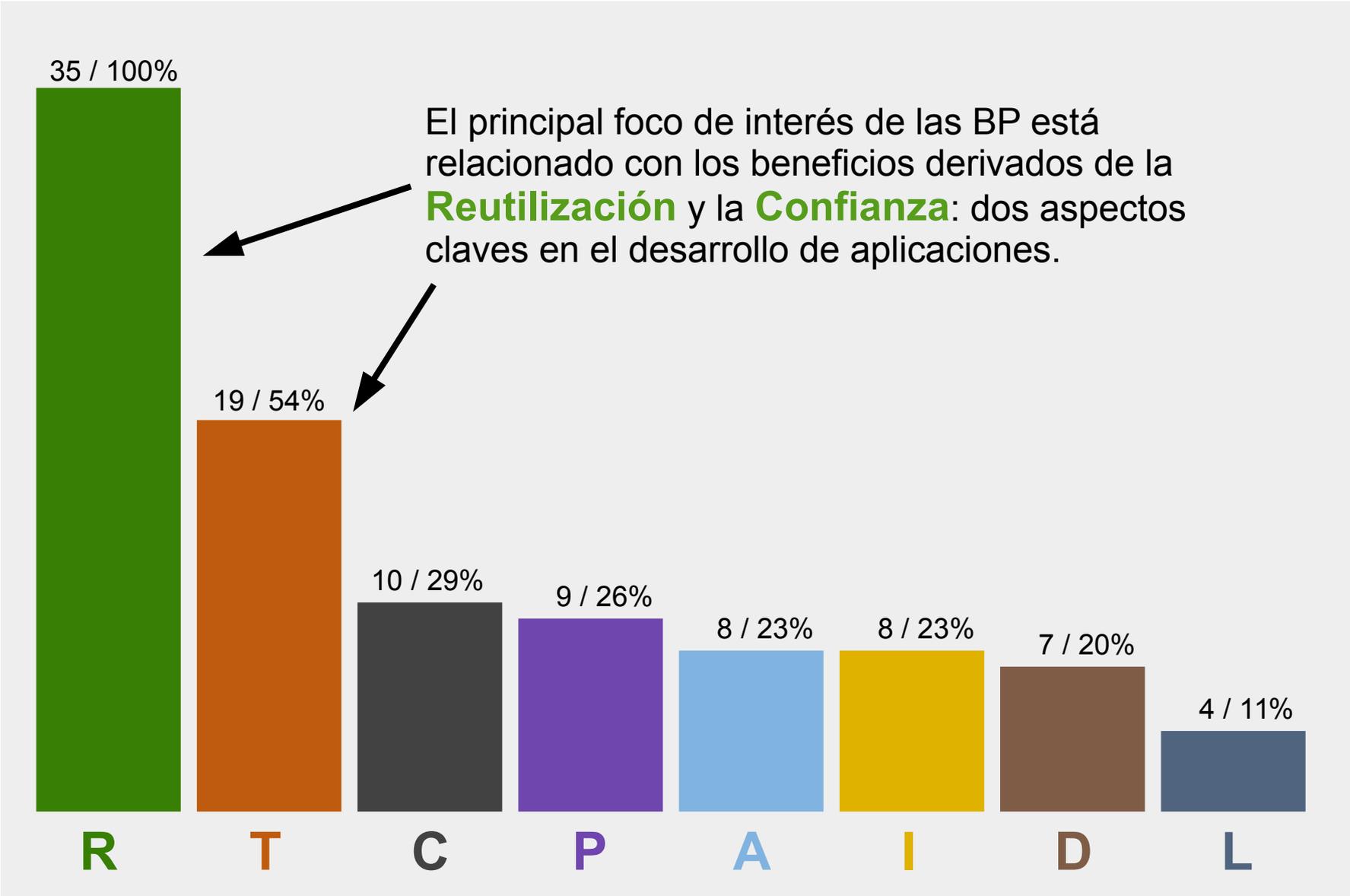
Retos y Buenas prácticas (1/2)

Metadatos Proporcionar metadatos para personas y máquinas	BP-1: Proporcionar metadatos	R C D P
	BP-2: Proporcionar metadatos descriptivos	R C D
	BP-3: Proporcionar metadatos estructurales	R C P
Licencia Permitir/Restringir el acceso/uso	BP-4: Proporcionar información sobre la licencia de uso de los datos	R T
Procedencia Aumentar confianza del usuario	BP-5: Proporcionar información sobre la procedencia de los datos	R C T
Calidad Añadir indicadores de calidad	BP-6: Proporcionar información sobre la calidad de los datos	R T
Versiones Seguimiento de las versiones y del historial de versiones	BP-7: Proporcionar un indicador de versión	R T
	BP-8: Proporcionar información sobre el historial de la versión	R T
Identificación Identificar conjuntos de datos y distribuciones	BP-9: Usar URIs persistentes como identificadores de los conjuntos de datos	R L D I
	BP-10: Usar URIs persistentes como identificadores dentro del conjunto de datos	R L D I
	BP-11: Asignar URIs a las versiones y series de los conjuntos de datos	R D T
Formatos Escoger los formatos de datos a utilizar	BP-12: Usar formatos estándar legibles por máquina	R P
	BP-13: Usar representaciones de datos localmente neutras	R C
	BP-14: Proporcionar datos en múltiples formatos	R P
Vocabularios Mejorar la interoperabilidad de los datos	BP-15: Reutilizar vocabularios, preferiblemente estandarizados	R C T I P
	BP-16: Escoger un nivel de formalización correcto	R C I

Retos y Buenas prácticas (2/2)

Acceso Proporcionar acceso a los datos	BP-17: Proporcionar el conjunto de datos en un único fichero para su descarga	R A
	BP-18: Proporcionar subconjuntos para grandes conjuntos de datos	R L A P
	BP-19: Negociación de contenido para suministrar datos en múltiples formatos	R A
	BP-20: Proporcionar acceso en tiempo real a los datos	R A
	BP-21: Suministrar los datos actualizados	R A
	BP-22: Proporcionar una explicación sobre los datos que no estén disponibles	R T
	BP-23: Hacer que los datos estén disponibles a través de una API	R A I P
	BP-24: Usar estándares Web como base de desarrollo de las APIs	R L D A I P
	BP-25: Proporcionar documentación completa para las APIs	R T
Preservación Archivar los datos	BP-26: Evitar actualizaciones de las APIs incompatibles con versiones anteriores	R T I
	BP-27: Preservar los identificadores	R T
Retroalimentación Involucrar a los usuarios	BP-28: Evaluar la cobertura de los datos	R T
	BP-29: Recopilar las retroalimentaciones de los consumidores de datos	R C T
Enriquecimiento Añadir valor a los datos	BP-30: Hacer que las retroalimentaciones estén disponibles públicamente	R T
	BP-31: Enriquecer los datos mediante la generación de nuevos datos	R C T P
Reedición Reutilizar los datos responsablemente	BP-32: Proporcionar presentaciones complementarias	R C T A
	BP-33: Permitir la retroalimentación con el editor original de los datos	R T I
	BP-34: Cumplir los términos de la licencia de uso	R T
	BP-35: Citar la edición original de los datos	R D T

Nota sobre los beneficios de las Buenas prácticas



Estructura de una Buena Práctica (1/2)

Best Practice 4: Provide data license information

Provide a link to or copy of the license agreement that controls use of the data.

Why

The presence of license information is essential for data consumers to assess the usability of data. User agents may use the presence/absence of license information as a trigger for inclusion or exclusion of data presented to a potential consumer.

Intended Outcome

Humans will be able to understand data license information describing possible restrictions placed on the use of a given distribution and software agents to automatically detect the data license of a distribution.

Possible Approach to Implementation

Data license information can be available via a link to, or embedded copy of, a human-readable license agreement. It can also be made available for processing via a link to, or embedded copy of, a machine-readable license agreement.

One of the following vocabularies that include properties for linking to a license can be used:

- Dublin Core [DCTERMS] (`dct:license`)
- Creative Commons [CCREL] (`cc:license`)
- schema.org [SCHEMA-ORG] (`schema:license`)
- XHTML [XHTML-VOCAB] (`xhtml:license`)

There are also a number of machine-readable rights languages, including:

- The Creative Commons Rights Expression Language [CCREL]
- The Open Digital Rights Language [ODRL-model]
- The Open Data Rights Statement Vocabulary [ODRS]

Denominación de la BP con una breve explicación de su contenido.

Justificación de la presencia de la BP y aspectos relacionados con su aplicación.

Resultados previstos

Explicación acerca de los resultados esperados tras aplicar la BP.

Posible implementación

Información detallada con soluciones para implementar la BP en una amplia variedad de situaciones. Esta sección incluye referencias a técnicas específicas y vocabularios concretos para aplicar la BP con tecnologías y recomendaciones existentes.

Estructura de una Buena Práctica (2/2)

EXAMPLE 4

Machine-readable

The CSV distribution of the bus stops dataset (`stops-2015-05-05.csv`) will be published under the [Creative Commons Attribution-ShareAlike 3.0 Unported](#) license. The property `dct:license` is used to include this information as part of the distribution metadata.

```
:stops-2015-05-05.csv
  a dcat:Distribution ;
  dct:title "CSV distribution of stops-2015-05-05 dataset" ;
  dct:description "CSV distribution of the bus stops dataset of MyCity" ;
  dcat:mediaType "text/csv;charset=UTF-8" ;
  dct:license <http://creativecommons.org/licenses/by-sa/3.0/> ;
.
```

Human-readable

[Example page](#) with human-readable data license information of the distribution.

How to Test

Check if the metadata for the dataset itself includes the data license information in a human-readable format.

Check if a user agent can automatically detect /discover the data license of the dataset.

Evidence

Relevant use cases: [R-LicenseAvailable](#), [R-MetadataMachineRead](#), [R-LicenseLiability](#)

Benefits



Reuse



Trust

Ejemplos

Las buenas prácticas incluyen ejemplos detallados para su implementación. Esta sección muestra desarrollos prácticos con versiones legibles tanto por personas como por máquinas. La versión legible por máquinas incluye un código de ejemplo. La versión legible por personas incluye un enlace a una página web que implementa la BP.

Como testar la BP

Técnicas o pasos para verificar o validar la aplicación de la BP en su conjunto de datos.

Evidencia

Cada BP está relacionada con uno o varios requisitos del documento de casos de uso.

Beneficios

Muestra las mejoras y beneficios relacionados con la disponibilidad del conjunto de datos en la Web.

Estructura de una Buena Práctica (2/2)

EXAMPLE 4

Machine-readable

The CSV distribution of the bus stops dataset (`stops-2015-05-05.csv`) will be published under the [Creative Commons Attribution-ShareAlike 3.0 Unported](#) license. The property `dct:license` is used to include this information as part of the distribution metadata.

```
:stops-2015-05-05.csv
  a dcat:Distribution ;
  dct:title "CSV distribution of stops-2015-05-05 dataset" ;
  dct:description "CSV distribution of the bus stops dataset of MyCity" ;
  dcat:mediaType "text/csv;charset=UTF-8" ;
  dct:license <http://creativecommons.org/licenses/by-sa/3.0/> ;
.
```

Human-readable

[Example page](#) with human-readable data license information of the distribution.

How to Test

Check if the metadata for the dataset itself includes the data license information in a human-readable format.

Check if a user agent can automatically detect /discover the data license of the dataset.

Evidence

Relevant use cases: [R-LicenseAvailable](#), [R-MetadataMachineRead](#), [R-LicenseLiability](#)

Benefits



Reuse



Trust

Ejemplos

Las buenas prácticas incluyen ejemplos detallados para su implementación. Esta sección muestra desarrollos prácticos con versiones legibles tanto por personas como por máquinas. La versión legible por máquinas incluye un código de ejemplo. La versión legible por personas incluye un enlace a una página web que implementa la BP.

Como testar la BP

Técnicas o pasos para verificar o validar la aplicación de la BP en su conjunto de datos.

Evidencia

Cada BP está relacionada con uno o varios requisitos del documento de casos de uso.

Beneficios

Muestra las mejoras y beneficios relacionados con la disponibilidad del conjunto de datos en la Web.

Vocabularios

Vocabulario diseñado para facilitar la interoperabilidad entre catálogos de datos publicados en la Web.

DCAT
(Data Catalog Vocabulary)

Extensión de DCAT para citar, proporcionar metadatos descriptivos para el consumidor sobre las distribuciones y uso de conjuntos de datos

DQV
(Data Quality vocabulary)

Extensión de DCAT para expresar la calidad de los datos, frecuencia de actualización, compromisos de mantenimiento, etc.

DUV
(Data usage vocabulary)

Vocabularios: estructura y finalidad

- **Data Catalog Vocabulary (DCAT):**

- DCAT permite describir **conjuntos de datos o distribuciones** en catálogos de datos.
- Los editores incrementan las posibilidades de que sus conjuntos de datos puedan ser encontrados.
- DCAT permite a las aplicaciones consumir metadatos de múltiples catálogos.
- Las clases de DCAT permiten representar: el catálogo en sí mismo, descripciones de conjuntos de datos y distribuciones y registros de los conjuntos de datos en el catálogo.

- **Data Quality Vocabulary (DQV):**

- Este vocabulario ayuda a los desarrolladores a **confiar** en los datos mediante determinadas propiedades para evaluar la **calidad** de un conjunto de datos o distribución.
- Para expresar dichas propiedades los conjuntos de datos o distribuciones se relacionan con cinco tipos de información: **retroalimentación y certificaciones de calidad, estándares cumplidos, políticas sobre la calidad de los datos, métricas cualitativas o cuantitativas y entidades involucradas en la procedencia del conjunto de datos o distribución.**
- DQV considera que las medidas de calidad pueden definirse y derivarse a partir de dos tipos: dimensiones (características de calidad) y métricas (procedimientos para medir dimensiones de calidad).

- **Data Usage Vocabulary (DUV):**

- El **submodelo de citación** establece como los usuarios pueden citar los datos, como los editores los anotan con referencias bibliográficas y como los editores pueden incluir referencias bibliográficas de los usuarios.
- El **submodelo de uso** permite realizar un seguimiento del uso de los datos y conocer como pueden ser utilizados por los consumidores a través de herramientas desarrolladas por terceros.
- El **submodelo de retroalimentación** permiten a los usuarios valorar los datos capturando el contexto de uso y comunicar sugerencias al editor.

Ejemplos de uso: DCAT

```
ejemplo:paradasBus-2016-01-01 a dcat:Dataset ;  
  dcterms:title "Paradas de Autobús" ;  
  dcterms:issued "2016-01-01"^^xsd:date ;  
  pav:version "1.0" ;  
  dcat:distribution :paradasBus-2016-01-01.csv .
```

Descripción del
Conjunto de
Datos

```
ejemplo:paradasBus-2016-01-01.csv a dcat:Distribution ;  
  dcat:downloadURL ejemplo:paradasBus-2015-05-05.csv> ;  
  dcterms:title "Distribución CSV del conjunto de datos  
    http://datos.ejemplo.org/paradasBus-2015-05-05"@es ;  
  dcterms:description "Distribución CSV de un conjunto de datos  
    con paradas de autobús"@es ;  
  dcat:mediaType "text/csv;charset=UTF-8" ;  
  dcterms:license <http://creativecommons.org/licenses/by-sa/3.0/> .
```

Descripción de
la distribución

PREFIJOS

```
ejemplo: <http://datos.ejemplo.org/dataset/> .  
dcterms: <http://purl.org/dc/terms/> .  
oa: <https://www.w3.org/ns/oa#> .  
pav: <http://purl.org/pav/> .  
dcat: <http://www.w3.org/ns/dcat#> .
```

Ejemplos de uso: DQV

ejemplo:paradasBus-2016-01-01 dqv:hasQualityAnnotation calidad:classificationQA .

```
calidad:disponibilidad
  a dqv:Dimension ;
  skos:prefLabel "Disponibilidad"@es ;
  skos:definition "Disponibilidad de un conjunto de datos presente,
    disponible y preparado para su uso."@en ;
```

Dimensión del
criterio de
anotación

```
calidad:classificationQA a dqv:UserQualityFeedback ;
  dqv:inDimension calidad:availability .
  oa:hasTarget ejemplo:paradasBus-2016-01-01 ;
  oa:motivatedBy dqv:qualityAssessment, oa:classifying ;
  oa:hasBody :cuatro_estrellas ;
```

Criterio de
anotación sobre
calidad

```
calidad:cuatro_estrellas a skos:Concept;
  skos:prefLabel "Cuatro Estrellas"@en ;
  skos:definition "Conjunto de datos disponible con URIs que pueden
    ser utilizadas para identificar objetos y para acceder a los
    mismos a través de la web en un formato legible por máquina."@es .
```

Elemento del
criterio de
anotación

PREFIJOS

```
ejemplo: <http://datos.ejemplo.org/dataset/> .
calidad: <http://datos.ejemplo.org/quality/> .
oa: <https://www.w3.org/ns/oa#> .
skos: http://www.w3.org/2004/02/skos/core#
pav: <http://purl.org/pav/> .
dqv: <http://www.w3.org/ns/dqv#> .
```

Ejemplos de uso: DUV

ejemplo:paradasBus-2016-01-01.csv duv:hasUsage uso:vis-rutas .

```
uso:vis-rutas a duv:Usage;  
  dcterms:title "Visualizador de rutas de autobús"^^xsd:string ;  
  dcterms:created "2016-05-05"^^xsd:date ;  
  duv:refersTo :stops-2015-05-05.csv .  
duv:hasUsageTool uso:calculador-rutas ;
```

Uso específico
del conjunto de
datos

```
uso:calculador-rutas a duv:UsageTool ;  
  dcterms:title "Route Calculator"^^xsd:string ;  
  dcterms:created "2016-03-04"^^xsd:date ;  
  dcterms:identifier <http://dx.doi.org/10.0902/1975.16> .
```

Herramienta
que da soporte al
uso específico

PREFIJOS

```
ejemplo: <http://datos.ejemplo.org/dataset/> .  
uso: <http://datos.ejemplo.org/usage/> .  
dcterms: <http://purl.org/dc/terms/> .  
duv: <http://www.w3.org/ns/duv#> .  
dqv: <http://www.w3.org/ns/dqv#> .
```

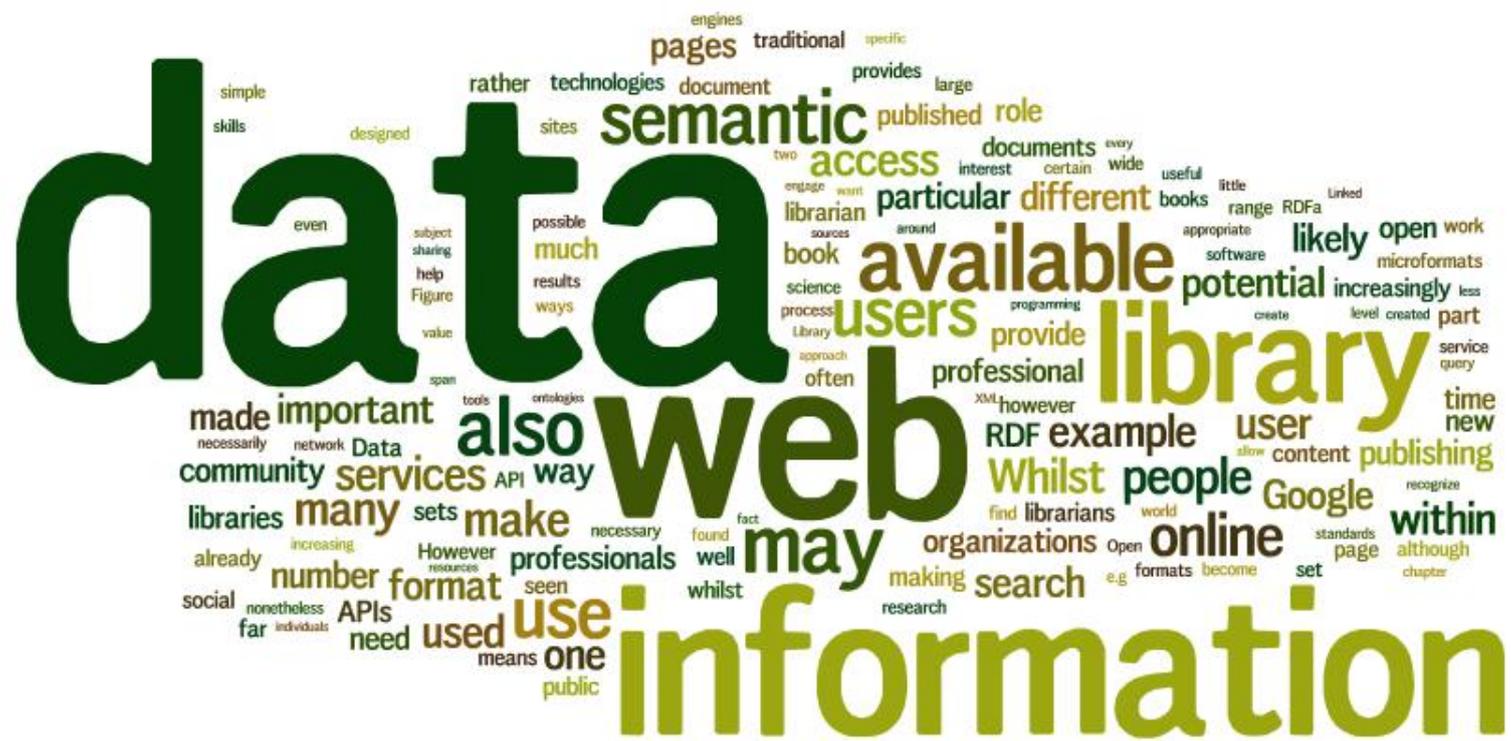
¿Qué es lo siguiente?

- **Aprobación** de la recomendación final sobre buenas prácticas.
- Revisión y actualización de **DCAT**.
- Elaboración y aprobación de las recomendaciones para **DQV y DUV**.
- **Despliegue** y aplicación efectivas de DCAT, DQV y DUV.
- **Metodologías** para la implementación de las buenas prácticas.
- **Herramientas** de análisis de conjuntos de datos y distribuciones.
- **Evaluación** de las buenas prácticas en conjuntos de datos y distribuciones ya publicadas.

En un futuro cercano la publicación y la descripción de datos en la Web según un conjunto de buenas prácticas es algo que se dará por hecho y la clave será el uso que se haga de los mismos mediante **APIs y aplicaciones incluso por terceros.**

Algunas referencias

- Heath, T. & Bizer, C. (2011). **Linked Data: Evolving the Web into a Global Data Space**. Synthesis Lectures on the Semantic Web: Theory and Technology, 1:1, 1-136. Morgan & Claypool.
<http://linkeddatabook.com/editions/1.0/>
- W3C (2013). **Data on the Web Best Practices Working Group**.
https://www.w3.org/2013/dwbp/wiki/Main_Page
- W3C (2014a). **RDF 1.1 Primer**. W3C Working Group Note 25 February 2014.
<https://www.w3.org/TR/2014/NOTE-rdf11-primer-20140225/>
- W3C (2014b). **Data Catalog Vocabulary (DCAT)**. W3C Recommendation 16 January 2014.
<https://www.w3.org/TR/vocab-dcat/>
- W3C (2016a). **Data on the Web Best Practices**. W3C Candidate Recommendation 30 August 2016.
<https://www.w3.org/TR/dwbp/>
- W3C (2016b). **Data on the Web Best Practices: Data Quality Vocabulary**. W3C Working Group Note 30 August 2016.
<https://www.w3.org/TR/vocab-dqv/>
- W3C (2016c). **Data on the Web Best Practices: Dataset Usage Vocabulary**. W3C Working Group Note 30 August 2016.
<https://www.w3.org/TR/vocab-duv/>



Muchas gracias
por su atención
¿Preguntas?

