



# Legislative data portals and linked data quality

Jose Emilio Labra Gayo  
WESO

**WE**b **S**emantics **O**viedo  
University of Oviedo, Spain



# About me...

In 2004, founded WESO (Web Semantics Oviedo) research group

Goal: Practical applications of semantic technologies

Several domains: e-Government, e-Health,...

2 books:

"*Web semántica*" (in Spanish), 2012

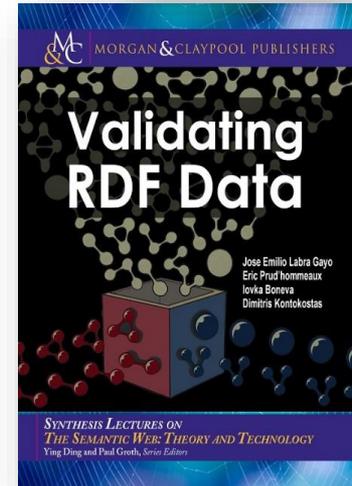
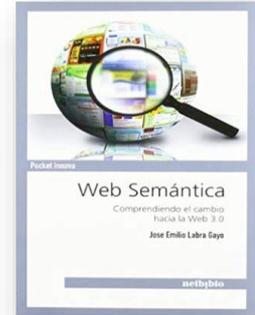
"*Validating RDF data*", 2017

...and software:

SHaclEX (Scala library, implements ShEx & SHACL)

RDFShape (RDF playground)

WikiShape (Wikidata playground)



HTML version: <http://book.validatingrdf.com>

Examples: <https://github.com/labra/validatingRDFBookExamples>



# About this talk...

It will be divided in 2 parts

## 1<sup>st</sup> part

Legislative linked data portals

Chilean National library of Congress

Presented at ESWC'19\*



## 2<sup>nd</sup> part

Linked data quality

RDF description and validation

ShEx & SHACL overview



# Legislative linked data portals

Processing the History of the Law at Chile

**Francisco Cifuentes Silva**

Library of Congress, Chile  
PhD Student  
WESO research group

**Jose Emilio Labra Gayo**

WESO research group  
University of Oviedo, Spain

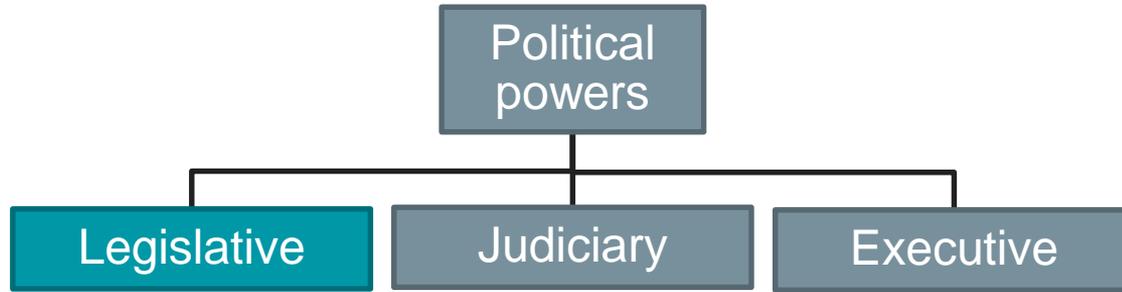
More information:

Cifuentes-Silva F., Labra Gayo J.E. (2019) Legislative Document Content Extraction Based on Semantic Web Technologies. In: Hitzler P. et al. (eds) The Semantic Web. ESWC 2019. Lecture Notes in Computer Science, vol 11503. Springer, Cham

# Chilean Library of Congress

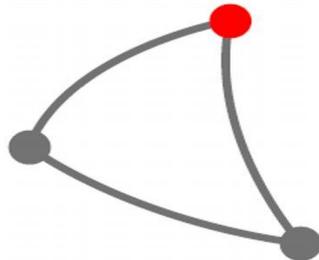


In Spanish: BCN (***B**iblioteca del **C**ongreso **N**acional de Chile*)



Independent body inside the Legislative power

Advices the parliament and gives services to citizens



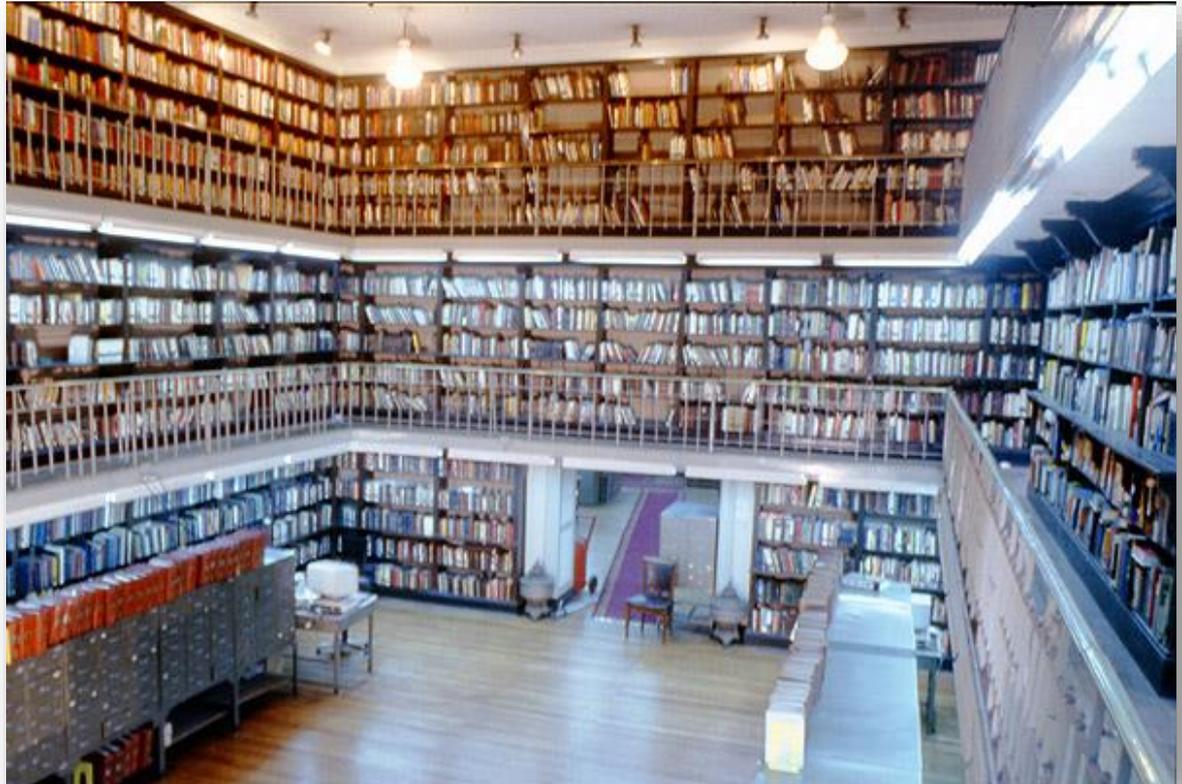
REPÚBLICA DE CHILE  
**SENADO**

<http://www.bcn.cl>

# 2 projects at library of congress (BCN)



History of the Law  
Parliamentary work



# History of the Law (LeyChile)



Collect all documents generated during a law legislative process

Phases:

An initiative sees life as a draft bill

Subject to debates

Validity time (it is published)

Modifications, additions,...

Derogation

Goal:

Capture *the spirit* of the law

Traceability

<https://www.bcn.cl/historiadelaley>

# Parliamentary work



Collect all legislative activity by each Member of Parliament

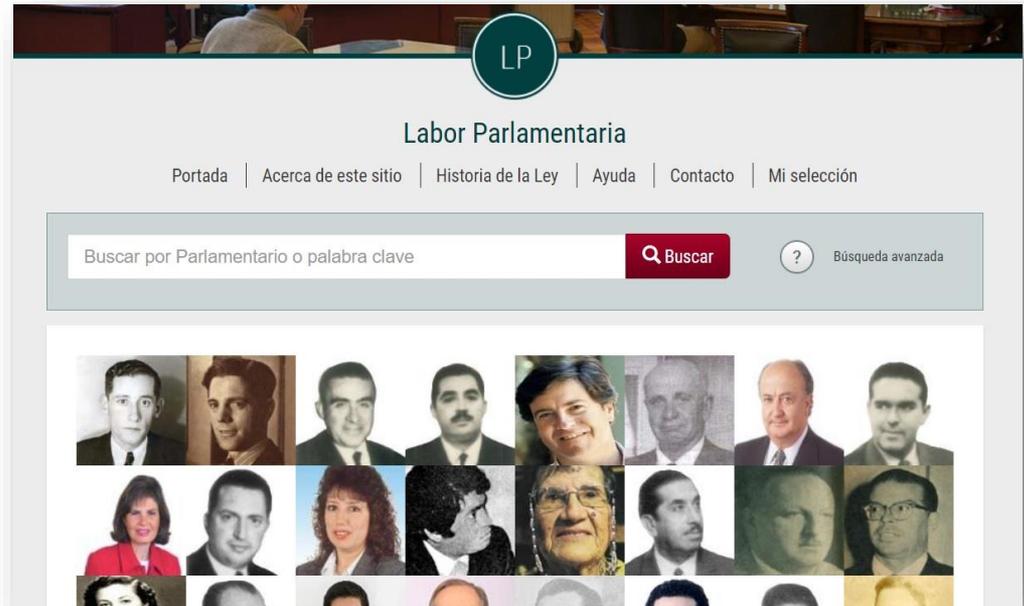
Retrieve all interventions made

Parliamentary motion

Session journal

Commission report

Ordered and categorised



<https://www.bcn.cl/laborparlamentaria/>

# Both projects adopted semantic technologies

Some initial reasons:

Semantic technologies considered one pillar of strategic plan (in 2014)

Innovative action to generate new products

Improve interoperability mechanisms

Sem. Web aligned well with open & public data



# Which semantic technologies?

An icon of a yellow folder containing several white documents, positioned to the left of the first text box.

## Text mining and content enrichment

Entity extraction

Topic identification

Automatic markup

Classification

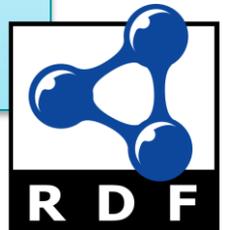
## Machine readable info

XML & URIs

RDF

Ontologies

Linked Open Data



# Workflow pipelines

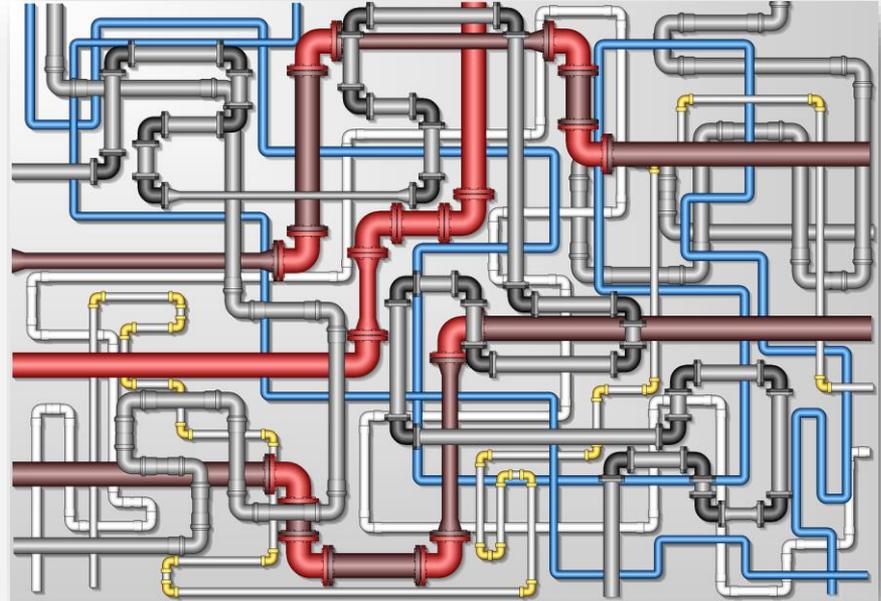


## 3 main steps

Automatic XML Marker

RDF & Linked data generation

Content delivery



# Workflow overview

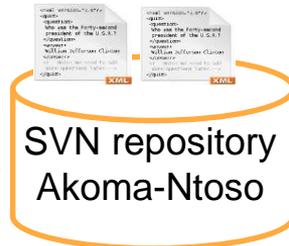
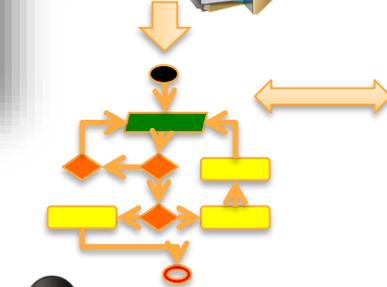
## Legislative documents

- Some docs in paper (requires OCR)
- Text documents

National library

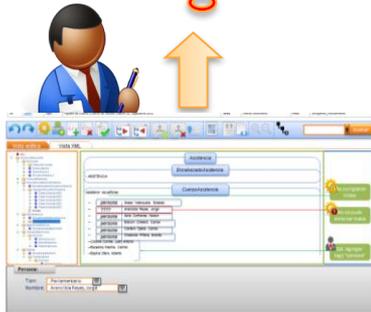


Automatic XML marker



Publishing  
(RDF extraction  
From Akoma-Ntoso)

XML editor & tools



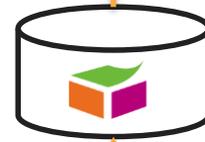
Content portals



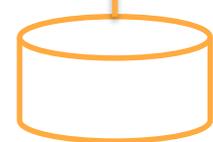
Services layer



Linked Open Data

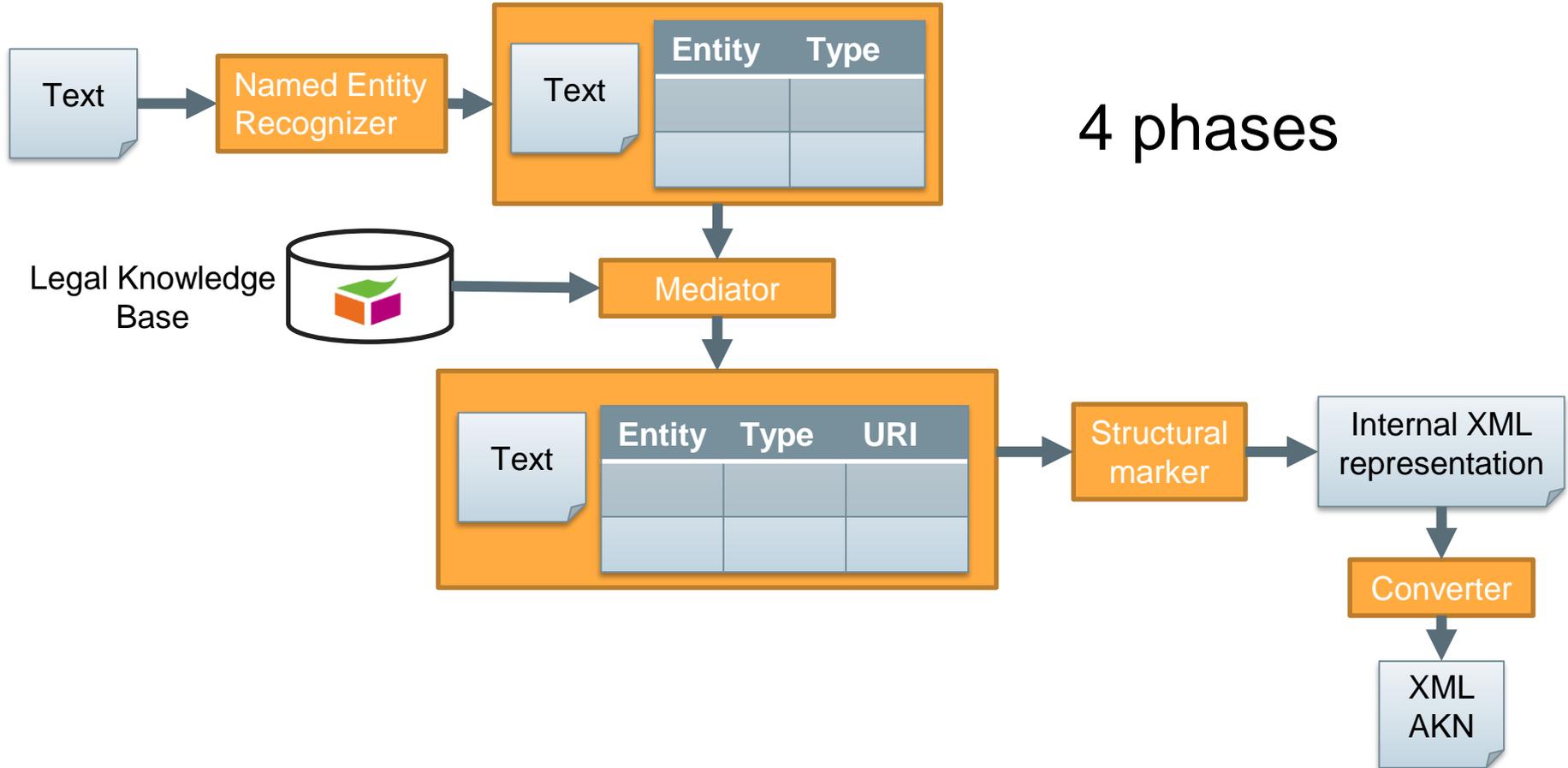


Query DB





# Automatic XML marker



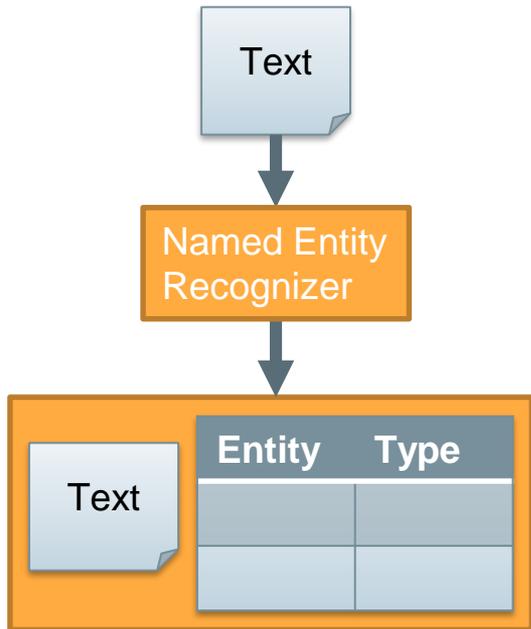


# 1. Named Entity Recognizer

## Detection of entities & types of entities

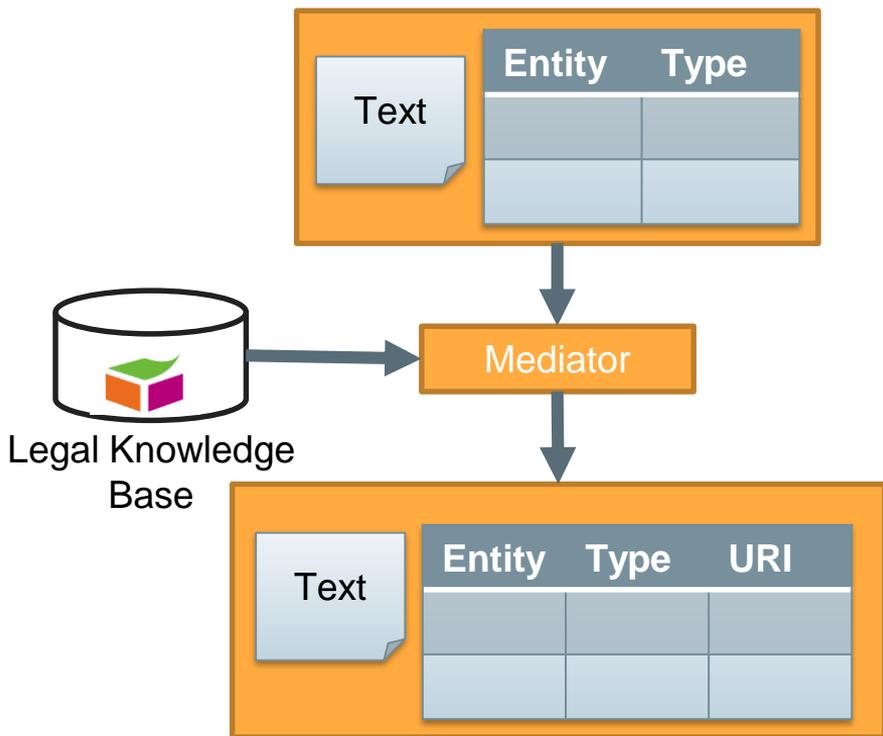
Web service implementing the Stanford NER with a CRF classifier

Evaluation in production: detects 97% entities



| Type         | Some examples  | # of entities |
|--------------|--|---------------|
| Person       | Salvador Allende, Sebastián Piñera   | 5.139         |
| Organization | Ministerio de Salud, SERNATUR  | 2.848         |
| Location     | Valparaíso, Santiago de Chile  | 1.251         |
| Document     | PROYECTOS DE LEY violencia que ha golpeado a las familias El   |               |
| Role         | <code>&lt;entity&gt;</code><br><code>&lt;body bestClass="ROL" knownFraction="1"&gt;Secretario&lt;/body&gt;</code><br><code>&lt;class value="ROL" probability="0.2201459321"/&gt;</code><br><code>&lt;/entity&gt;</code>                      |               |
| Events       | , señor  |               |
| Law          | <code>&lt;entity&gt;</code><br><code>&lt;body bestClass="PER" knownFraction="1"&gt;Raúl Guerrero Guerrero&lt;/body&gt;</code><br><code>&lt;class value="PER" probability="0.83023132"/&gt;</code><br><code>&lt;/entity&gt;</code>            |               |
| Dates        | y ei rrosecretario,  |               |
|              | <code>&lt;entity&gt;</code><br><code>&lt;body bestClass="PER" knownFraction="0.5"&gt;don Fernando Parga Santelices&lt;/body&gt;</code><br><code>&lt;class value="PER" probability="0.9555160392"/&gt;</code><br><code>&lt;/entity&gt;</code> |               |
|              | .  |               |

## 2. Mediator



Entity linking and disambiguation

Text similarity algorithms

Based on Apache Lucene

In-house development

- Use of context information to narrow list of candidates
- Custom filters and association heuristics
- Specialized web services

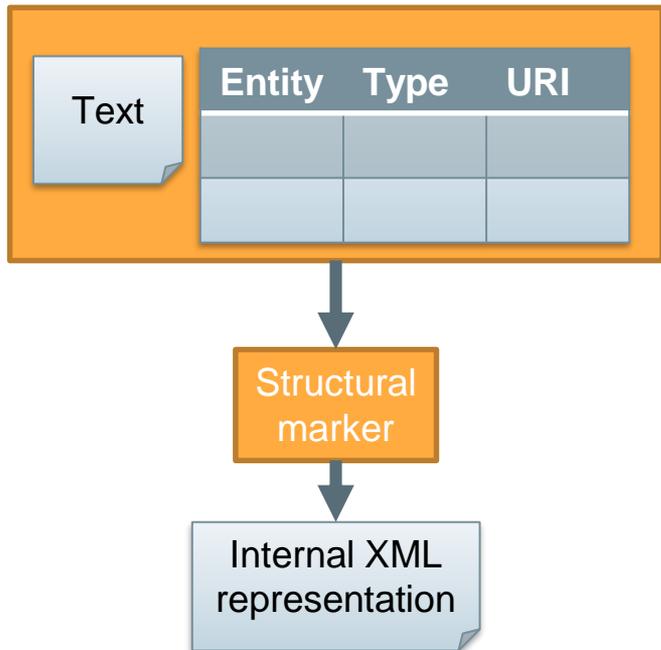
# 3. Structural marker

Detect structures in the text

Titles, subtitles, paragraphs, sections,...

Special structure for debates: participation

Regular expressions + custom rules



```
</Portada>
▼<Indice>
  <Titulo>INDICE GENERAL DE LA SESION</Titulo>
  ▼<CuerpoIndice>
    I.- SUMARIO DEL DEBATE II.- SUMARIO DE DOCUMENTOS III.- DOCUMENTOS
    DE LA CUENTA IV.- ASISTENCIA V.- TEXTO DEL DEBATE
    ▼<entity>
      <body bestClass="ORG" found="true" knownFraction="1"
      rdfLabel="camara de diputados"
      uri="http://datos.bcn.cl/recurso/cl/organismo/camara-de-
      diputados" uriProbability="0.8999994993209839">CAMARA DE
      DIPUTADOS</body>
      <class probability="0.8170063361" value="ORG"/>
    </entity>
  </CuerpoIndice>
</Indice>
▼<SumarioDelDebate>
  <Titulo>I.- SUMARIO DEL DEBATE Pág.</Titulo>
  ▼<CuerpoSumarioDelDebate>
    1.-La
```

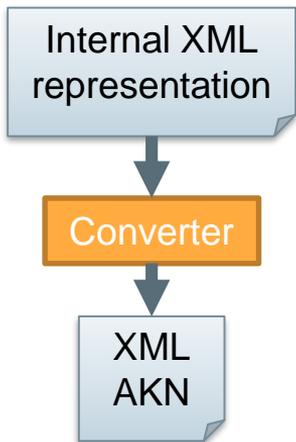


## 4. XML converter to Akom-Ntoso

Programmatic approach

Internal XML representation similar to DOM

Each node converted to text in AKN-XML



```
. PROSECRETARIO, el señor
<person bcn:classProbability="0.8929788338" bcn:found="true"
bcn:uriProbability="1.0466665840148925" id="instance_8_per"
refersTo="#per8">Parga, don Fernando</person>
~
</coverPage>
▼<preface>
  ▼<longTitle>
    <p>INDICE GENERAL DE LA SESION</p>
  </longTitle>
  ▼<toc>
    <tocItem href="/" level="0">I.- SUMARIO DEL DEBATE</tocItem>
    <tocItem href="/" level="0">II.- SUMARIO DE DOCUMENTOS</tocItem>
    <tocItem href="/" level="0">III.- DOCUMENTOS DE LA CUENTA</tocItem>
    <tocItem href="/" level="0">IV.- ASISTENCIA</tocItem>
    <tocItem href="/" level="0">V.- TEXTO DEL DEBATE CAMARA DE
DIPUTADOS</tocItem>
  </toc>
  ▼<longTitle>
    <p>I.- SUMARIO DEL DEBATE Pág.</p>
  </longTitle>
```

# Human edition of AKN-Documents



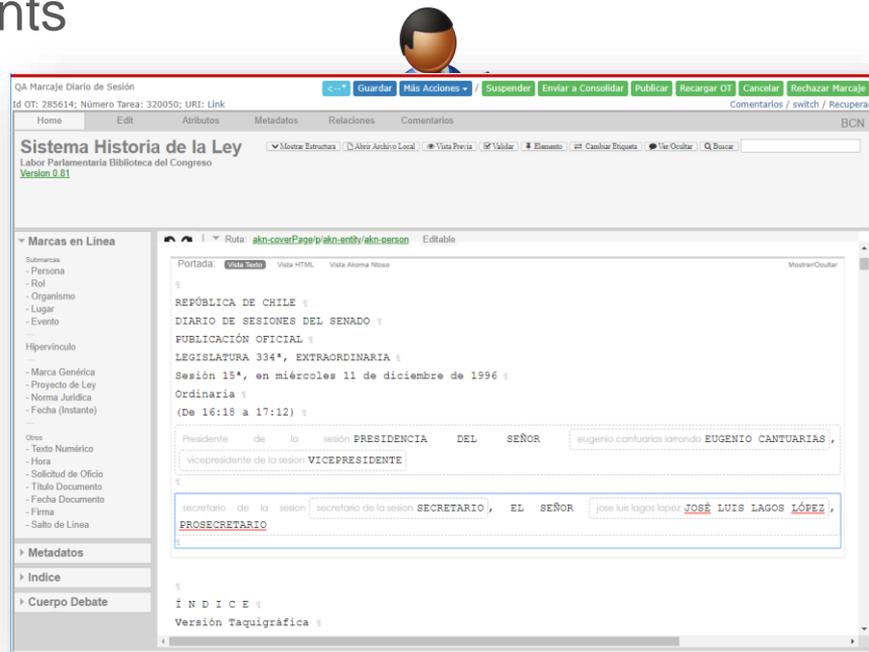
Quality assurance by human analysts

They review the generated XML documents

2 editors:

Ad-hoc XML editor

Commercial editor: LegisPro (Xcential)





# Linked data generation

The pilot project (2011) carefully defined a stable URI model

URIs have been maintained since then

URIs = IDs in the whole system

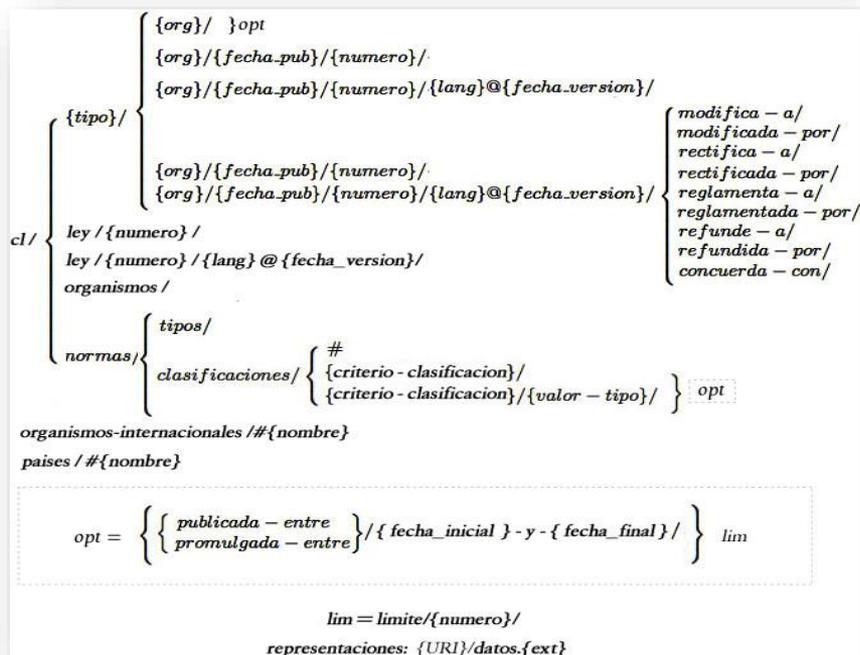
URIs are dereferentiable

Content negotiation

Custom linked data browser

Documentation (*in Spanish*)

<http://datos.bcn.cl/es/documentacion>

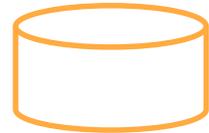


# AKN2RDF



## RDF extraction from Akoma-Ntoso XML

- Custom-made converter (XSL discarded for perceived complexity)
- Each XML tag implemented in one Class
- Extracted data saved into multiple databases (Relational and RDF)





# Linked data generation

Source: AKN XML documents

Target: RDF data

<http://datos.bcn.cl/recurso/cl/documento/579095.xml>

<http://datos.bcn.cl/recurso/cl/documento/579095/>



```

del Alfo Arturo Merino Benitez, al Aeropuerto de Los Cerrillos en Santiago y Aeropuerto Pudahuel
Libertador O'Higgins al Aeropuerto Internacional de Pudahuel de Santiago" name="Proyecto de ley"/>
<TLReference id="pinCHA0LS7A" href="http://datos.bcn.cl/recurso/temporal/1125" showAs="Autoriza la
Expropiación en Favores del Fisco de un Terreno Ubicado en la Localidad de Allinco, Comuna de Galvarino"
name="Proyecto de ley"/>
<TLReference id="r_nulo_aut" showAs="Recurso Nulo" name="r_nulo_aut"
href="http://datos.bcn.cl/recurso/nulo"/>
<TLReference id="seAcuerdaInsistir_aut" showAs="Se acuerda insistir" name="seAcuerdaInsistir_aut"
href="http://datos.bcn.cl/ontologias/bcn-resources/seAcuerdaInsistir"/>
</references>
</meta>
▼<coverPage id="akn579095-cp1">
  <p id="akn579095-cp1-p1">7REPUBLICA DE CHILE</p>
  <p id="akn579095-cp1-p2">CAMARA DE DIPUTADOS</p>
  <p id="akn579095-cp1-p3">LEGISLATURA ORDINARIA</p>
  <p id="akn579095-cp1-p4">
    <span id="ididm169254592"/>
  </p>
  <p id="akn579095-cp1-p5">Sesión 26, en martes 31 de julio de 1973</p>
  <p id="akn579095-cp1-p6">
    {Ordinaria: de
    <recordedTime id="tmpLMSULIHS" type="startEvent" time="1973-07-31T16:00:00Z">16</recordedTime>
    a
    <recordedTime id="tmp2H565JAJ" type="endEvent" time="1973-07-31T17:07:00Z">17. 07</recordedTime>
    </p>
  <p id="akn579095-cp1-p7">
    <entity name="AgrupadoresSemanticosPortada" referTo="#r01129" id="entityDRW4HQPQ">
      <role referTo="#r01129" id="instance_2_rol" bcn:classProbability="0.1409836338" bcn:found="true"
      bcn:uriProbability="1.0">Presidencia</role>
      de los señores
      <person referTo="#per3" id="instance_3_per" bcn:classProbability="0.9214812769" bcn:found="true"
      bcn:uriProbability="1.8622583150863647">Pareto, don Luis</person>
    }
    <person referTo="#per4" id="instance_4_per" bcn:classProbability="0.8893615685" bcn:found="true"
    bcn:uriProbability="1.9408557415008545">Lorca, don Gustavo</person>
    y
    <person referTo="#per6BLUYDWO" id="instance_5_per" bcn:classProbability="0.9372152293"
    bcn:found="true" bcn:uriProbability="1.0">Muñoz Barra, don Roberto</person>
  </entity>
  <entity name="AgrupadoresSemanticosPortada" referTo="#r016" id="entityURSIFFP2V">
    <role referTo="#r016" id="instance_6_rol" bcn:classProbability="0.2042947686" bcn:found="true"
    bcn:uriProbability="0.888884782791138">Secretario</rol>
    el señor
    <person referTo="#per7" id="instance_7_per" bcn:classProbability="0.9083771063" bcn:found="true"
    bcn:uriProbability="1.0262798285484314">Guerrero, don Radi</person>
  </entity>
  <entity name="AgrupadoresSemanticosPortada" referTo="#r018" id="entity2QL7FWCK">
    <role referTo="#r018" id="instance_8_rol" bcn:classProbability="0.1724517129" bcn:found="true"
    bcn:uriProbability="0.9999990463256836">Prosecretario</role>
    el señor
    <person referTo="#per9" id="instance_9_per" bcn:classProbability="0.9029382733" bcn:found="true"
    bcn:uriProbability="1.0456879949569702">Parga, don Fernando</person>
  </entity>

```

Biblioteca del Congreso Nacional de Chile   Datos Abiertos Enlazados

<http://datos.bcn.cl/recurso/cl/documento/579095>

- bcnres:hasWork = <http://datos.bcn.cl/recurso/cl/documento/579095/work>
- frbr:creator = <http://datos.bcn.cl/recurso/cl/organismo/camara-de-diputados>
- bcnres:tieneSeccionRecurso = <http://datos.bcn.cl/recurso/cl/documento/579095/seccion/akn579095-rc1>
- bcnres:tieneSeccionRecurso = <http://datos.bcn.cl/recurso/cl/documento/579095/seccion/akn579095-rc1>
- bcnres:tieneSeccionRecurso = <http://datos.bcn.cl/recurso/cl/documento/579095/seccion/akn579095-ds1>
- bcnres:tieneSeccionRecurso = <http://datos.bcn.cl/recurso/cl/documento/579095/seccion/akn579095-ad1>
- bcnres:tieneSeccionRecurso = <http://datos.bcn.cl/recurso/cl/documento/579095/seccion/akn579095-ds1>
- bcnres:tieneSeccionRecurso = <http://datos.bcn.cl/recurso/cl/documento/579095/seccion/akn579095-ds2>
- dc:identifier = "579095"^^xsd:integer
- bcnres:hasManifestation = <http://datos.bcn.cl/recurso/cl/documento/579095/manifestation>
- bcnres:tieneSeccionRecurso = <http://datos.bcn.cl/recurso/cl/legislatura/318/camara-de-diputados/sesion/ordinaria/26@1973-07-31>
- bcnres:tienePresidente = <http://datos.bcn.cl/recurso/persona/1557>
- bcnres:tienePresidente = <http://datos.bcn.cl/recurso/persona/635>
- bcnres:tienePresidente = <http://datos.bcn.cl/recurso/persona/972>
- rdfs:label = "Diario de sesión: Sesión ordinaria N° 26 de la Cámara de Diputados, Legislatura 1973"^^xsd:string
- dc:language = "es"^^xsd:string
- bcncon:perteneceA = <http://datos.bcn.cl/recurso/cl/legislatura/318>
- bcnres:tieneDocumentoTXT = <http://datos.bcn.cl/recurso/cl/documento/579095.txt>
- bcnres:tieneIntegranteDirectiva = <http://datos.bcn.cl/recurso/persona/1557>
- bcnres:tieneIntegranteDirectiva = <http://datos.bcn.cl/recurso/persona/635>
- bcnres:tieneIntegranteDirectiva = <http://datos.bcn.cl/recurso/persona/4465>
- bcnres:tieneIntegranteDirectiva = <http://datos.bcn.cl/recurso/persona/4464>
- bcnres:tieneIntegranteDirectiva = <http://datos.bcn.cl/recurso/persona/972>
- bcnres:tieneDocumentoAkomaNtoso = <http://datos.bcn.cl/recurso/cl/documento/579095.xml>
- dc:title = "Diario de sesión: Sesión ordinaria N° 26 de la Cámara de Diputados, Legislatura 1973"^^xsd:string
- rdf:type = foaf:Document
- rdf:type = bcnres:DiarioDeSesion
- bcnres:tieneSecretario = <http://datos.bcn.cl/recurso/persona/4464>
- bcnres:tieneResultado = <http://datos.bcn.cl/recurso/nulo>

Linked data browser (WESO-DESH)



# SPARQL endpoint

RDF triples are published as a public SPARQL endpoint

<http://datos.bcn.cl/sparql>

```
https://datos.bcn.cl/sparql

1 PREFIX bcnnorms: <http://datos.bcn.cl/ontologies/bcn-norms#>
2 PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
3 PREFIX dc: <http://purl.org/dc/elements/1.1/>
4
5 SELECT DISTINCT ?name (count(?norm) as ?number)
6 WHERE {
7   ?norm bcnnorms:createdBy ?organization .
8   ?organization bcnnorms:hasName ?name .
9   ?organization rdf:type bcnnorms:GovernmentalOrganization.
10  FILTER regex(?name, "^MUNICIPALIDAD")
11 }
12 ORDER BY DESC (?number)
13
```

Number of norms by municipality

|   | name                           | number |
|---|--------------------------------|--------|
| 1 | "MUNICIPALIDAD DE PROVIDENCIA" | "1257" |
| 2 | "MUNICIPALIDAD DE LAS CONDES"  | "930"  |
| 3 | "MUNICIPALIDAD DE SANTIAGO"    | "783"  |
| 4 | "MUNICIPALIDAD DE RANCAGUA"    | "546"  |
| 5 | "MUNICIPALIDAD DE LO PRADO"    | "383"  |
| 6 | "MUNICIPALIDAD DE VITACURA"    | "327"  |
| 7 | "MUNICIPALIDAD DE LA FLORIDA"  | "289"  |
| 8 | "MUNICIPALIDAD DE MACUL"       | "271"  |
| 9 | "MUNICIPALIDAD DE CONCHALÍ"    | "247"  |

# Content delivery



Web portals using Open Source Technologies

CMS (Typo3)

Python/Java

Varnish

Apache Lucene

REST Web service layers which connect to RDF triplestore and DB

Data exports to PDF, Doc and XML formats

URIs of parliamentary profiles = URIs in triplestore

# History of the Law portal



Historia de la Ley: Historia de la Ley N° 21.156

Biblioteca De https://www.bcn...

Usted está en: Portada Historia de la Ley N° 21.156

Buscador

Ingrese su búsqueda Buscar

Búsqueda Avanzada

Historia de la Ley N° 21.156

Establece la obligación de disponer de desfibriladores externos automáticos portátiles en los establecimientos y recintos que indica

Publicada: 20-05-2019

Pantalla completa Fabrica tu dossier

Tramitación Información adicional

1. Primer Trámite Constitucional: Senado

1.1. Moción Parlamentaria. Moción Parlamentaria: Sesión Ordinaria N° 38; Moción 9014-03

1.1. Moción Parlamentaria. Moción de los señores Senadores Gonzalo Uriarte Herrera, Guido Girardi Lavín, Fulvio Rossi Ciocca y Francisco Chahuán Chahuán. Fecha 03 de julio, 2013. Moción Parlamentaria en Sesión 38. Legislatura 361. Boletín N° 9.014-03 Proyecto de ley, iniciado en moción d...

Ver documento completo

1.2. Primer Informe de Comisión de Salud. Senado. Fecha 18 de mayo, 2018. Informe de Comisión de Salud en Sesión 20. Legislatura 366.

1.3. Discusión en Sala. Fecha 05 de junio, 2018. Diario de Sesión en Sesión 21. Legislatura 366. Discusión General. Se aprueba en general.

1.4. Boletín de Indicaciones. Fecha 11 de junio, 2018. Boletín de Indicaciones

Volver arriba

Historia de la Ley: Historia de la Ley N° 21.156

Biblioteca De https://www.bcn...

Usted está en: Portada Historia de la Ley N° 21.156

Buscador

Ingrese su búsqueda Buscar

Búsqueda Avanzada

Historia de la Ley N° 21.156

Establece la obligación de disponer de desfibriladores externos automáticos portátiles en los establecimientos y recintos que indica

Publicada: 20-05-2019

Pantalla completa Fabrica tu dossier

Tramitación Información adicional

Información de la norma

Id en Leychile: 1131706

Enlace a Leychile: https://www.leychile.cl/Navegar?IdNorma=1131706&IdVersion=2019-05-20

URI: http://datos.bcn.cl/recurso/cl/ley/ministerio-de-salud/2019-05-20/21156

Información del Proyecto de Ley

Número de boletín: 9014-03

Tramitación: Web Senado

URI: http://datos.bcn.cl/recurso/cl/proyecto-de-ley/9014-03

Iniciativa: Moción

Fecha de publicación: 2013-07-03

Gonzalo Uriarte Herrera

Francisco Chahuán Chahuán

Guido Girardi Lavín

Fulvio Rossi Ciocca

Volver arriba

https://www.bcn.cl/laborparlamentaria/wsgl/consulta/verLaborParlamentaria.py?IdPersona=1467

<https://www.bcn.cl/historiadelaley>



# History of the Law portal

Historia de la Ley: Navegar Ver

Senado Cámara de Diputados Transparencia Contacto/Sugerencias

BCN Ley Chile Observatorio Información Territorial Historia Política Formación Cívica

Búsqueda Normas destacadas Tratados Historia de la Ley Lévelo Personaliza Ayuda

HL

Historia de la Ley

Portada Acerca de este sitio

Usted está en: Portada Historia de la Ley N° 21.155

Navegar

Navegar versiones Comparar versiones

Tramitación de la Ley

Versiones Historia de la Ley N° 21.155

- Primer Trámite Constitucional (desde: 09-04-20)
- Primer Trámite Ingreso Proyecto

Artículo 2°

Artículo 3°

Artículo 4°

Artículo 5°

Artículo 6°

Artículo 7°

Artículo 8°

Artículo 9°

Artículo 10°

Oficio de Ley a Cámara Revisora

Segundo Trámite Constitucional (desde: 09-04-20)

Trámite Finalización (desde: 09-04-20)

Enviar a mi selección

Prim (Pri) Fecl (F) Cdm (C) Ses (S)

Legislatura: Legislatura Ordinaria número 362

Volver a Historia de la Ley completa

Proyecto de ley

**Artículo 1°.**

La presente ley tiene por objetivos principales:

- 1) Asimilar la lactancia con leche materna o lactancia materna como un derecho fundamental de la infancia, y como un derecho de las madres y a la vez como deber de éstas en caso que las condiciones de salud no lo impidan;
- 2) Fomentar, promover y proteger el amamantamiento o la lactancia por leche materna en todos los sectores de la sociedad como el medio óptimo e insustituible para la alimentación y desarrollo integral de niños y niñas a lo menos hasta sus dos años de edad, y además para la protección de su salud y la de sus madres;

Fomentar, promover y proteger el amamantamiento o la lactancia por leche materna en todos los sectores de la sociedad como el medio óptimo e insustituible para la alimentación y desarrollo integral de niños y niñas a lo menos hasta sus dos años de edad, y además para la protección de su salud y la de sus madres;

Historia de la Ley: Comparar Ver

Biblioteca Del C

Para dar cumplimiento a las disposiciones de la ley N° 11.902, de 18 de octubre de 1955, la Corporación de Servicios Habitacionales, dentro del plazo de ciento ochenta días contados desde la vigencia de esta ley, otorgará títulos de dominio en forma gratuita a los actuales ocupantes de la Población "Nueva Matucana", ubicada en la comuna de Quinta Normal y cuya propiedad se encuentra inscrita en el Registro de Propiedades del Conservador de Bienes Raíces del Departamento de Santiago a fojas 6.011, N° 7.106 del año 1961 y a fojas 10.584 vta. N° 12.740 del año 1961.

Facílitase a la Corporación de Servicios Habitacionales para que determine los deslindes particulares y cabida del predio que se transferirá a cada poblador.

**Artículo 2°.-**

Derégase el artículo 6° de la ley N° 11.902, de 18 de octubre de 1955.

**Artículo 3°.-**

La Corporación de Servicios Habitacionales, dentro del plazo de ciento ochenta días, contados desde la vigencia de esta ley, otorgará títulos de dominio en forma gratuita a los actuales ocupantes de la Población "El Pantano", ubicada en la ciudad de Valdivia, cuyos deslindes son: "Norte, Hospital John Kenney y Calle Bueras; Sur, calle Santa María; Oriente, calle 8 de octubre; y Poniente, varios propietarios.

Facílitase a la Corporación de Servicios Habitacionales para que determine los deslindes particulares y cabida del predio que se transferirá a cada poblador.

**Artículo 4°.-**

La Corporación de Servicios Habitacionales deberá otorgar títulos de dominio a los actuales ocupantes de las Poblaciones "Irene Frey" y "Alcalde Dr. Enrique Dintrans" de Rancagua, dentro del plazo de 180 días contado desde la fecha de vigencia de la presente ley.

Facílitase a la Junta Directiva de la Corporación de Servicios Habitacionales para convenir condiciones de pago especiales con los actuales ocupantes de la Población "Irene Frey" que se encuentren atrasados en el pago del dividendo provisorio para los efectos de acceder al inciso anterior.

**Artículo 5°.-**

Si por razones de orden técnico para el desarrollo de obras de

El Servicio Nacional de Salud o el Servicio Público que corresponda, dentro del plazo de ciento ochenta días contados desde la vigencia de esta ley, otorgará títulos de dominio, en forma gratuita, a los actuales ocupantes de la Población "El Pantano", ubicada en la ciudad de Valdivia, cuyos deslindes son: Norte, Hospital John Kennedy y calle Bueras; Sur, calle Santa María, Oriente, calle 8 de octubre, y Poniente, varios propietarios.

Facílitase a la Corporación de Servicios Habitacionales para que determine los deslindes particulares y cabida del predio que se transferirá a cada poblador.

**Artículo 2°.-**

La Corporación de Servicios Habitacionales deberá otorgar títulos de dominio a los actuales ocupantes de las poblaciones "Irene Frey" y "Alcalde Dr. Enrique Dintrans", de Rancagua, dentro del plazo de 180 días, contado desde la fecha de vigencia de la presente ley.

Facílitase a la Junta Directiva de la Corporación de Servicios Habitacionales para convenir condiciones de pago especiales con los actuales ocupantes de la población "Irene Frey" que se encuentren atrasados en el pago del dividendo provisorio para los efectos de acceder al inciso anterior.

**Artículo 3°.-**

Si por razones de orden técnico para el desarrollo de obras de urbanización en la población El Pantano, de Valdivia, no se pudiere otorgar títulos de dominio a determinados ocupantes, se les dará preferencia para ser ubicados en otros sectores, dentro de los planes de asignación de terrenos o viviendas de la Corporación de Servicios Habitacionales."

En estos casos, la Corporación de Servicios Habitacionales deberá conceder condiciones especiales de pago, sin sujeción a su ley orgánica o a los reglamentos vigentes sobre la materia.

**Artículo 4°.-**

La Corporación de Servicios Habitacionales transferirá gratuitamente, a sus actuales ocupantes, los terrenos que se expropien al Servicio Agrícola y Ganadero y en los cuales se encuentran ubicadas las poblaciones "Manuel Rodríguez", "Teniente Merino" y "Pampa Irigoin", de la comuna de Puerto Montt.

En iguales condiciones, transferirá los terrenos que se desalfecten de su calidad de bienes nacionales de uso público y que estén ocupados por las poblaciones mencionadas en el inciso anterior.

<https://www.bcn.cl/historiadelaley>

# Parliamentary Work



This screenshot shows the profile page for Gabriel Valdés Subercaseaux on the website <https://www.bcn.cl/>. The page includes a portrait of the member, his name, and biographical information: "Nacimiento: 03 de julio de 1919, Santiago Chile" and "Fallecimiento: 07 de septiembre de 2011, Santiago". There are social media sharing buttons for Facebook, Twitter, and a '+1' button. Below the profile, there are tabs for "Pantalla Completa", " Trayectoria parlamentaria", " Biografía", and " Mociónes en Ley". The main content area is titled "Labor parlamentaria completa" and features a navigation menu with "Labor Legislativa", " Incidentes", " Homenajes", " Comisiones y grupos", and " Permisos y otros". A list of legislative acts is shown, with the first entry being "Moción Inadmisible" under the "Intervención" category, dated "Página 1 de 78". The entry text reads: "DESIGNACIÓN DE SEÑOR SERGIO MUÑOZ GAJARDO COMO MINISTRO DE LA CORTE SUPREMA" and "Intervención de Gabriel Valdés Subercaseaux en Senado el 04 de octubre de 2005". A "Ver más" button is visible below the text.

Show participation of each Member of Parliament

This screenshot shows the "Participación de Gabriel Valdés Subercaseaux" page on the website. At the top, there is a search bar with the text "Buscar por Parlamentario o palabra clave" and a "Buscar" button. Below the search bar, there is a navigation menu with "Volver", "Opciones", and "Buscar dentro del texto". The main heading is "Participación de Gabriel Valdés Subercaseaux", followed by social media sharing buttons. The content area is titled "DESIGNACIÓN DE SEÑOR SERGIO MUÑOZ GAJARDO COMO MINISTRO DE LA CORTE SUPREMA" and includes a "Ver antecedente" button. The text of the intervention is displayed, starting with "El señor VALDÉS.- Señor Presidente, creo que esta discusión ha excedido la naturaleza de los hechos. Aquí no se trata de probidad ni de situaciones mejores o peores. Están vigentes la capacidad y la atribución del Senado en orden a votar 'sí' o 'no' por el nombre que sugirió el Jefe del Estado para ocupar el cargo de Ministro de la Corte Suprema...". The text continues with "Sin embargo, el Presidente del Senado ha traído a colación una reforma a nuestro sistema, la cual ha merecido objeciones. Y hay prácticamente acuerdo para que la interpretación pertinente, que es sustancial, vaya a la Comisión de Constitución. Si el señor Presidente comparte este criterio, no la apliquemos ahora." and "El señor ZALDÍVAR (don Adolfo) - ¡Obvio! El señor VALDÉS.- No tiene sentido hacerlo. Por lo tanto, o el asunto no va a Comisión y el Presidente del Senado obtiene un pronunciamiento favorable a su tesis, o bien, debe remitirse al mencionado órgano técnico. De ocurrir esto último, queda suspendida la proposición de la Mesa, correspondiendo a la Sala pronunciarse, conforme al antiguo sistema, sobre la designación de Ministro de la Corte Suprema." and "El señor ZALDÍVAR (don Adolfo) - ¡Es que no se respetó! El señor VALDÉS.-".



# Some experimental visualizations

Relationships between laws

Historical Parliament

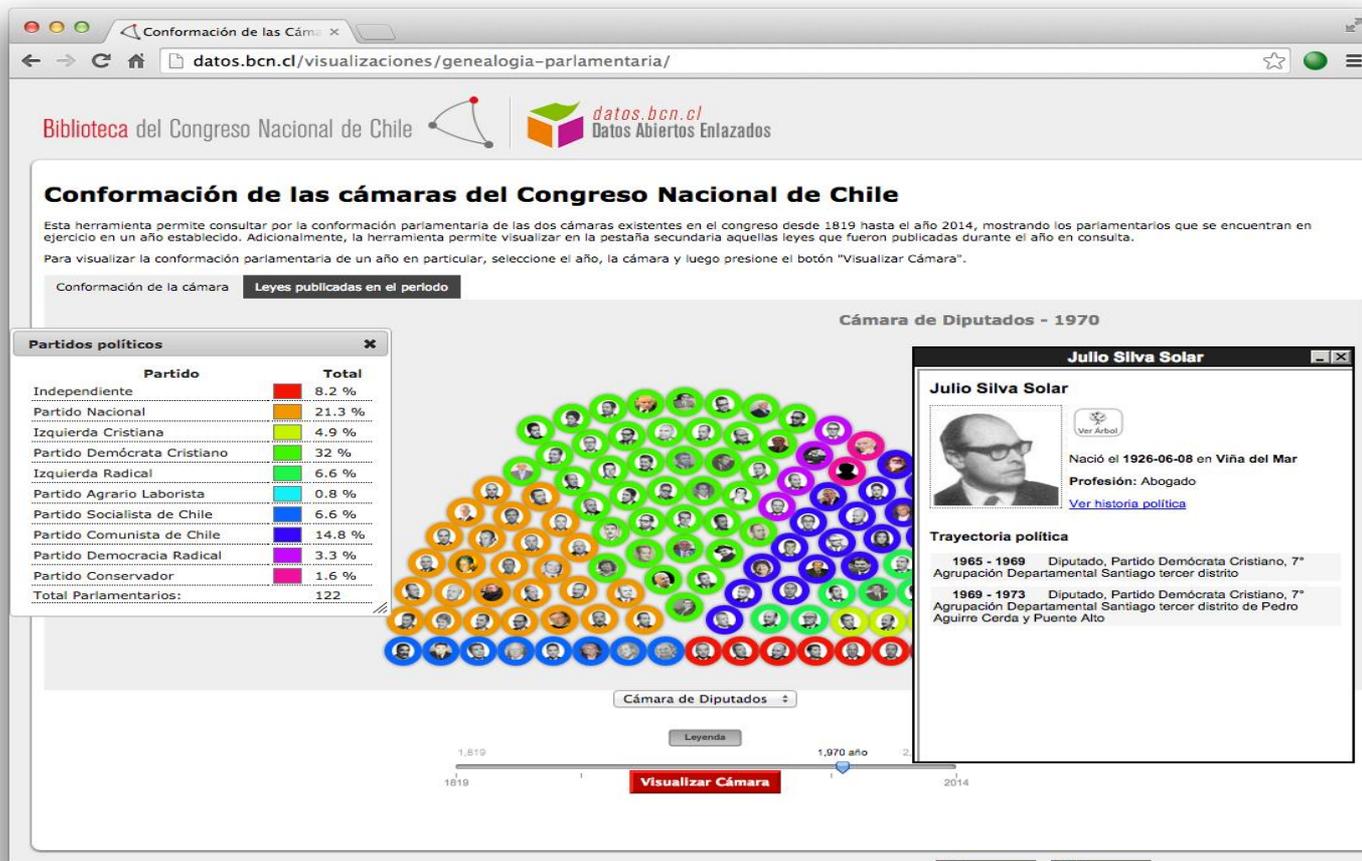
Parliamentary genealogy (family relationships)

Regions mentioned in laws (legislative hackathon)





# Historical parliament



# Parliamentary genealogy



Herramienta de verificación de x +  
https://datos.bcn.cl/visualizaciones/genealo...

## Buscador de relaciones familiares

Escriba los nombres de políticos en cada una de las cajas y presione "Buscar relación familiar".

**Salvador Allende Gossens**



Fecha nacimiento: 1908-07-26  
Fecha fallecimiento: 1973-10-11  
Lugar nacimiento: Valparaíso  
Lugar fallecimiento: Santiago

**Sebastián Piñera Echeñique**



Fecha nacimiento: 1949-12-01  
Lugar nacimiento: Santiago

### ¿Cómo interpretar el resultado?

Los siguientes íconos representan la relación existente entre dos personas

| Familiar      | relación  | Familiar |
|---------------|---|----------|
| hijo          |  | padre    |
| hijo          |  | madre    |
| padre o madre |  | hijo     |
| cónyugue      |  | cónyugue |

**Ejemplo:** Pedro es hijo de María  
hijo  madre

**Buscar relación familiar**

**¿qué es esto?**

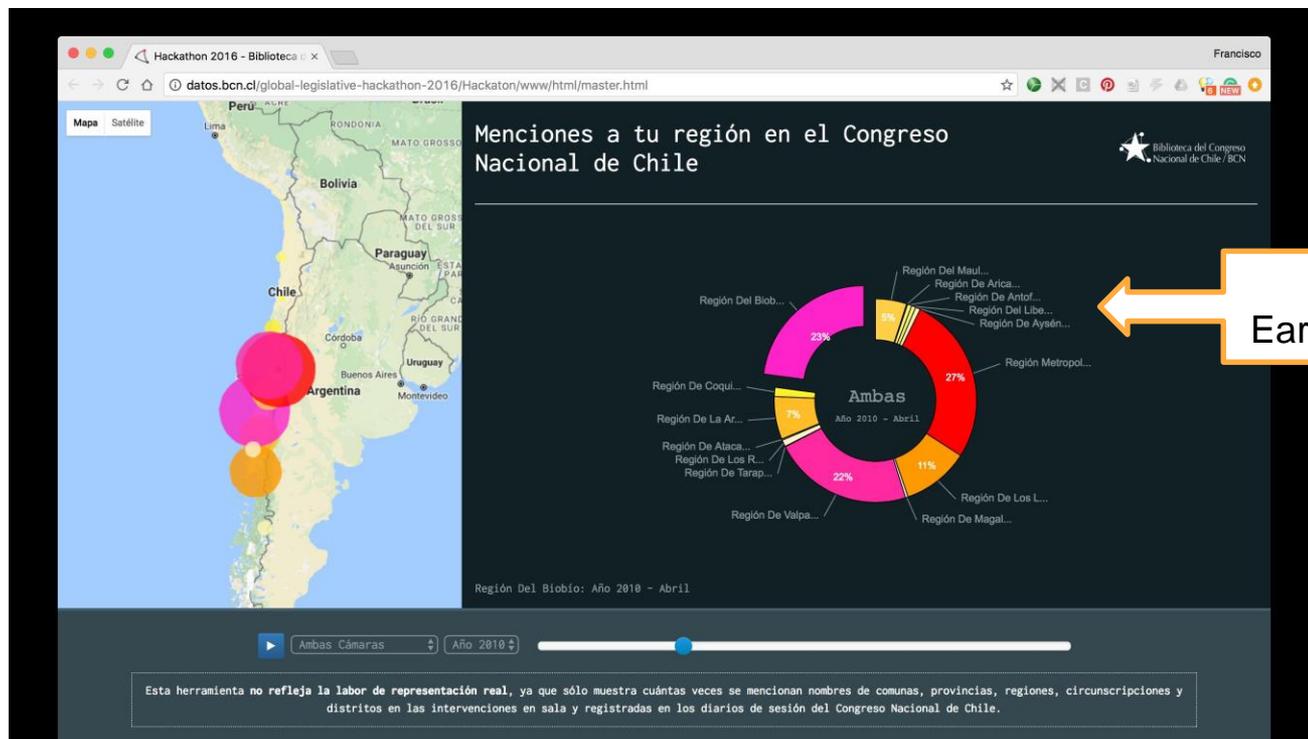
Salvador Allende Gossens  Laura Gossens Uribe  Laura Allende Gossens  Denise Pascal Allende  Jorge Chadwick Vergara  Jorge Enrique Chadwick Valdés  Roberto Tomás Chadwick Castro  Herman Chadwick Valdés  Polette Piñera Carvalho  José Manuel Piñera Figueroa  José Piñera Carvalho  Sebastián Piñera Echeñique

Sebastián Piñera Echeñique es **Sobrino del Cuñado del Suegro del Sobrino** de Salvador Allende Gossens

# Regions mentioned by law



Result of a legislative hackathon



In 2010 there was an Earthquake in BioBio region

<http://datos.bcn.cl/global-legislative-hackathon-2016/Hackaton/www/html/master.html>



# Statistics

24.368 documents (nov. 2018)

Number of RDF triples: 28 millions

According to Google analytics

Average browsing time: 2min 26s

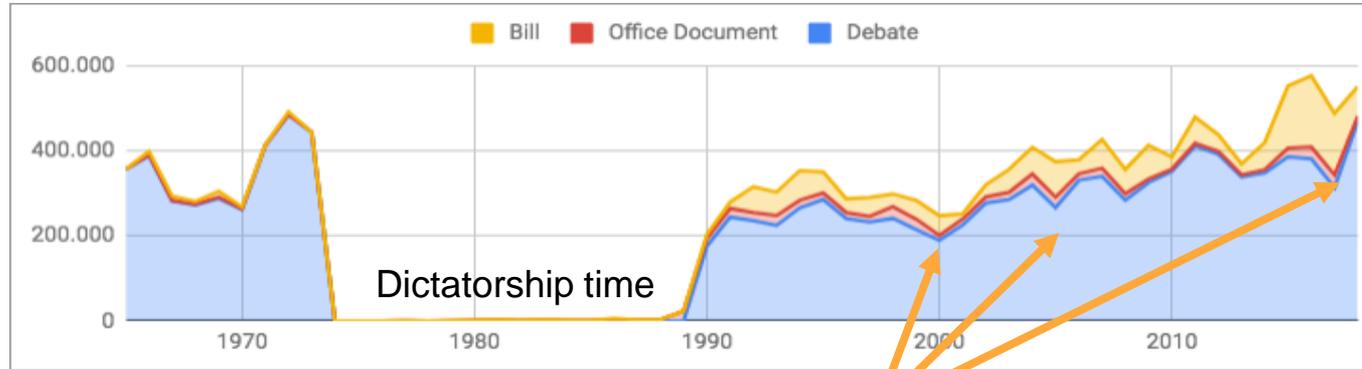
Visits received 331,481 (nov. 2016-2017)  $\Rightarrow$  476,241 (nov. 2017-2018)

| Dataset                       |                    | RDF triples by document |                |               |             |                |             |               |
|-------------------------------|--------------------|-------------------------|----------------|---------------|-------------|----------------|-------------|---------------|
| <i>Document type</i>          | <i>Total docs.</i> | <i>Min.</i>             | <i>1st Qu.</i> | <i>Median</i> | <i>Mean</i> | <i>3rd Qu.</i> | <i>Max.</i> | <i>Stdev.</i> |
| Debate Senate                 | 3.614              | 35                      | 665            | 1.001         | 1.097       | 1.353          | 6.883       | 668,67        |
| Debate Chamber of<br>Deputies | 4.298              | 35                      | 1.166          | 1.640         | 1.788       | 2.199          | 11.771      | 961,42        |
| Bills                         | 2.514              | 37                      | 109            | 308           | 689         | 864            | 15.934      | 1055,93       |
| Others types                  | 13.942             | 37                      | 40             | 40            | 49,78       | 40             | 15.720      | 245,68        |
| Total                         | 24.368             |                         |                |               |             |                |             |               |



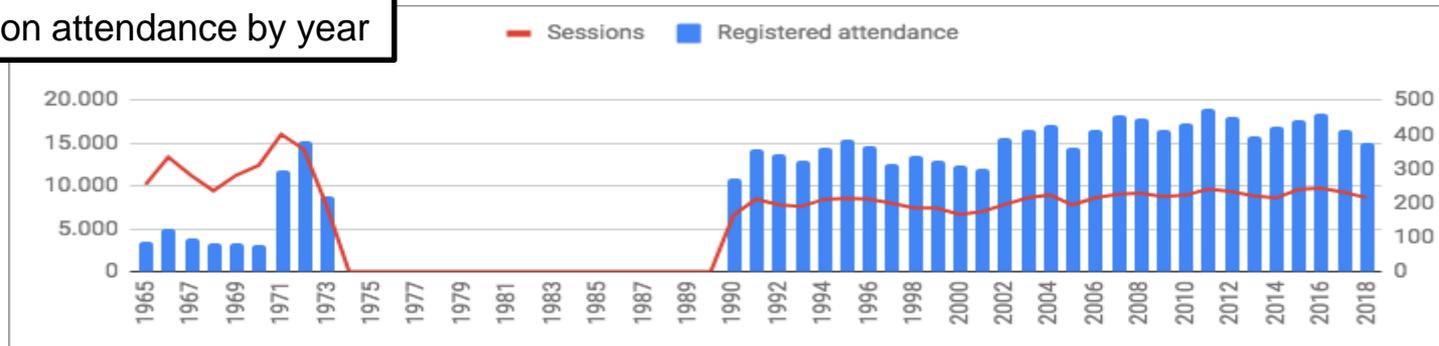
# ...and some findings...

RDF triples generated by year



Question: why are there some valleys?

Session attendance by year



# Lessons learnt



## RDF granularity & inference trade-off

RDF statements + inference (high running times...queries that didn't terminate)

*A priori* inferred triples added to triple store (high response times for large docs)

Small subset of RDF triples (structural parts of docs and metadata)

## Performance problems in XML editor browsing long docs (>1000pages)

## Low SPARQL endpoint usage by external apps

If we could start again, I would recommend ShEx

**Personal note:** These kind of projects led to my interest in ShEx

# Conclusions & future projects



Well designed URIs can act as a perfect glue for interoperability

Automatic workflow pipelines help long-term survival of LD-based projects

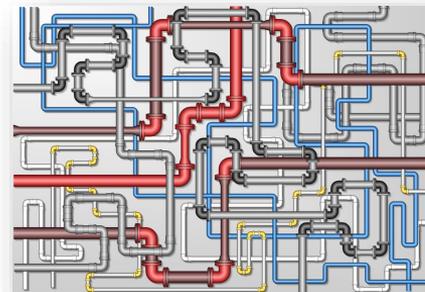
SPARQL endpoint since 2011

Future projects on top of existing ones

National Budget as Linked data

Diana Project: Members of Parliament linked to social network analysis

New portal: User customization & recommender systems





2<sup>nd</sup> part

# Linked data quality and Shapes



# RDF, the good parts...

RDF as an integration language

RDF as a *lingua franca* for semantic web and linked data

RDF flexibility & integration

- Data can be adapted to multiple environments

- Open and reusable data by default

RDF for knowledge representation

RDF data stores & SPARQL

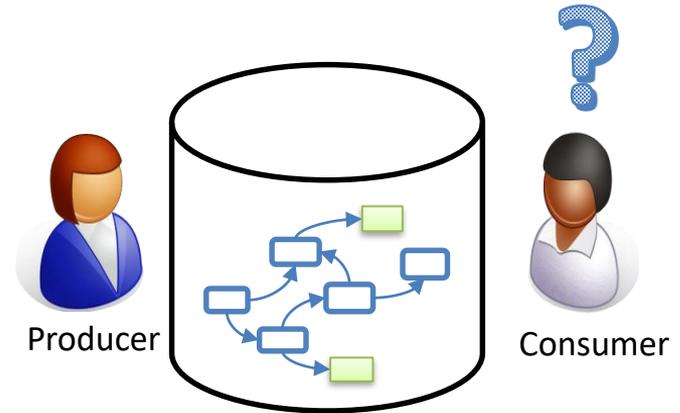
# RDF, the other parts

## Consuming & producing RDF

Multiple syntaxes: Turtle, RDF/XML, JSON-LD, ...

Embedding RDF in HTML

Describing and validating RDF content



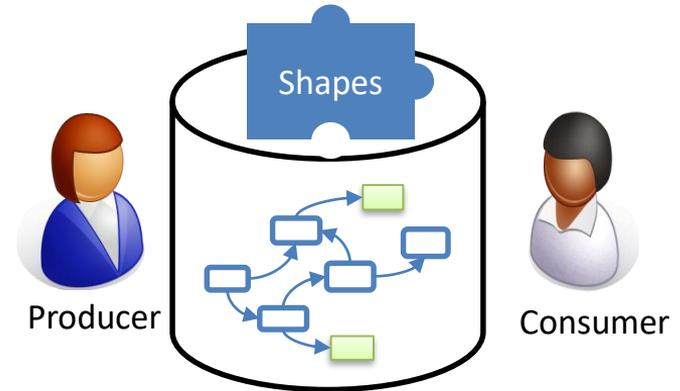
# Why describe & validate RDF?

## For producers

- Developers can understand the contents they are going to produce
- Ensure they produce the expected structure
- Advertise and document the structure
- Generate interfaces

## For consumers

- Understand the contents
- Verify the structure before processing it
- Query generation & optimization





# Similar technologies

| Technology           | Schema                               |
|----------------------|--------------------------------------|
| Relational Databases | DDL                                  |
| XML                  | DTD, XML Schema, RelaxNG, Schematron |
| Json                 | Json Schema                          |
| RDF                  | ?                                    |

Fill that gap



# ShEx and SHACL

2013 RDF Validation Workshop

Conclusions of the workshop:

*There is a need of a higher level, concise language for RDF Validation*

ShEx initially proposed (v 1.0)

2014 W3c Data Shapes WG chartered

2017 SHACL accepted as W3C recommendation

2017 ShEx 2.0 released as Community group draft

2019 ShEx adopted by Wikidata



# Short intro to ShEx

ShEx (Shape Expressions Language)

Concise and human-readable language for RDF validation & description

Syntax similar to SPARQL, Turtle

Semantics inspired by regular expressions & RelaxNG

2 syntaxes: Compact and RDF/JSON-LD

Official info: <http://shex.io>

Semantics: <http://shex.io/shex-semantics/>, primer: <http://shex.io/shex-primer>

# Simple example

Prefix declarations  
as in  
Turtle/SPARQL

```
prefix schema: <http://schema.org/>
prefix xsd:    <http://www.w3.org/2001/XMLSchema#>

<User> IRI {
  schema:name  xsd:string    ;
  schema:knows @<User>      *
}
```

Nodes conforming to **<User>** shape must:

- Be IRIs
- Have exactly one `schema:name` with a value of type `xsd:string`
- Have zero or more `schema:knows` whose values conform to **<User>**

# RDF Validation using ShEx

Schema

```
<User> IRI {  
  schema:name xsd:string ;  
  schema:knows @<User> *  
}
```

Shape map

```
:alice@<User> ✓  
:bob @<User> ✓  
:carol@<User> ✗  
:dave @<User> ✗  
:emily@<User> ✗  
:frank@<User> ✓  
:grace@<User> ✗
```

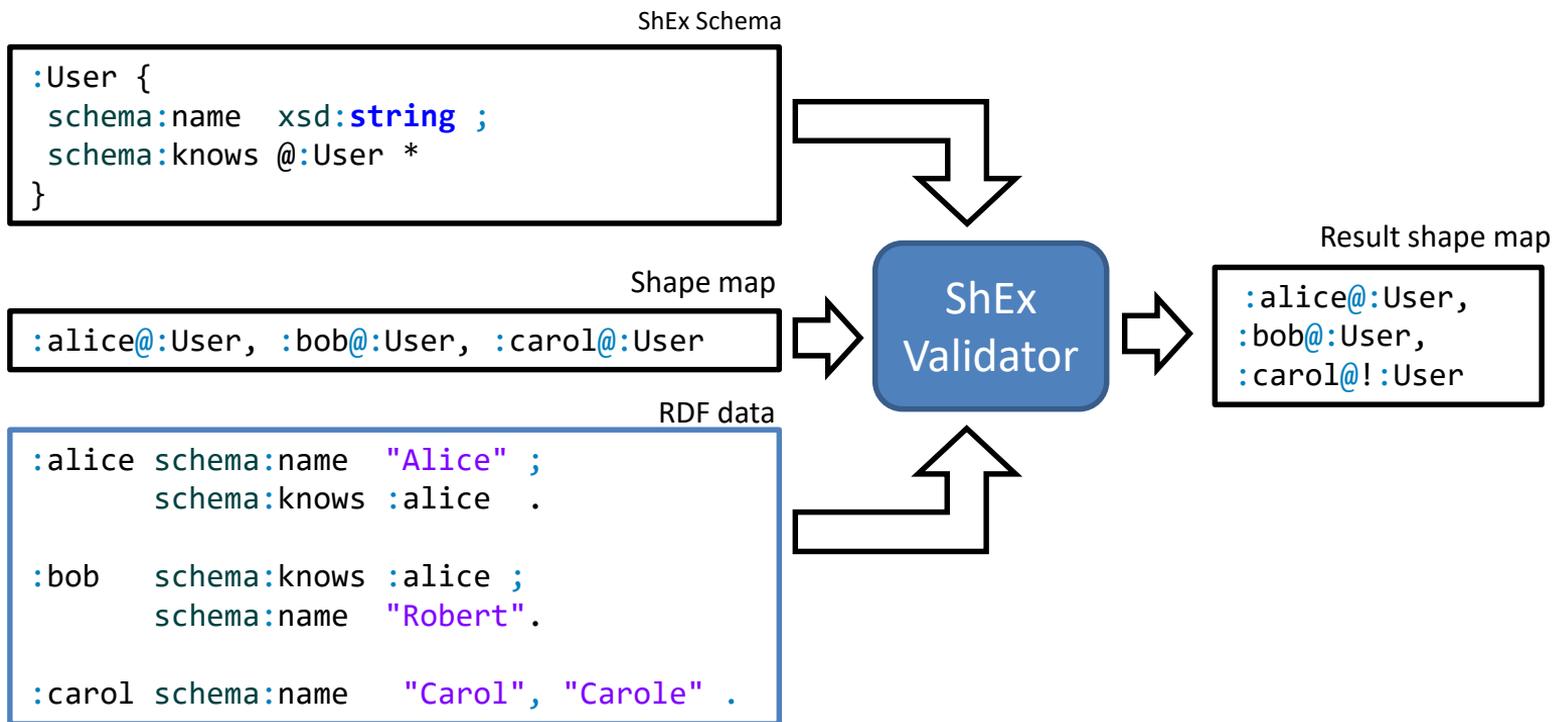
```
:alice schema:name "Alice" ;  
      schema:knows :alice .  
  
:bob schema:knows :alice ;  
     schema:name "Robert".  
  
:carol schema:name "Carol", "Carole" .  
  
:dave schema:name 234 .  
  
:emily foaf:name "Emily" .  
  
:frank schema:name "Frank" ;  
       schema:email <mailto:frank@example.org> ;  
       schema:knows :alice, :bob .  
  
:grace schema:name "Grace" ;  
       schema:knows :alice, _:1 .  
  
_:1 schema:name "Unknown" .
```

Try it (RDFShape): <https://goo.gl/97bYdv>Try it (ShExDemo): <https://goo.gl/Y8hBsw>

# Validation process

**Input:** RDF data, ShEx schema, Shape map

**Output:** Result shape map

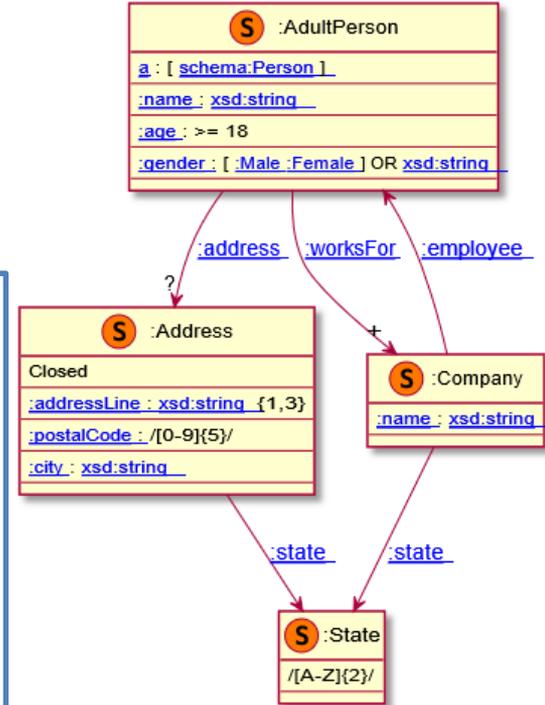




# Example with more ShEx features

```
:AdultPerson EXTRA rdf:type {  
  rdf:type [ schema:Person ] ;  
  :name xsd:string ;  
  :age MinInclusive 18 ;  
  :gender [ :Male :Female ] OR xsd:string ;  
  :address @:Address ? ;  
  :worksFor @:Company +  
}  
:Address CLOSED {  
  :addressLine xsd:string {1,3}  
  :postalCode /[0-9]{5}/  
  :state @:State  
  :city xsd:string  
}  
:Company {  
  :name xsd:string  
  :state @:State  
  :employee @:AdultPerson *  
}  
:State /[A-Z]{2}/
```

```
:alice rdf:type :Student, schema:Person ;  
  :name "Alice" ;  
  :age 20 ;  
  :gender :Male ;  
  :address [  
    :addressLine "Bancroft Way" ;  
    :city "Berkeley" ;  
    :postalCode "55123" ;  
    :state "CA"  
  ] ;  
  :worksFor [  
    :name "Company" ;  
    :state "CA" ;  
    :employee :alice  
  ] .
```



Try it: <https://tinyurl.com/yd5hp9z4>



# SHACL

SHACL (Shapes Constraint Language)

W3C recommendation:

<https://www.w3.org/TR/shacl/> (July 2017)

RDF vocabulary

2 parts: SHACL-Core, SHACL-SPARQL



# Same example with blank nodes

```
prefix :      <http://example.org/>
prefix sh:    <http://www.w3.org/ns/shacl#>
prefix xsd:   <http://www.w3.org/2001/XMLSchema#>
prefix schema: <http://schema.org/>
```

```
:UserShape a sh:NodeShape ;
  sh:targetNode :alice, :bob, :carol ;
  sh:nodeKind sh:IRI ;
  sh:property [
    sh:path schema:name ;
    sh:minCount 1; sh:maxCount 1;
    sh:datatype xsd:string ;
  ] ;
  sh:property [
    sh:path schema:email ;
    sh:minCount 1; sh:maxCount 1;
    sh:nodeKind sh:IRI ;
  ] .
```

```
:alice schema:name "Alice Cooper" ;
        schema:email <mailto:alice@mail.org> .

:bob    schema:firstName "Bob" ;
        schema:email <mailto:bob@mail.org> . ☹️

:carol  schema:name "Carol" ;
        schema:email "carol@mail.org" . ☹️
```

Data graph

Shapes graph

Try it. RDFShape <https://goo.gl/ukY5vq>

# Longer example

## In ShEx

```

:AdultPerson EXTRA a {
  a      [ schema:Person ] ;
  :name  xsd:string ;
  :age   MinInclusive 18 ;
  :gender [ :Male :Female ] OR xsd:string ;
  :address @:Address ? ;
  :worksFor @:Company + ;
}
:Address CLOSED {
  :addressLine xsd:string {1,3} ;
  :postalCode  /[0-9]{5}/ ;
  :state       @:State ;
  :city        xsd:string ;
}
:Company {
  :name      xsd:string ;
  :state     @:State ;
  :employee  @:AdultPerson * ;
}
:State      /[A-Z]{2}/

```

## In SHACL

```

:AdultPerson a sh:NodeShape ;
  sh:property [
    sh:path rdf:type ;
    sh:qualifiedValueShape [
      sh:hasValue schema:Person
    ] ;
  sh:qualifiedPropertyPath sh:addressLine ;
  sh:qualifiedPropertyPath sh:worksFor ;
] ;
:Address a sh:NodeShape ;
  sh:closed true ;
  sh:property [ sh:path :addressLine ;
    sh:datatype xsd:string ;
  ] ;
  sh:property [ sh:path :state ;
    sh:node :State ;
  ] ;
  sh:property [ sh:path :employee ;
    sh:node :AdultPerson ;
  ] ;
] ;
:Company a sh:NodeShape ;
  sh:property [ sh:path :name ;
    sh:datatype xsd:string ;
  ] ;
  sh:property [ sh:path :state ;
    sh:node :State ;
  ] ;
  sh:property [ sh:path :employee ;
    sh:node :AdultPerson ;
  ] ;
] ;
:State a sh:NodeShape ;
  sh:pattern "[A-Z]{2}" .

```

Its recursive!!! (not well defined SHACL)  
Implementation dependent feature

Try it: <https://tinyurl.com/ycl3mkzr>

# Some challenges and perspectives

Theoretical foundations of ShEx/SHACL

Generating shapes from data

Validation Usability

RDF Stream validation

Schema ecosystems

Wikidata

Solid



# Theoretical foundations of ShEx/SHACL

## Conversion between ShEx and SHACL

SHaclEX library converts subsets of both

### Challenges

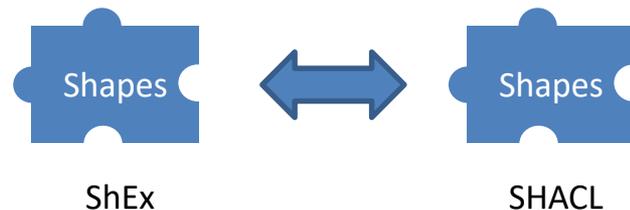
Recursion and negation

Performance and algorithmic complexity

Detect useful subsets of the languages

Convert to SPARQL

## Schema/data mapping

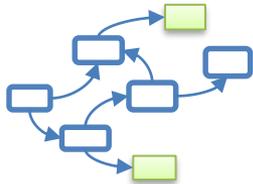
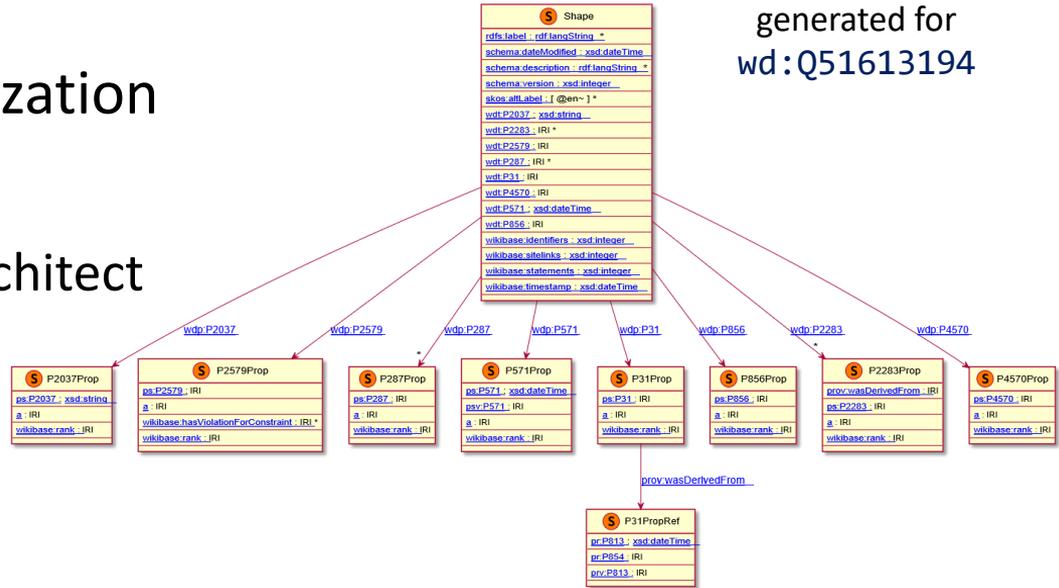




# Generating Shapes from Data

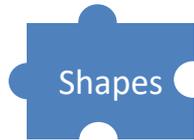
Useful use case in practice  
Knowledge Graph summarization  
Some prototypes:  
ShExer, RDFShape, ShapeArchitect

Shape Expression  
generated for  
wd:Q51613194



RDF data

infer



Try it with RDFShape:  
<https://tinyurl.com/y8pjcbfy>

# Validation usability

## Learning from users

Early adopters: WebIndex, HL7 FHIR, Eclipse Lyo, GenWiki,...

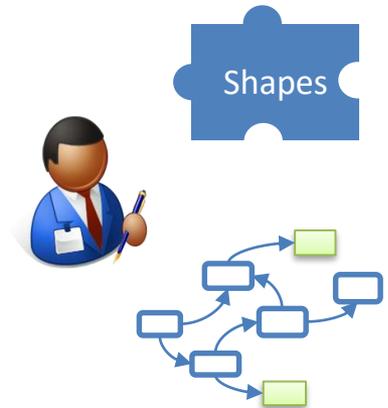
Improve error information/visualization/navigation/repairing

## Authoring/visualization tools

## Propose annotation sets

UI generation

Error reporting/suggestion (SHOULD/MUST/...)



# RDF Stream validation

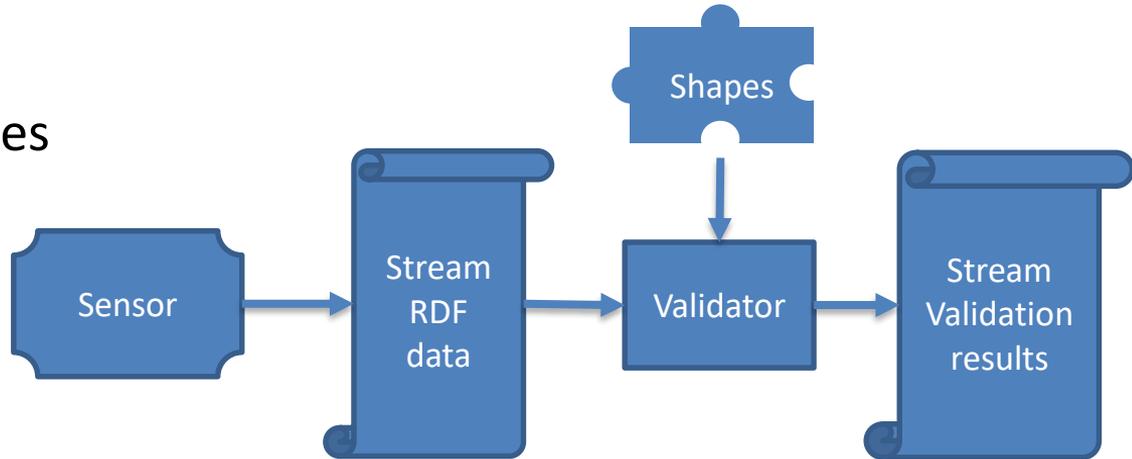
## Validation of RDF streams

### Challenges:

Incremental validation

Named graphs

Addition/removal of triples



# Schema ecosystems: Wikidata



In May, 2019, Wikidata announced ShEx adoption

New namespace for schemas

Example: <https://www.wikidata.org/wiki/EntitySchema:E2>

It opens lots of opportunities/challenges

Schema evolution and comparison

A screenshot of a web browser showing the Wikidata page for EntitySchema:E2. The browser's address bar shows the URL 'https://wikidata-shex.wmflabs.org/wiki/EntitySchema:E2'. The page features a blue hexagonal logo with a red hexagon inside. Below the logo are links for 'Main page', 'Recent changes', 'Random page', and 'Help about MediaWiki'. The main content area has tabs for 'EntitySchema' and 'Discussion', with 'EntitySchema' selected. Below the tabs is a search bar and a table with the following data:

| language code | label   | description  | aliases                  | edit                 |
|---------------|---------|--|--------------------------|----------------------|
| en            | Disease | Shape Expression for diseases coming from Disease Ontology in Wikidata ( Source: <a href="https://github.com/andrawaag/Genewiki-ShEx/blob/master/diseases/Wikidata-Disease-Ontology.shex">https://github.com/andrawaag/Genewiki-ShEx/blob/master/diseases/Wikidata-Disease-Ontology.shex</a> ) | Sickness<br> <br>Illness | <a href="#">edit</a> |



# Schema ecosystems: Solid project

SOLID (SOcial Linked Data): Promoted by Tim Berners-Lee

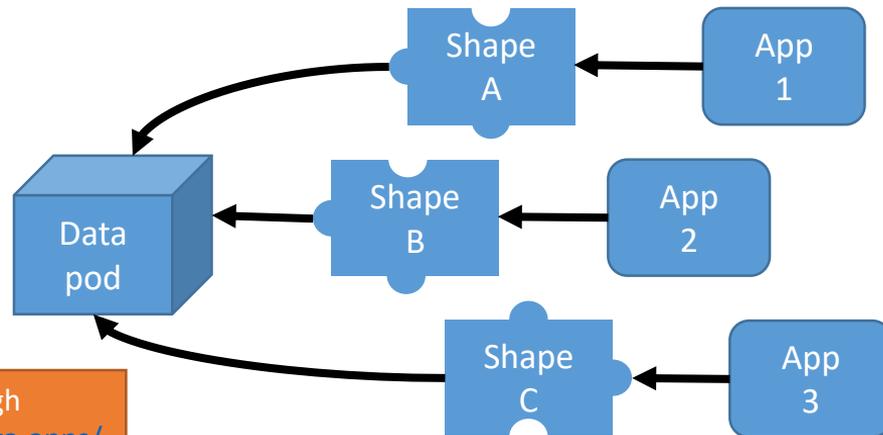
Goal: Re-decentralize the Web

Separate data from apps

Give users more control about their data

Internally using linked data & RDF

Shapes needed for interoperability



"...I just can't stop thinking about shapes.", Ruben Verborgh  
<https://ruben.verborgh.org/blog/2019/06/17/shaping-linked-data-apps/>



# Conclusions

Explicit schemas (shapes) can improve linked data quality

2 languages proposed: ShEx/SHACL

Towards an ecosystem of shapes for data portals

New challenges and opportunities



# More info

## About ShEx

ShEx by Example (slides):

[https://figshare.com/articles/ShExByExample\\_pptx/6291464](https://figshare.com/articles/ShExByExample_pptx/6291464)

ShEx chapter from Validating RDF data book:

<http://book.validatingrdf.com/bookHtml010.html>

## About SHACL

SHACL by example (slides):

[https://figshare.com/articles/SHACL\\_by\\_example/6449645](https://figshare.com/articles/SHACL_by_example/6449645)

SHACL chapter at Validating RDF data book

<http://book.validatingrdf.com/bookHtml011.html>

## Comparing ShEx and SHACL

<http://book.validatingrdf.com/bookHtml013.html>

# End of presentation

Acknowledgements:

David Vilches, Eridan Otto, Christian Sifaqui