



**LOD** BRASIL  
LINKED OPEN DATA

**19a21** novembro 2014

---

# IodKEM – Linked Open Data & Knowledge Engineering and Management

<José Leomar Todesco>

<LEC/EGC/UFSC>

# AGENDA

- Sobre a W3C e a apresentação do Phil Archer;
- Alguns grupos de trabalho da W3c em Linked Data
- Pesquisa do INESC
- Iniciativas do LEC
  - O projeto ontoKEM
  - O projeto lodKEM

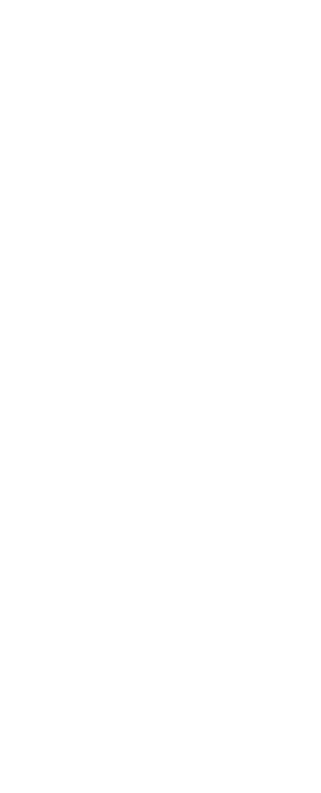
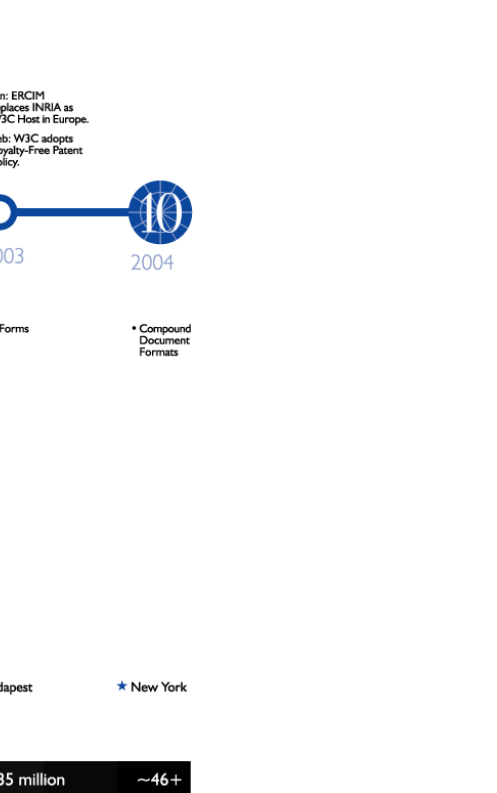
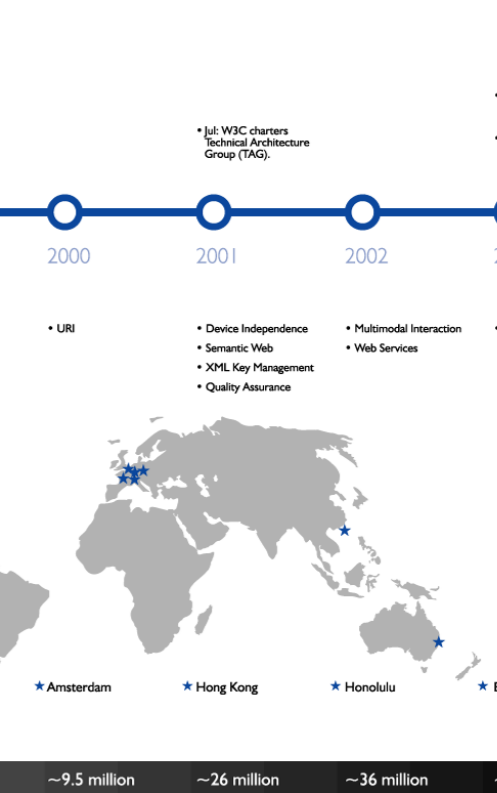
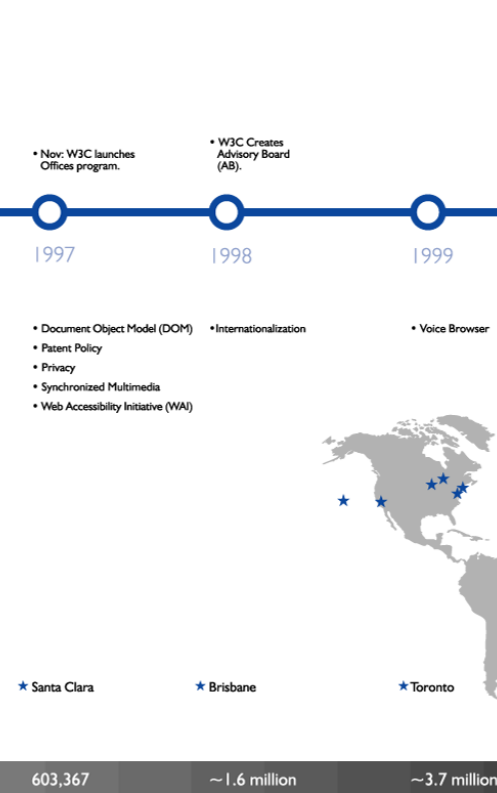
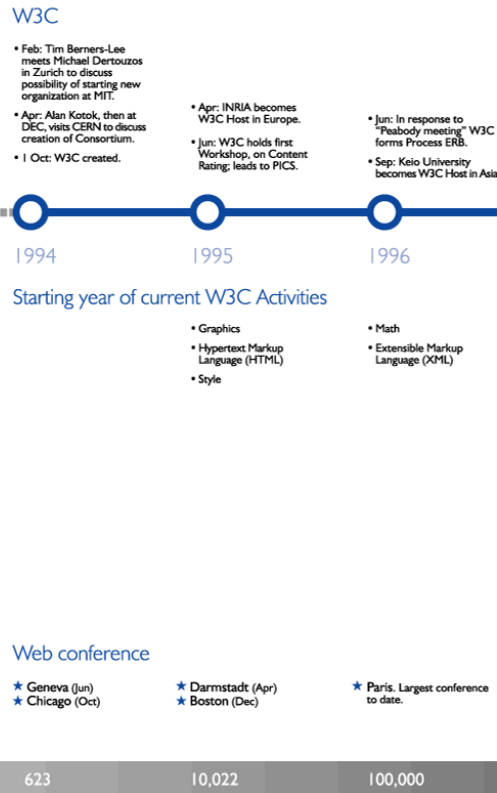
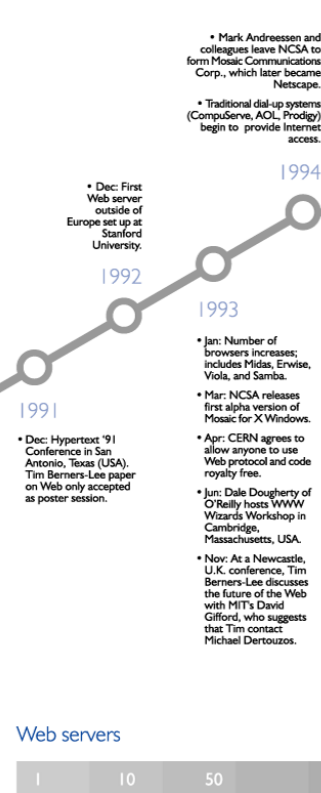
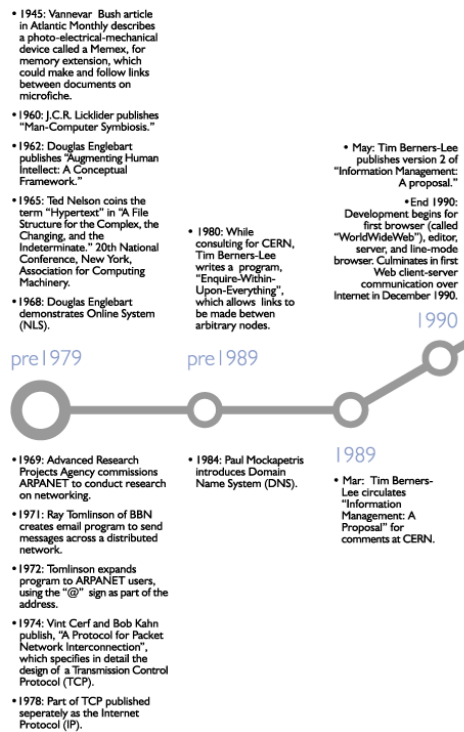
## Sobre a W3C e a apresentação do Phil Archer

- De fato devemos comemorar os 25 anos da Web, muitos resultados.
- O papel importante da W3C nesses 20 anos e sua missão  
*Open Standards* (2012) - <http://open-stand.org/about-us/principles/>
- **Princípios de design que orientam a W3C:**
  - *Web for all (Web Acessibility Initiative, Internationalization and Mobile Web for social development)*
  - *Web on Everything (Web of Devices, Mobile Web Initiative and Browsers and Other Agents)*
- **Visão:**
  - *Web for Rich Interaction (Web Design and Application and Web Architecture)*
  - *Web of Data and Services (Essential XML Technologies, Semantic Web and Web of Services)*
  - *Web of Trust (Semantic Web, XML Security, Web of Services Security and Privacy)*



# W3C 10 WORLD WIDE WEB CONSORTIUM Tenth Anniversary

## Pre-W3C Web and Internet Background



### Web conference

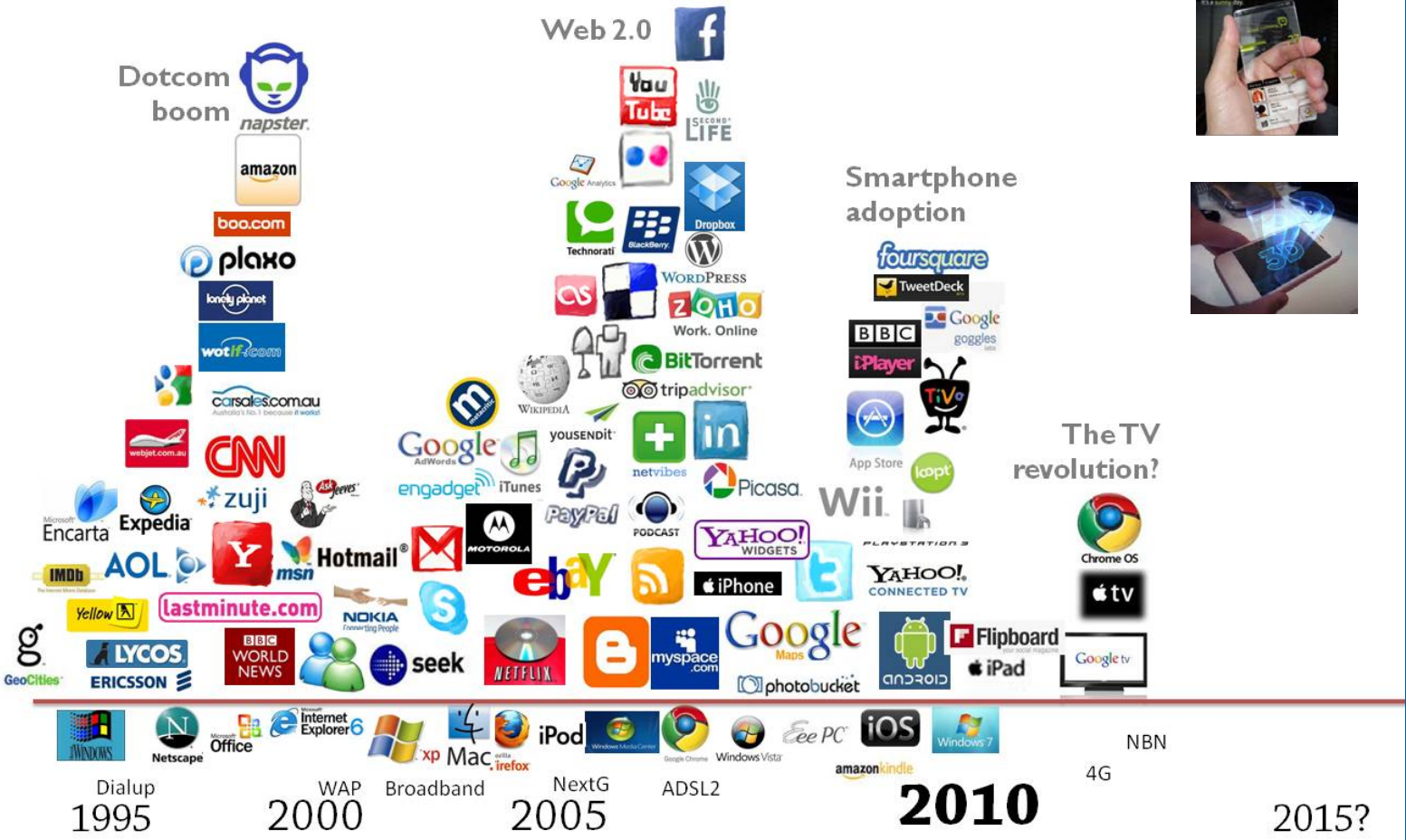
- ★ Geneva (Jun)
- ★ Chicago (Oct)
- ★ Darmstadt (Apr)
- ★ Boston (Dec)
- ★ Paris. Largest conference to date.
- ★ Santa Clara
- ★ Brisbane
- ★ Toronto
- ★ Amsterdam
- ★ Hong Kong
- ★ Honolulu
- ★ Budapest
- ★ New York

### Web servers

1	10	50	623	10,022	100,000	603,367	~1.6 million	~3.7 million	~9.5 million	~26 million	~36 million	~35 million	~46+
---	----	----	-----	--------	---------	---------	--------------	--------------	--------------	-------------	-------------	-------------	------

# We Have Added Or Evolved Roughly One Key Behaviour Per Year To Internet Capability

Link+Research+Browse+Read+Book+Search+Email+Buy+Write+Tag+Share+Watch+Socialise+Map+Aggregate+Locate+Stream



## Top 10 Strategic Technology Trends for 2015

**Merging the Real World and the Virtual World**

- Computing Everywhere
- The Internet of Things
- 3D Printing

**Intelligence Everywhere**

- Advanced, Pervasive and Invisible Analytics
- Context-Rich Systems
- Smart Machines

**The New IT Reality Emerges**

- Cloud/Client Computing
- Software-Defined Applications and Infrastructure
- Web-Scale IT
- Risk-Based Security and Self-Protection

Gartner

## Alguns grupos de trabalho da W3C relacionados a LD

- [Best Practices for Multilingual Linked Open Data](#): envolver crowd-source da comunidade sobre as melhores práticas para a produção de dados abertos ligados multilingues, que estão principalmente concentradas em nomear, rotulagem, articulação e qualidade de dados ligados multilingues, entre outros.



- [Declarative Linked Data Apps](#): produzir uma especificação que descreve como aplicativos *Linked Data Web* podem ser construídos usando tecnologias declarativas, minimizando a necessidade de código-fonte. Modelos de desenvolvimento de software atuais envolvem escrever código fonte (principalmente em linguagens imperativas) e programas de construção a partir dele. O código-fonte é propenso a erros, e administrá-lo requer grande esforço dos desenvolvedores.



- [Decoders W3C Brasil](#): discussões sobre os resultados dos eventos de desenvolvimento realizados pelo W3C Brasil e seus produtos resultantes, bem como incentivar reflexões sobre como melhorar, evoluir e manter os aplicativos feitos em atividades como *hackdays* e *hackatons*.



## Alguns grupos de trabalho da W3C relacionados a LD

- [Development Linked Data](#): discutir as normas e as melhores práticas para expor o desenvolvimento de *Open Data* como Linked Data.



- [ETL Markup Language](#): discutir os requisitos para um padrão aberto para descrever projetos de ETL, incluindo estrutura do projeto, sequenciamento, transformações de fluxo de dados, conectores de fonte de dados, com a finalidade de transportar projetos ETL entre as ferramentas de fonte ETL comerciais e abertas.

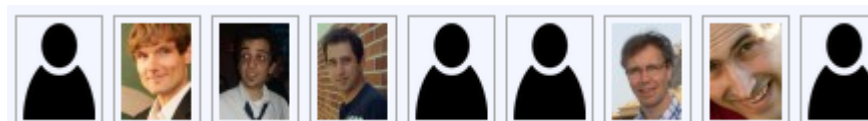


- [JSON for Linking Data](#): (JavaScript Object Notation for Linking Data) é um formato de dados ligados, leve que dá ao seu contexto de dados. É fácil para os seres humanos ler e escrever e fácil para máquinas para analisar e gerar. Ele é baseado no formato JSON já bem sucedido e fornece uma maneira de ajudar a dados JSON interoperar em escala na Web. Para quem já está familiarizado com JSON, escrever JSON-LD é muito fácil, o Estas propriedades fazem JSON-LD uma linguagem de ligação ideal para intercâmbio de dados para ambientes de JavaScript, serviço Web e bancos de dados não estruturados, como CouchDB e MongoDB.



## Alguns grupos de trabalho da W3C relacionados a LOD

- [Linked Building Data](#): Este grupo reúne especialistas na área de modelagem de informações (BIM) e tecnologias da Web de dados para definir os casos de uso e requisitos atuais e futuros de aplicações de *Linked Data* baseado em todo o ciclo de vida dos edifícios (construções).



- [Linked Data Models for Emotion and Sentiment Analysis](#): é um fórum para promover a pesquisa em análise de sentimento. Os tópicos abordados são: Definição de um vocabulário Linked Data com base em emoção e análise de sentimento; Requisitos além da análise baseada em texto, ou seja, a emoção / sentimento análise de imagens, vídeo, análise de redes sociais, etc.; Esclarecer exigências e a necessidade de consenso, como por exemplo, sistemas atualmente usar recursos muito diferentes para descrever valores de polaridade (1-5, -2 / -1 / 0/1/2, positivo neutro negativo, bom, muito bom, etc . -Marl e Onyx são vocabulários de emoção e análise de sentimentos que podem ser tomadas como ponto de partida para a discussão no grupo.



- [Open Data Directory](#): O diretório de dados abertos lista produtos, serviços e projetos de pesquisa que utilizam *Linked Data*. Atualmente, o diretório serve como um agregador de casos de uso e web sites usando *Linked Data* e deverá evoluir ao longo do tempo em resposta as necessidades dos utilizadores.





## Alguns grupos de trabalho da W3C relacionados a LOD

- [Open Government](#): discutir e preparar os dados e especificações API relacionados para abrir as informações do governo, que podem incluir: \* pessoas, como legisladores \* organizações, tais como assembleias ou comitês \* posições das pessoas dentro das organizações \* áreas, tais como distritos eleitorais \* Eventos , tais como eleições \* documentos, tais como notas ou agendas \* discursos, tais como aquelas dadas pelos legisladores em legislaturas, ...



- [Open Linked Education](#): Como uma área crescente e emergente, dados ligados abertos para a educação está experimentando atualmente impulso em várias iniciativas e organizações, incluindo a Educação Aberta, *LinkedUp*, *LinkedUniversities* ou *LinkedEducation*, e a *Open Knowledge Foundation*, são alguns exemplos.



- [Semantic Open Data](#): explorar maneiras de alavancar Dados Abertos como uma contribuição para a Web Semântica e para ajudar a alcançar um entendimento comum de ambas as tecnologias, bem como a sua relação.



# Pesquisa do INESC suportada pelo Open Data Research Network



## **Coordenação Geral**

Nathalie Beghin – Inesc

## **Pesquisadores**

Carmela Zigoni – Inesc

Gisele Craveiro – GPOPAI/USP

Marcelo Tavares – GPOPAI/USP

## **Organização**

Nathalie Beghin – Inesc

Carmela Zigoni – Inesc

# Pesquisa do INESC suportada pelo Open Data Research Network

## 1. Etapa quantitativa: Dados abertos e orçamento público: avaliando a performance dos governos

Fazer um comparativo quantitativo, tomando como base os oito princípios de dados abertos, os quais são interpretados para o contexto específico dos dados orçamentários com base nos códigos de boas práticas, regulamentações e leis aplicáveis, a saber:

- ✓ Code of Good Practices on Fiscal Transparency (FMI, 2007);
- ✓ Constituição da República Federativa do Brasil;
- ✓ Lei Complementar nº 101/2000;
- ✓ Lei Complementar nº 131/2009;
- ✓ Lei nº 4.320/1964;
- ✓ Decreto 7.185/2010;
- ✓ Manual técnico do orçamento 2013;
- ✓ Lei de Acesso à Informação (Lei nº 12.527/2011).

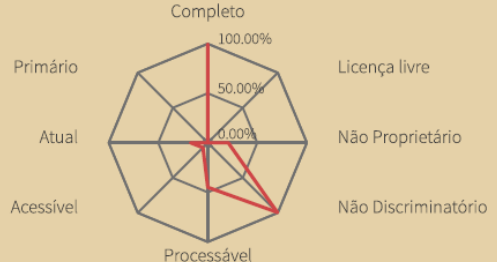
Tabela 1: Arcabouço de Validação	
Princípio	CrITÉrios de Pontuação
1. Completo	Presença de informações sobre receitas e despesas para o ano de 2013 (pontuação 0 ou 1).
2. Primário	Observação das categorias e fases das receitas e despesas (pontuação 0 ou 1; para estas, os 15 campos de informações de receitas e despesas previstos no Decreto 7.185/2010 devem ser publicados).
3. Atual	A data da última atualização de 2013 deve ser de 24 horas. (pontuação 0 ou 1)
4. Acessível	Não apresentar erros do tipo P1, aqueles que impedem o acesso à informação, pontuação 1; conforme os padrões da e-MAG avaliados por meio do programa de validação automática ASES do Governo Eletrônico (pontuação 0 ou 1).
5. Processável por máquina	PDF não pontua, XLS pontua 1, CSV ou XML pontuam 2, se existir CSV e outras alternativas mais sofisticadas, equivalentes a 4 e 5 estrelas do modelo de Tim Berners-Lee, pontua 3.
6. Não discriminatório	Foi verificado se havia necessidade de algum acesso especial ou cadastro (pontuação 0 ou 1).
7. Não Proprietário	PDF e XLS (sem pontuação) e CSV (pontua 1).
8. Livre de licença	Foi verificada a existência de licenças de uso dos dados e se elas previam alguma restrição (pontuação 0 ou 1).

Fonte e Elaboração: Gpopai/Usp, 2014.

# Pesquisa do INESC suportada pelo Open Data Research Network

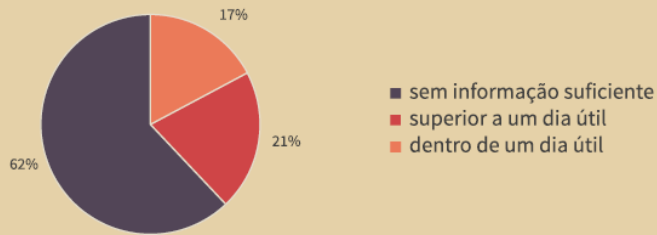
## Resultados

Figura 1: Conformidade dos sites de transparência governamentais pesquisados com Princípios de Dados Governamentais Abertos (DGA)



Fonte e Elaboração: Gpopai/Usp, 2014.

Figura 2: Percentual da avaliação do princípio 'atual'



Fonte e Elaboração: Gpopai/Usp, 2014.

Tabela 8: Ranking dos portais de transparência

Posição	Ente da federação	Pontuação	Completo	Primário	Atual	Acessível	Processável	Não discriminatório	Não proprietário	Licença livre
1ª	Prefeitura do Rio de Janeiro	6	1	-	-	-	3	1	1	-
	Prefeitura de São Luis	6	1	-	-	1	2	1	1	-
	Prefeitura de João Pessoa	6	1	-	-	-	3	1	1	-
2ª	Controladoria Geral da União	5	1	-	-	-	2	1	1	-
	Prefeitura de Teresina	5	1	-	-	-	2	1	1	-
	Senado Federal	5	1	-	-	-	2	1	1	-
3ª	Prefeitura de Palmas	4	1	-	1	-	1	1	-	-
	Prefeitura de Maceió	4	1	-	1	-	1	1	-	-
	Prefeitura de Curitiba	4	1	-	-	1	1	1	-	-
4ª	Prefeitura de Salvador	3	1	-	-	-	1	1	-	-
	Prefeitura de São Paulo	3	1	-	-	-	1	1	-	-
	Prefeitura de Boa Vista	3	1	-	1	-	-	1	-	-
	Prefeitura de Recife	3	1	-	-	-	1	1	-	-
	Prefeitura de Belém	3	1	-	-	-	1	1	-	-
	Prefeitura de Florianópolis	3	1	-	1	-	-	1	-	-
	Prefeitura de Macapá	3	1	-	1	-	-	1	-	-
	Prefeitura de Fortaleza	2	1	-	-	-	-	1	-	-
5ª	Prefeitura de Porto Alegre	2	1	-	-	-	-	1	-	-
	Prefeitura de Cuiabá	2	1	-	-	-	-	1	-	-
	Governo do Distrito Federal	2	1	-	-	-	-	1	-	-
	Prefeitura de Vitória	2	1	-	-	-	-	1	-	-
	Prefeitura de Goiânia	2	1	-	-	-	-	1	-	-
	Prefeitura de Porto Velho	2	1	-	-	-	-	1	-	-
	Prefeitura de Belo Horizonte	2	1	-	-	-	-	1	-	-
	Prefeitura de Natal	2	1	-	-	-	-	1	-	-
	Prefeitura de Aracajú	2	1	-	-	-	-	1	-	-
	Prefeitura de Campo Grande	2	1	-	-	-	-	1	-	-
6ª	Prefeitura de Rio Branco	2	1	-	-	-	-	1	-	-
	Prefeitura de Manaus	1	-	-	-	-	-	1	-	-

Fonte e Elaboração: Gpopai/Usp, 2014.

# Pesquisa do INESC suportada pelo Open Data Research Network

## 2. Etapa qualitativa: Dados abertos e orçamento público: impactos na percepção de intermediários

O objetivo da etapa qualitativa da pesquisa foi avaliar/dimensionar, por meio de entrevistas com *intermediários* de dados abertos (de orçamento, mas não só), o impacto desses dados no alargamento dos direitos humanos e da cidadania no Brasil. Os intermediários de dados abertos são aqueles que “*se apoderam dos dados e os transformam e recombina para, então, rerepresentá-los ao público*”, uma vez que muitos dados não atingirão diretamente os usuários finais<sup>29</sup>, ou seja, organizações ou agentes individuais que de alguma forma processam, analisam e/ou reusam os dados existentes de forma a potencializar o acesso às informações disponíveis à sociedade.

- a) Processos: estudos sobre os pontos que favorecem ou dificultam os processos de implementação de uma política ou programa, aí incluídos seus desenhos e dimensões organizacionais e institucionais.
- b) Resultados, que englobam:
  - i) Desempenho (ou resultados em um sentido estrito, *outcomes*), que se refere aos “produtos” do programa, tais como definidos em suas metas;
  - ii) Impacto: mudanças na situação dos beneficiários, provocadas diretamente pelo programa;
  - iii) Efeitos: outros resultados do programa, sociais ou institucionais, esperados ou não, que acabam se produzindo em decorrência do programa.

# Pesquisa do INESC suportada pelo Open Data Research Network

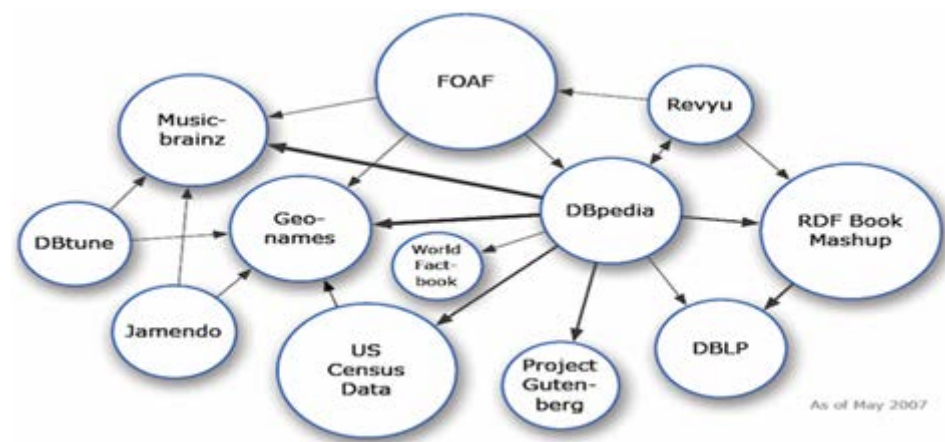
## Resultados

Quadro 4. Percepção dos Usuários sobre os Websites de Dados Abertos que Acessam								
Princípio Portal	Completos	Primários	Atuais	Accessíveis	Processáveis por máquina	Não discriminatório	Não proprietário	Livre de licença
Câmara dos Deputados (atividade legislativa)	Sim	Não	Sim	Sim	Não	Sim	Sim	Sim
Senado Federal (atividade legislativa)	Sim	Não	Sim	Sim	Não	Sim	Sim	Sim
Câmara dos Deputados (dados orçamentários do órgão)	Não	Sim	Sim	Sim	Sim	Sim	Não sabe	Não sabe
Senado Federal (dados orçamentários do órgão)	Não	Sim	Sim	Sim	Sim	Sim	Não sabe	Não sabe
Portal da Transparência	Sim/não sabe	Sim/não sabe	Sim	Sim/não	Sim	Sim	Não sabe	Não explicitado no site
Siga Brasil	Sim	Sim	Sim	Sim	Sim	Sim/não (cadastro para usuário avançado)	Sim	Não sabem.
SIOP	Sim	Sim	Sim	Sim	Sim	Sim	Não sabe	Não sabe
Sistema de Votação da Plenária da Câmara	Sim	Sim	Sim	Sim	Sim	Não sabe	Não sabe	Não sabe
Receita Federal	Sim	Sim	Sim	Sim	Sim	Não (necessita CNPJ e captcha)	Não sabe	Não sabe
SISTN (Caixa Econômica Federal)	Não	Sim	Não	Sim	Não	Sim	Sim	Não sabe

Fonte/Elaboração: Inesc, 2014.

# Linked Open Data Cloud

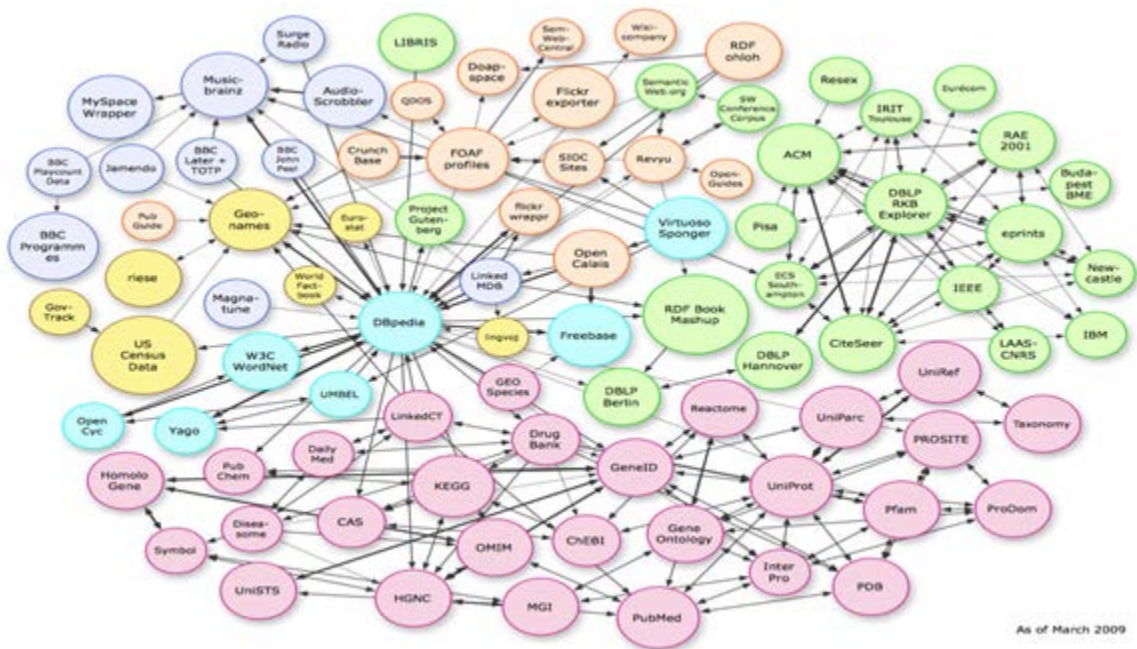
2007



As of May 2007

2009

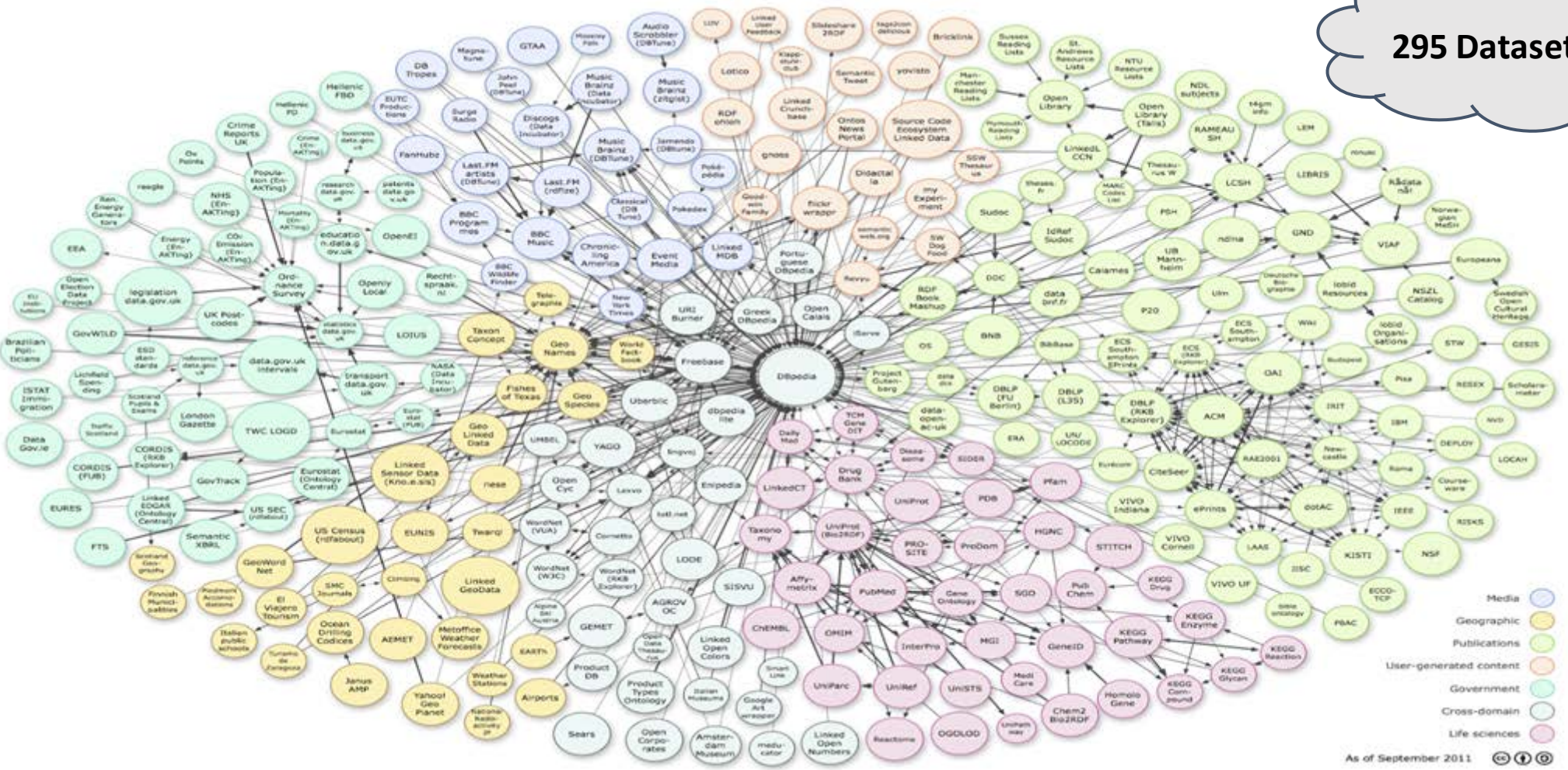
93 Datasets



As of March 2009

# Linked Open Data Cloud 2011

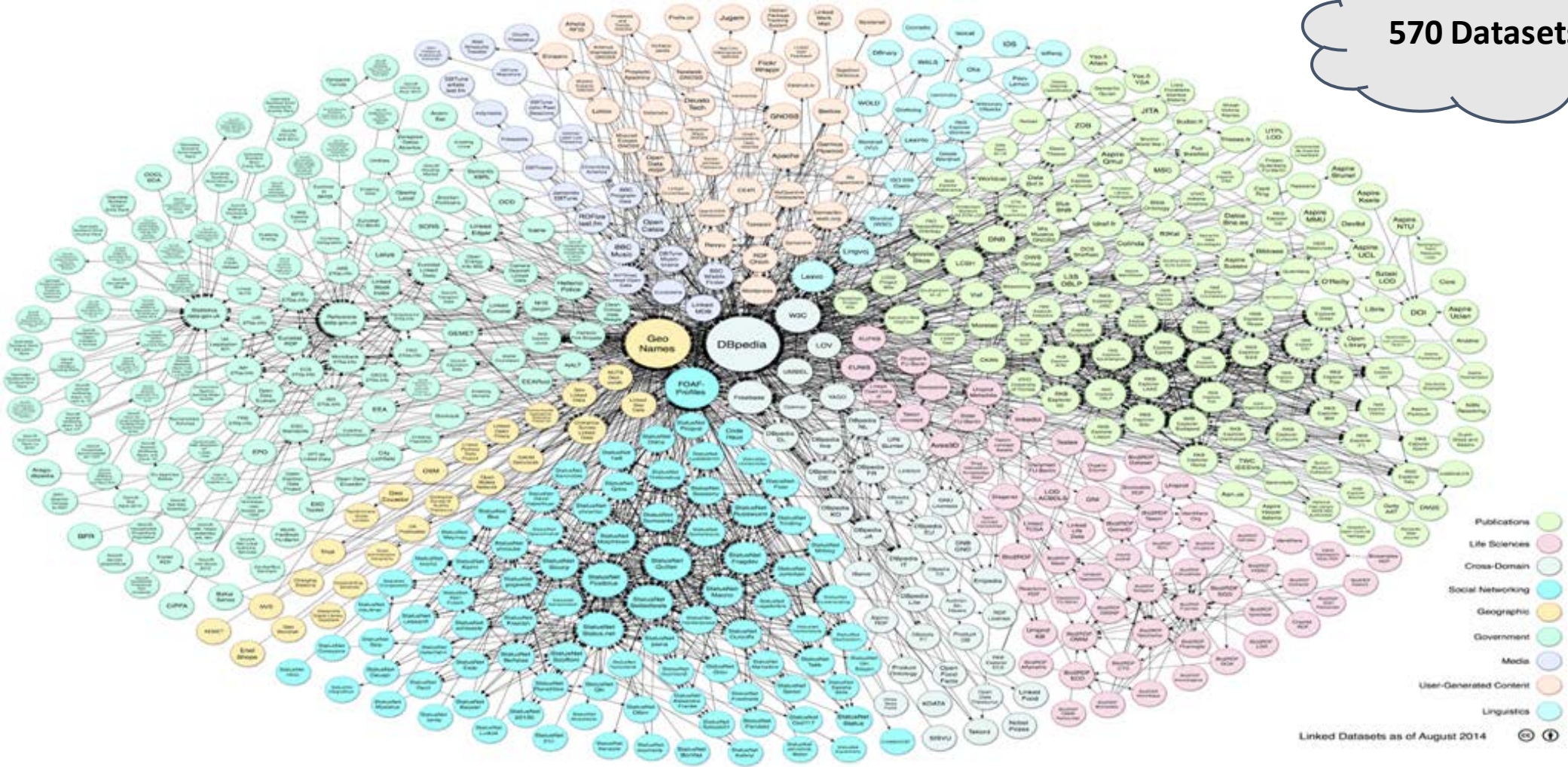
295 Datasets





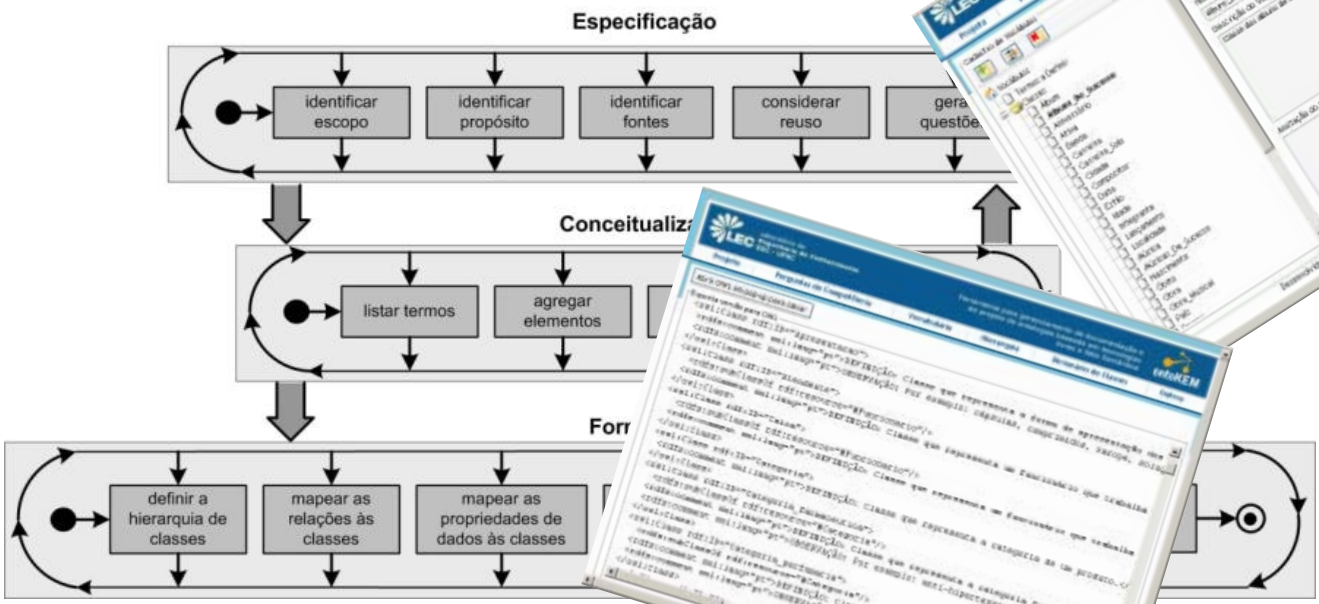
# Linked Open Data Cloud 2014

570 Datasets



# Iniciativas do LEC - ontoKEM

Ferramenta para EC que suporta um processo de construção e documentação de ontologias desenvolvida em 2007.



# Iniciativas do LEC - IodKEM

Projeto iniciado em 2011 que visa contribuir para o uso de LOD com alta qualidade promovendo seu uso e publicação.



**lodKEM**  
linked open data & knowledge engineering and management

HOME | QUEM SOMOS | DOWNLOADS | PROJETOS | PARCEIROS | NOTÍCIAS | LINKS | LOGIN

**DADOS**  
Conheça os dados disponíveis compilados pelo IodKEM e por parceiros

**APLICAÇÕES**  
Obtenha informações sobre as aplicações de Open Linked Data

**DESENVOLVEDORES**  
Acesse materiais exclusivos e direcionados a desenvolvedores

**ONTOLOGIAS**  
Acesse as ontologias disponíveis

**APRENDENDO**  
Tenha acesso a materiais, tutoriais e artigos para facilitar seu aprendizado

**PROJETOS DE P&D**  
Projetos de Pesquisa e Desenvolvimento

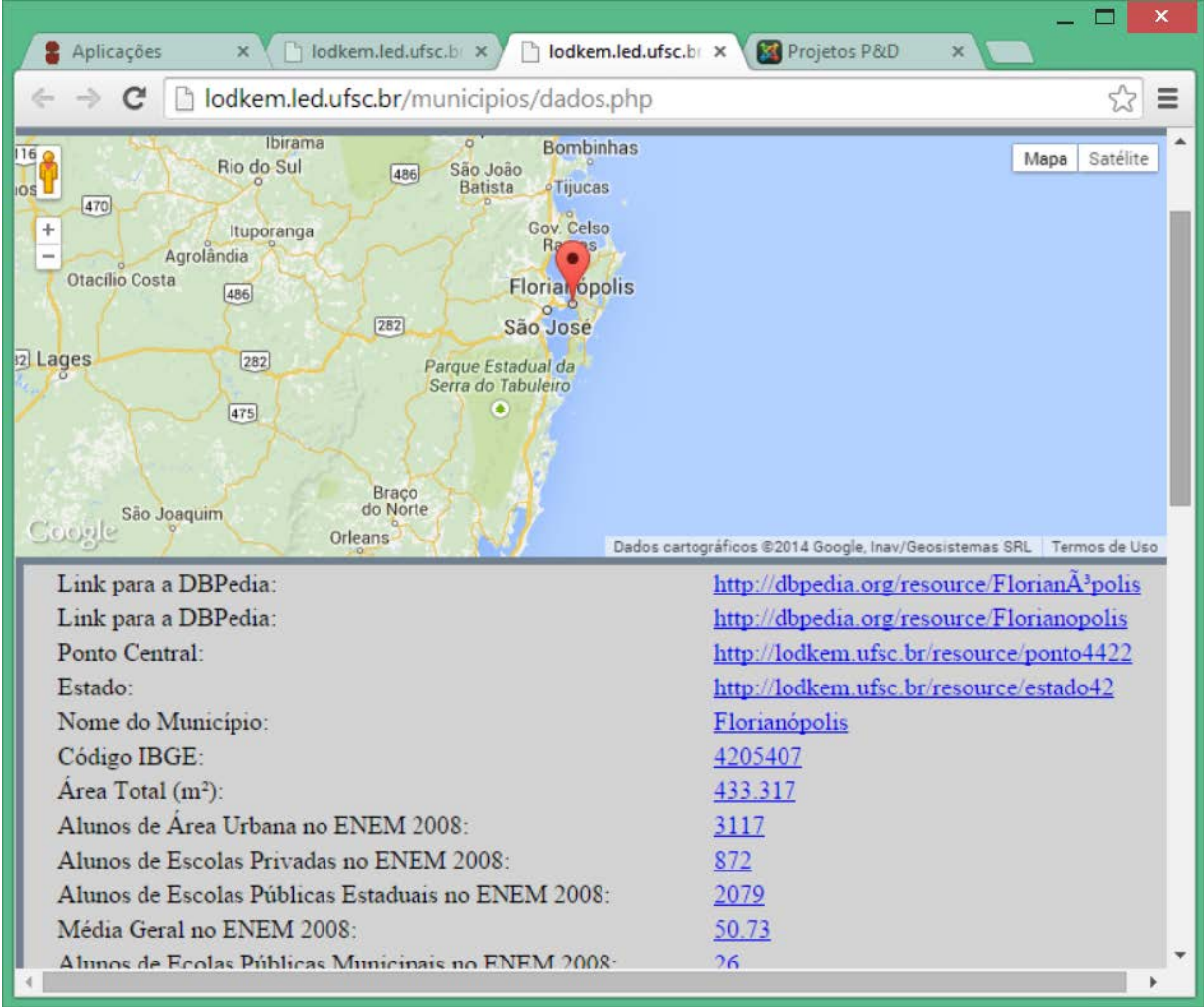
**Últimas Notícias**

- [Joined Up Data for Joined Up Problem Solving](#)
- [Swirri awarded new Scottish Government linked open data project](#)

Há alguns anos ocorre o debate sobre dados governamentais abertos, sobre informação pública livre para qualquer um acessar e reutilizar com qualquer finalidade (BERNERS-LEE, 2009). Em 2009, o tema começou a aparecer na grande mídia, com governos de vários países (como Estados Unidos, Reino Unido, Canadá e Nova Zelândia) anunciando iniciativas voltadas a abrir sua informação pública. Entretanto, esses dados, em sua grande maioria, têm sido publicados de maneira não estruturada, em diferentes padrões e sem qualquer conexão semântica entre si. Isso cria uma grande dificuldade para que pesquisadores e outros interessados possam manipular e extrair novos conhecimentos a partir deles. Um exemplo é a

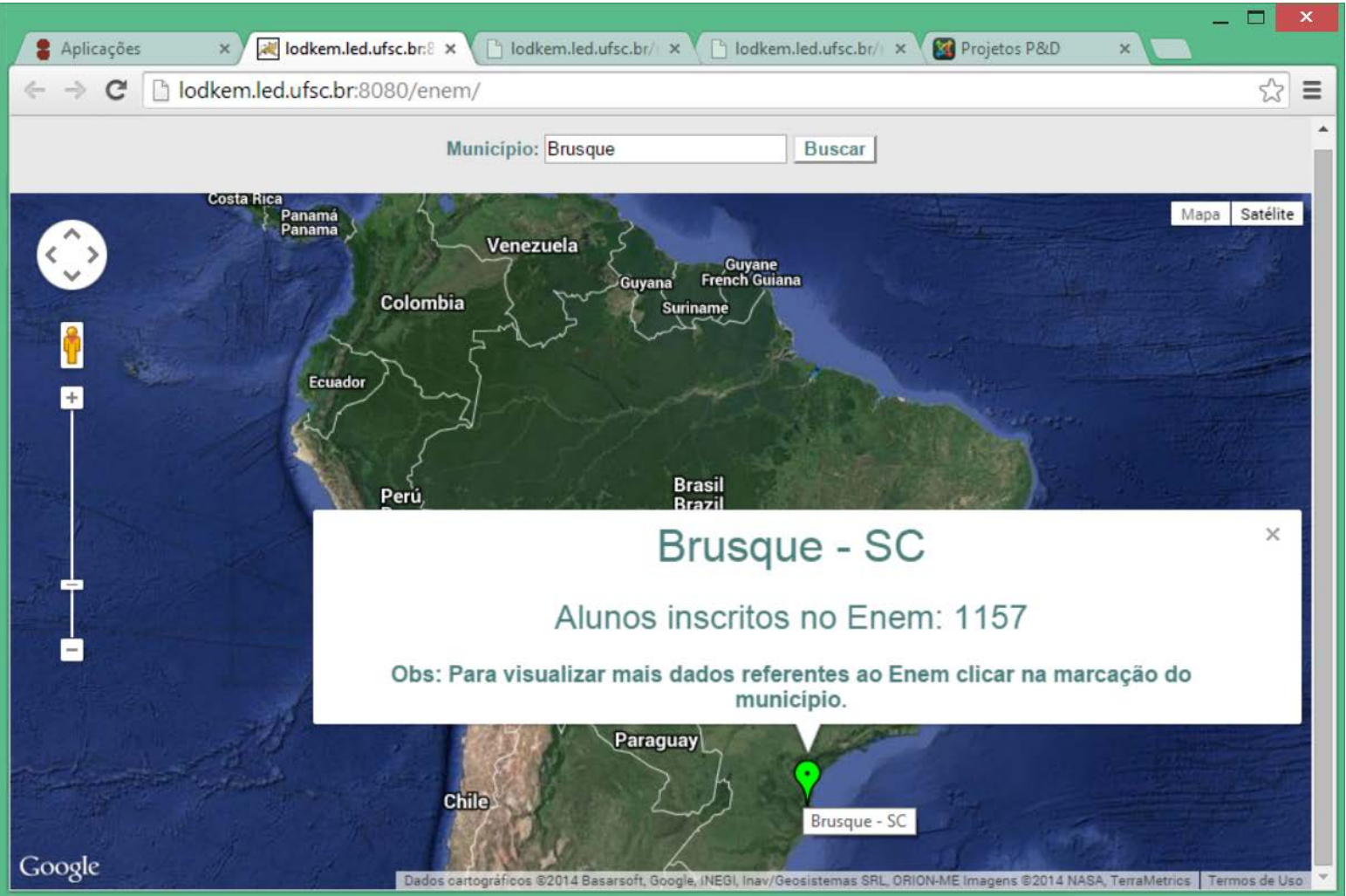
# lodKEM – GeoPoliticaBR

Data set de municípios brasileiros ligados a DBPedia.



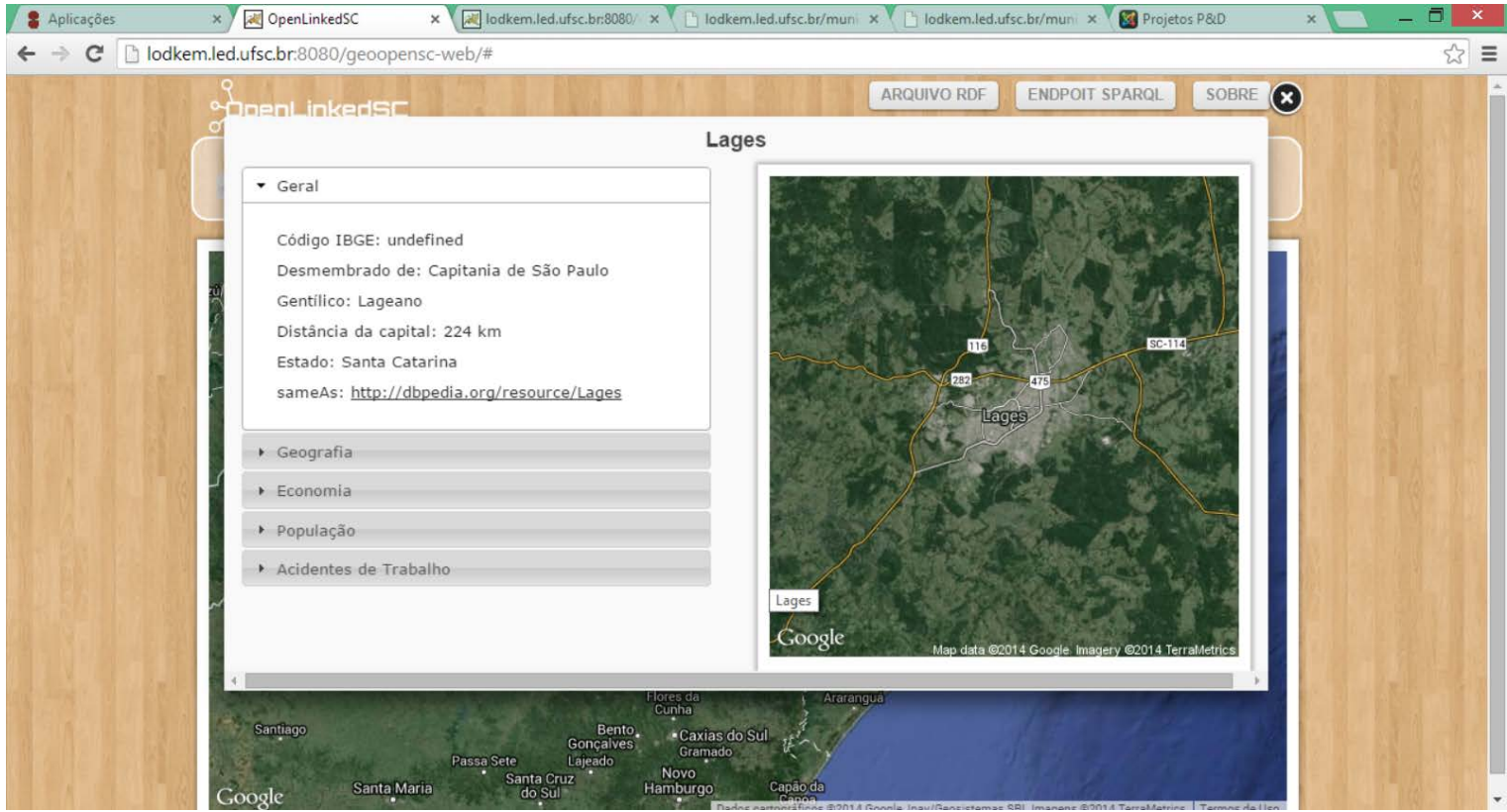
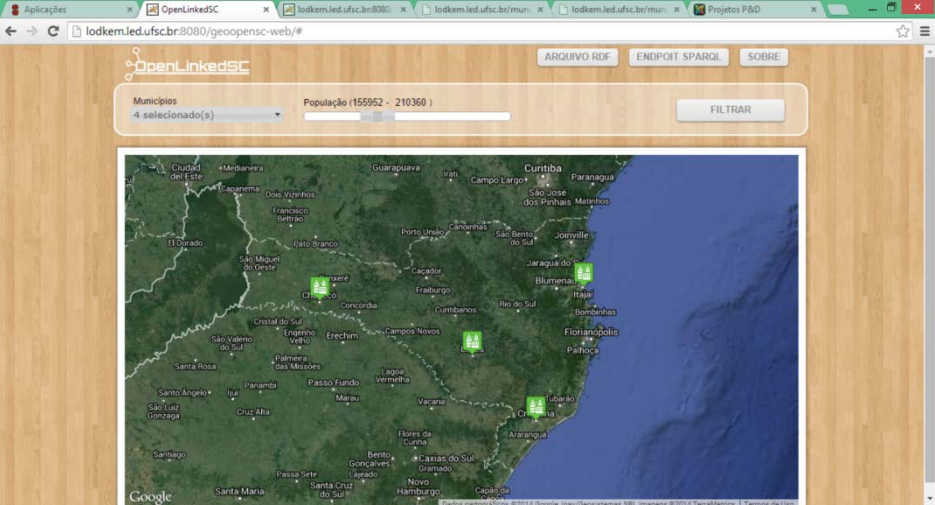
# IodKEM – ENEM

Data set do resultado do ENEM de 2008 ligado a GeoPoliticaBR.



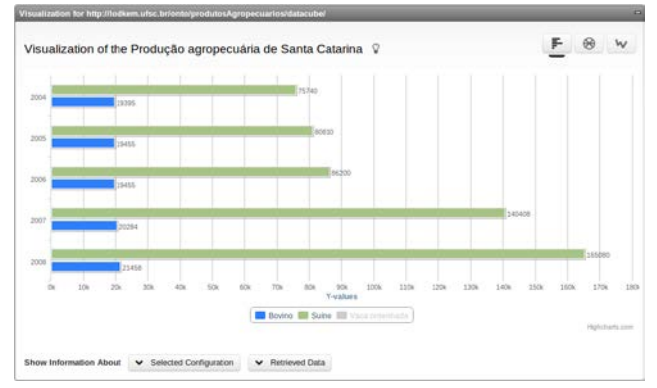
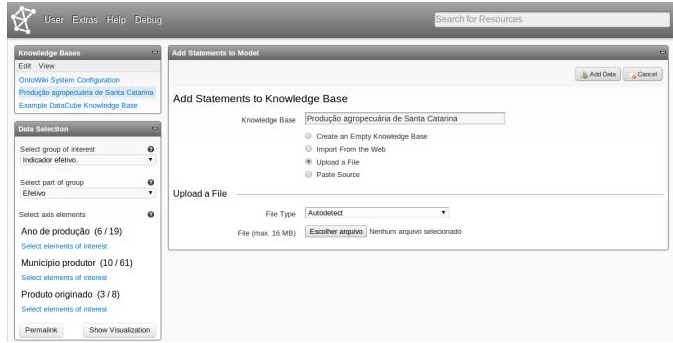
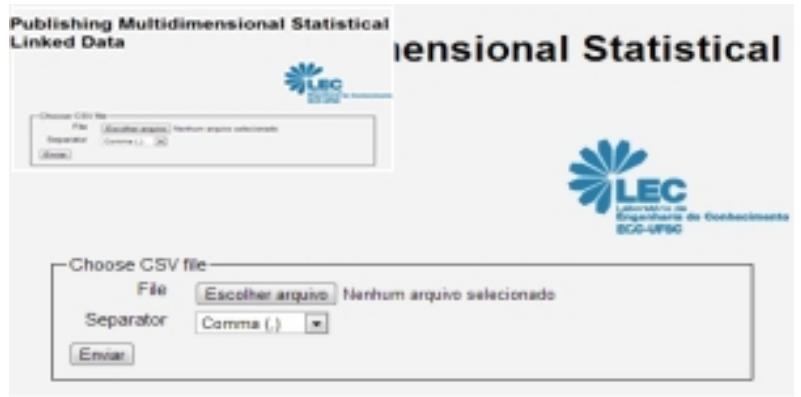
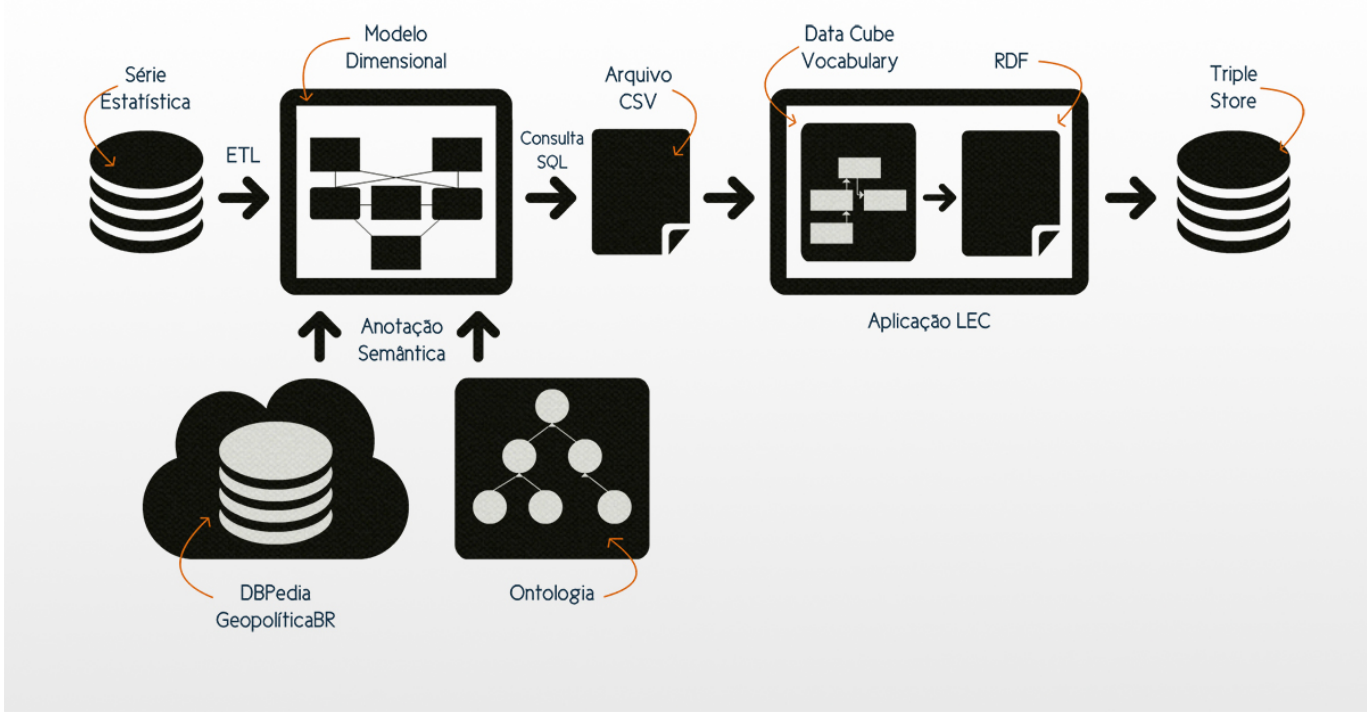
# IodKEM – OpenLinked SC

Data set com dados abertos de SC



# IodKEM – RDF Data Cube Vocabulary

Data set de dados de produção agrícola de SC – Série 1972 - 21012





# LOD BRASIL

## LINKED OPEN DATA

19a21 novembro 2014

### Realização



UNIVERSIDADE FEDERAL  
DE SANTA CATARINA



INSTITUTO FEDERAL  
SANTA CATARINA

### Apoio

