



# Enfrentando os Desafios da Interoperabilidade Semântica de Dados

## Governamentais

Prof. Dr. João Paulo A. Almeida

([jp Almeida@inf.ufes.br](mailto:jp Almeida@inf.ufes.br))

<http://nemo.inf.ufes.br>

Departamento de Informática

Universidade Federal do Espírito Santo

# Agenda



- Visão: **Onde queremos chegar... ?**
    - Aprofundamento da “sociedade da informação”
    - **Web**
  - Diagnóstico: **Onde estamos... ?**
    - Dados governamentais
  - Desafios: O que precisamos fazer...?
-

# Visão: Onde queremos chegar...?



- Ter **acesso** a dados produzidos por uma enormidade de fontes (inclusive governamentais, mas também da sociedade civil, dos cidadão)
  - Ligar ou **combinar** dados produzidos das várias fontes
  - Produzir ou **publicar** dados
    - sobre absolutamente qualquer coisa que seja do interesse
  - Princípio básico:
  - Uso dos dados **muito além** do propósito original para sua coleção
  - **combinações de potencial “infinito”**
-

# Visão: Onde queremos chegar...?



- No caso de dados governamentais fomentar:
  - **Transparência** ativa
  - **Pesquisa** acadêmica e **jornalismo de dados**
  - **Formulação de políticas públicas** e tomada de decisão com base em evidências
  - **Participação** de cidadãos e do setor privado



## Visão: ... **queremos saber!**



- Onde foi parar o meu dinheiro?
- Qual o melhor investimento para o recurso público?
- Os professores da educação básica em diferentes unidades da federação tem salários diferentes?
  - Isto leva a desempenho diferente em testes educacionais?
- Qual a relação entre criminalidade e iluminação pública?
- Qual a relação entre saúde e saneamento básico?
  - Por doença?
  - Por região geográfica?
  - Por IDH?
- Etc., etc., etc.

# Diagnóstico: Onde estamos?



- Uma **infinidade de “ilhas”** de sistemas de informação
- Só no caso governamental:
  - Diferentes esferas (federal, estadual, municipal)
  - Diferentes poderes (executivo, legislativo, judiciário)
  - Diferentes órgãos com complexa estrutura organizacional
  - Cada qual mantendo informações diferentes, porém complementares
    - Sobre saúde, segurança pública, economia, educação, mobilidade urbana, **etc. etc. e etc.**
  - Cada qual com seus “formatos” de dados
    - **Não padronizados**
    - **Pouco ou mal documentados**
    - **Desarticulados**

# Diagnóstico: **...no meio do caminho!**



- Lei de Acesso a Informação de 2011

“Os sítios [...] deverão [...] atender, entre outros, aos seguintes requisitos:

I - conter ferramenta de pesquisa de conteúdo [...]

II - possibilitar a gravação de **relatórios em diversos formatos eletrônicos, inclusive abertos e não proprietários**, tais como planilhas e texto, de modo a facilitar a análise das informações;

III - possibilitar o **acesso automatizado por sistemas externos em formatos abertos, estruturados e legíveis por máquina**;

IV - **divulgar em detalhes os formatos utilizados para estruturação da informação;**“



# Diagnóstico: ...no meio do caminho!



- **Não há um “formato de dados abertos”**
- Há na verdade **“meta-formatos”**
- CSV (Comma Separated Values)
- XML (Extensible Markup Language)
- JSON (JavaScript Object Notation)
- RDF (Resource Description Framework)
- OWL (Web Ontology Language)





# Diagnóstico: **...no meio do caminho!**



- Integração é desafiadora dentro de uma organização
- E agora estamos cruzando fronteiras organizacionais, culturais, políticas, ...
  
- Esforço heróico de “hackers” especialistas
- Não “escalável”
- Pouca precisão no uso dos dados
  - Potencialmente chegando a conclusões erradas!
- Enorme potencial pouco explorado



# Para onde foi o meu dinheiro ? BETA

Base: **Governo federal**

Período: **2010**

[Tweet](#) 328
 [Like](#) 272

## ENCARGOS ESPECIAIS, TRANSFERÊNCIAS, REPASSE

[Voltar para página anterior](#)



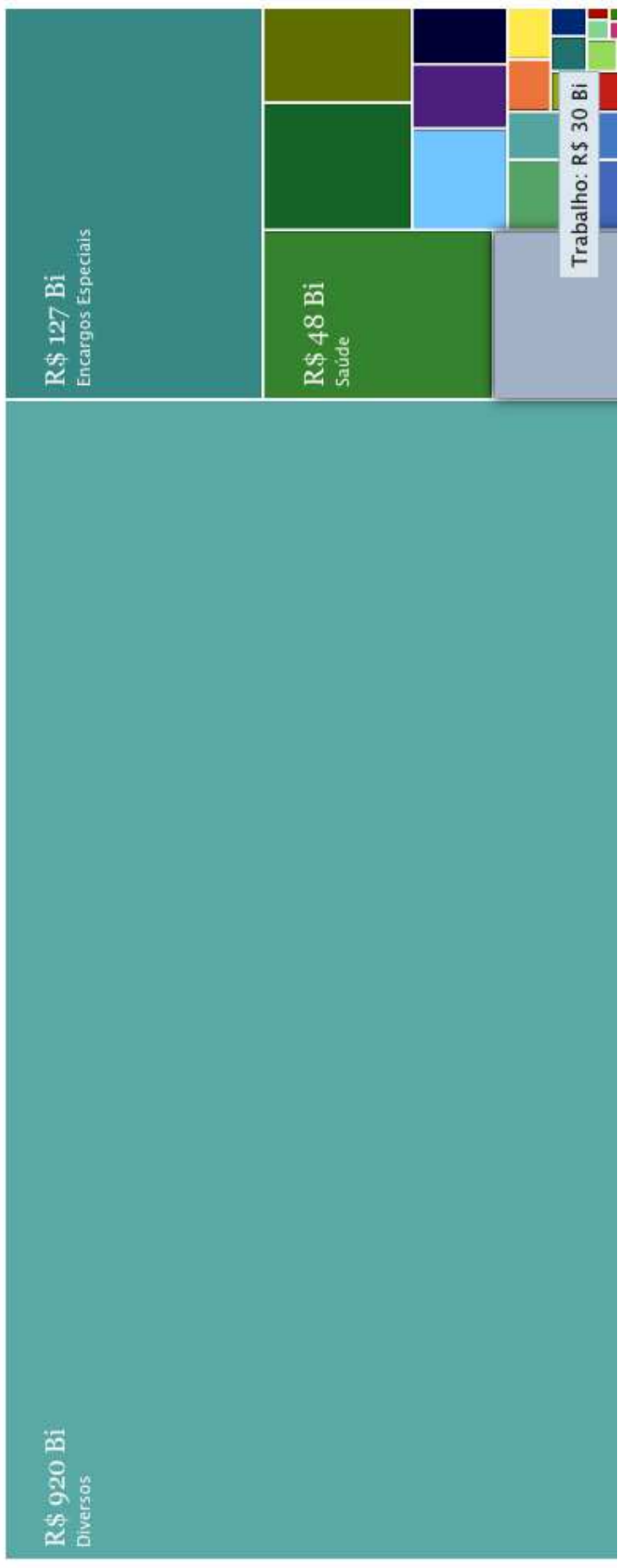
# Para onde foi o meu dinheiro ? BETA

Base: **Governo federal** | Período: **2010**

Tweet 328 | Like 272

ESTA DISPOSIÇÃO É BASEADA NA LEI DE DIRETRIZES ORÇAMENTARIAS E OS DADOS SÃO DO PORTAL DA TRANSPARÊNCIA, NO QUAL ESTÃO APENAS OS INVESTIMENTOS REALIZADOS PELO GOVERNO FEDERAL.

[Voltar para página anterior](#)



http://www.transparencia.es.gov.br



Governo do Estado do Espírito Santo  
**PORTAL DA TRANSPARÊNCIA**  
Secretaria de Estado de Controle e Transparência

O que é o Portal? | Glossário | Legislação | Manual de Navegação | Dados Abertos | Mapa do Site | Acessibilidade | A | A- | A+ |

Domingos Martins  
Romero Mendonça/SECOM

RECEITAS | DESPESAS | PESSOAL | COMPRAS | CONVÊNIO E TRANSFERÊNCIAS | ORÇAMENTO | INFORMAÇÕES GERAIS

Busque aqui...

**SECONT**  
Saiba mais sobre o órgão responsável pela Transparência do Estado.

**OUVIDORIA**  
O Estado quer te ouvir. Deixe aqui suas críticas e sugestões.

**PERGUNTAS FREQUENTES**  
Esclareça suas dúvidas sobre o Portal.

**Endereço**  
Av. Governador Bley, 236 - 8º andar  
Centro, Vitória - ES, CEP 29010-150

**Contato**  
transparencia@secont.es.gov.br  
0800 022 11 17

**WhatsApp**  
 (27)98849-4584

**Redes Sociais**  
 EStransparencia

2012. Conteúdo e dados do Portal da Transparência licenciados sob uma Licença Creative Commons

http://www.transparencia.es.gov.br



 Governo do Estado do Espírito Santo  
**PORTAL DA TRANSPARÊNCIA**  
Secretaria de Estado de Controle e Transparência

[O que é o Portal?](#) | [Glossário](#) | [Legislação](#) | [Manual de Navegação](#) | [Dados Abertos](#) | [Mapa do Site](#) | [Acessibilidade](#)

[RECEITAS](#) | [DESPESAS](#) | [PESSOAL](#) | [COMPRAS](#) | [CONVÊNIOS E TRANSFERÊNCIAS](#) | [ORÇAMENTO](#)

## Despesas

### Despesas por Órgão

Veja aqui quanto e com o quê estão sendo aplicados os recursos do Estado.

De  a  Seleccione um ano

(Formato: De dd/mm a dd/mm)

 Resultado em Gráfico

 Resultado em Tabela

 Download da base de dados

Aqui você pode acompanhar, de forma anual, os valores arrecadados pelo Estado, anteriores ao ano de 2009.

Seleccione um ano   Resultado em Tabela



RECEITAS

DESPESAS

PESSOAL

COMPRAS

CONVÊNIO E TRANSFERÊNCIAS

ORÇAMENTO

INFORMAÇÕES GERAIS

## Despesas

Despesas por Órgão  **Órgão**

Dados atualizados em: **21/05/2015**

Período da consulta: **01/01/2015 a 31/12/2015**

Total Empenhado pelo Governo:	<b>R\$ 9.798.899.355,41</b>
Total Liquidado pelo Governo:	<b>R\$ 3.980.625.239,10</b>
Total Pago pelo Governo:	<b>R\$ 3.787.116.518,80</b>

Código	Órgão	Valor Empenhado	Valor Liquidado	Valor Pago
60101	DEFENSORIA PÚBLICA DO ESTADO DO ESPÍRITO SANTO	15.742.392,33	15.500.407,82	15.497.778,93
60901	FUNDO DE APARELHAMENTO DA DEFENSORIA PÚBLICA	10.887.642,80	4.522.456,11	4.494.074,10
100101	SECRETARIA DA CASA CIVIL	2.312.469,26	1.988.048,77	1.962.845,53

# empenhado2015.csv



NumAnoExercicio;NumCodigoOrgao;NumCodigoOrgao;NumCodigoUnidadeGestora;NumCodigoGestaoEmitente;numDocumentoEmpenho;NumCodigoFuncao;NumCodigoSubFuncao;NumCodigoPrograma;NumCodigoAcao;NumCpfCnpjNis;StrNomeFavorecido;numTipoFavorecido;NumCodigoElementoDespesa;NumCodigoSubelementoDespesa;NumProcesso;NumTipoLicitacao;numValorEmpenho;datEmpenhoDocumento;NumCodigoGrupoDespesa;NumCodigoModalidade;historicoDocumentoEmpenho;NumCategoriaEconomica;NumCodigoSubtitulo;NumCodigoFonte;

2015;45;45;450101;1;;;;;;;;;;14;;;;;R\$ 163,80;11/05/2015;3;;;3;;;

2015;10;10;100102;1;;;;;;;;;;39;;;;;R\$ 6.196,00;03/02/2015;3;;;3;;;

2015;10;10;100102;1;;;;;;;;;;15;;;;;R\$ 56,00;02/02/2015;3;;;3;;;

2015;45;45;450101;1;;;;;;;;;;15;;;;;R\$ 56,00;04/02/2015;3;;;3;;;

## tborgao2015.CSV

numCodigoOrgao; numAnoExercicio; strNomeOrgao; numTipoPoder

10; 2014; SECRETARIA DA CASA CIVIL ; 4

10; 2015; DESCRIÇÃO PENDENTE DE DEFINIÇÃO; 4

45; 2014; SECRETARIA DE ESTADO DA SEGURANCA PÚBLICA E DEFESA SOCIAL ; 4

45; 2015; DESCRIÇÃO PENDENTE DE DEFINIÇÃO; 4



Controladoria-Geral da União  
**Portal da Transparência**  
GOVERNO FEDERAL

[Perguntas frequentes](#) | [Contato](#) | [Glossário](#) | [Links](#) | [Manual de navegação](#)

Acesso rápido

Você está em:  
Início» Download de Dados» **Servidores**

### Download de Dados

### Servidores Civis e Militares do Executivo Federal

Nesta seção estão disponíveis informações, em formato aberto, sobre cargo, função, situação funcional e remuneração dos servidores civis e militares, bem como dos agentes públicos do Poder Executivo Federal.

**Os arquivos abaixo apresentam:**

1) Servidores: NOME; CPF; MATRICULA; DESCRICAO\_CARGO; CLASSE\_CARGO; REFERENCIA\_CARGO; PADRAO\_CARGO; NIVEL\_CARGO; SIGLA\_FUNCAO; NIVEL\_FUNCAO; FUNCAO; CODIGO\_ATIVIDADE; ATIVIDADE; OPCA\_O\_FUNCAO\_TOTAL; UORG\_LOTACAO; COD\_ORG\_LOTACAO; ORG\_LOTACAO; COD\_ORGSUP\_LOTACAO; ORGSUP\_LOTACAO; UORG\_EXERCICIO; COD\_ORG\_EXERCICIO; ORG\_EXERCICIO ; COD\_ORGSUP\_EXERCICIO; ORGSUP\_EXERCICIO; TIPO\_VINCULO; SITUACAO\_VINCULO; COD\_GRUPO\_AFASTAMENTO; COD\_AFASTAMENTO; DATA\_INICIO\_AFASTAMENTO; DATA\_TERMINO\_AFASTAMENTO; REGIME\_JURIDICO; JORNADA\_DE\_TRABALHO; DATA\_INGRESSO\_CARGOFUNCAO; DATA\_NOMEACAO\_CARGOFUNCAO; DATA\_INGRESSO\_ORGAO; DOCUMENTO\_INGRESSO\_SERVICOPUBLICO; DATA\_DIPLOMA\_INGRESSO\_SERVICOPUBLICO; DIPLOMA\_INGRESSO\_CARGOFUNCAO; DIPLOMA\_INGRESSO\_ORGAO; DIPLOMA\_INGRESSO\_SERVICOPUBLICO.

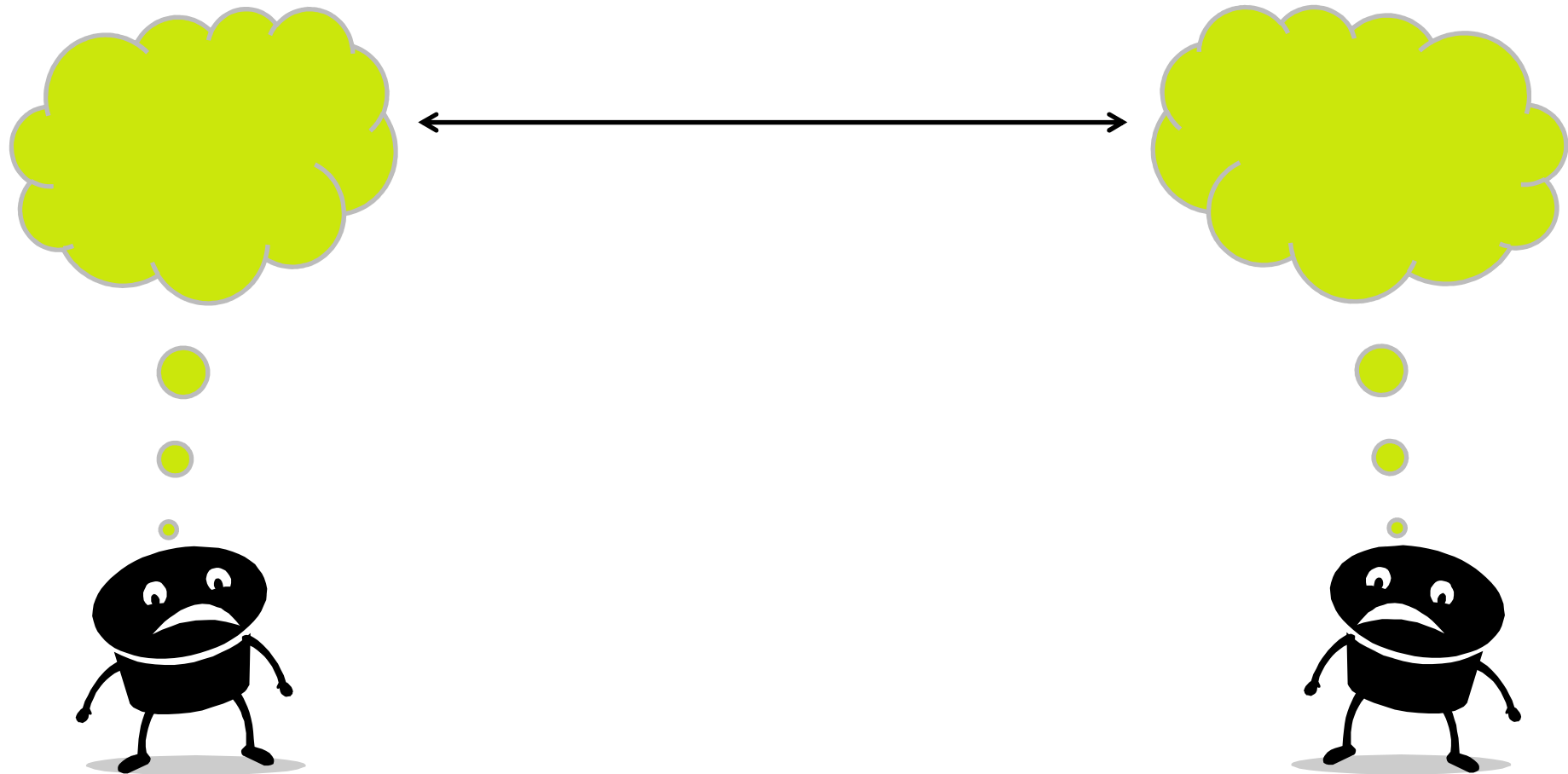




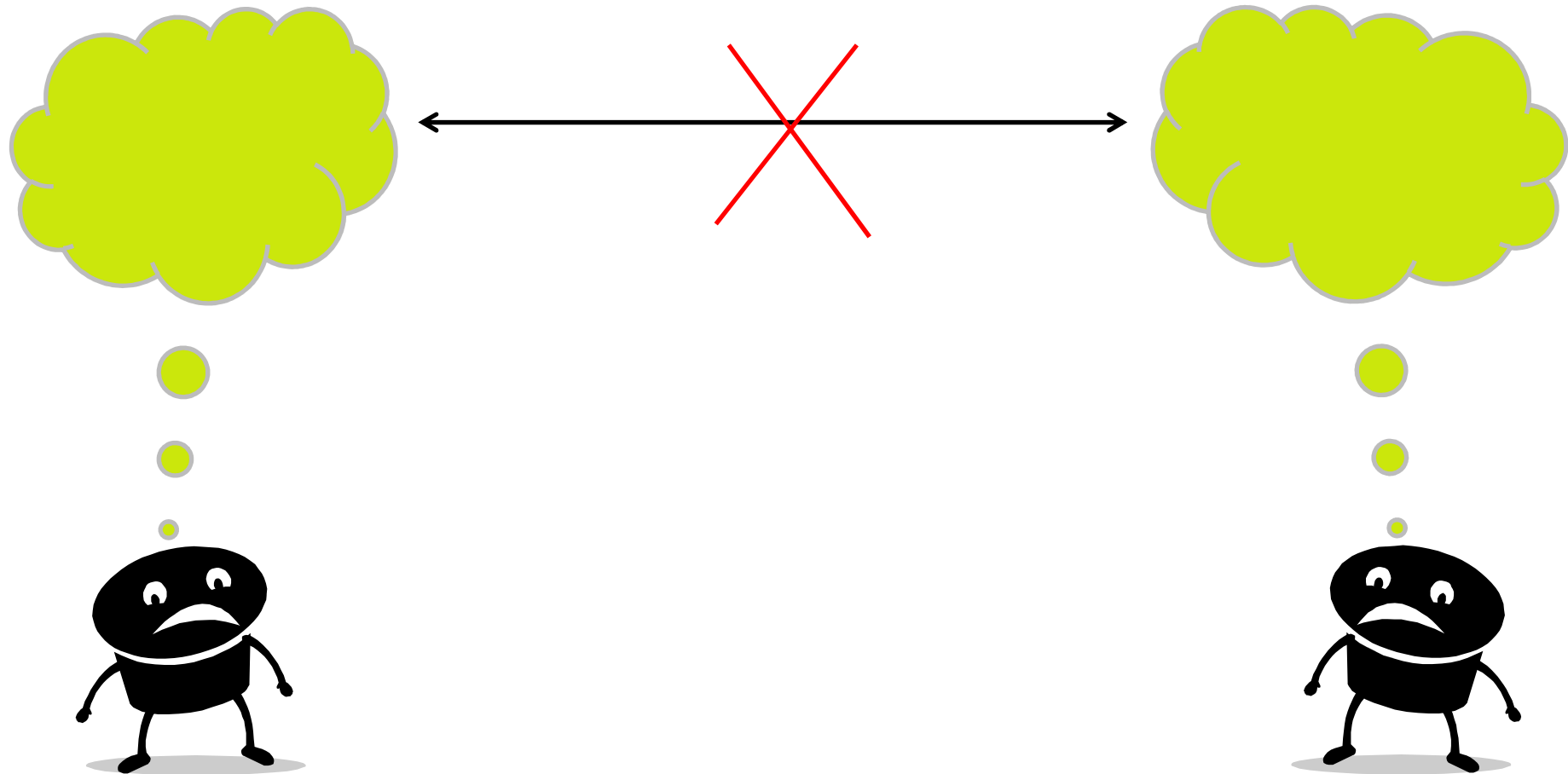
Para onde vamos???



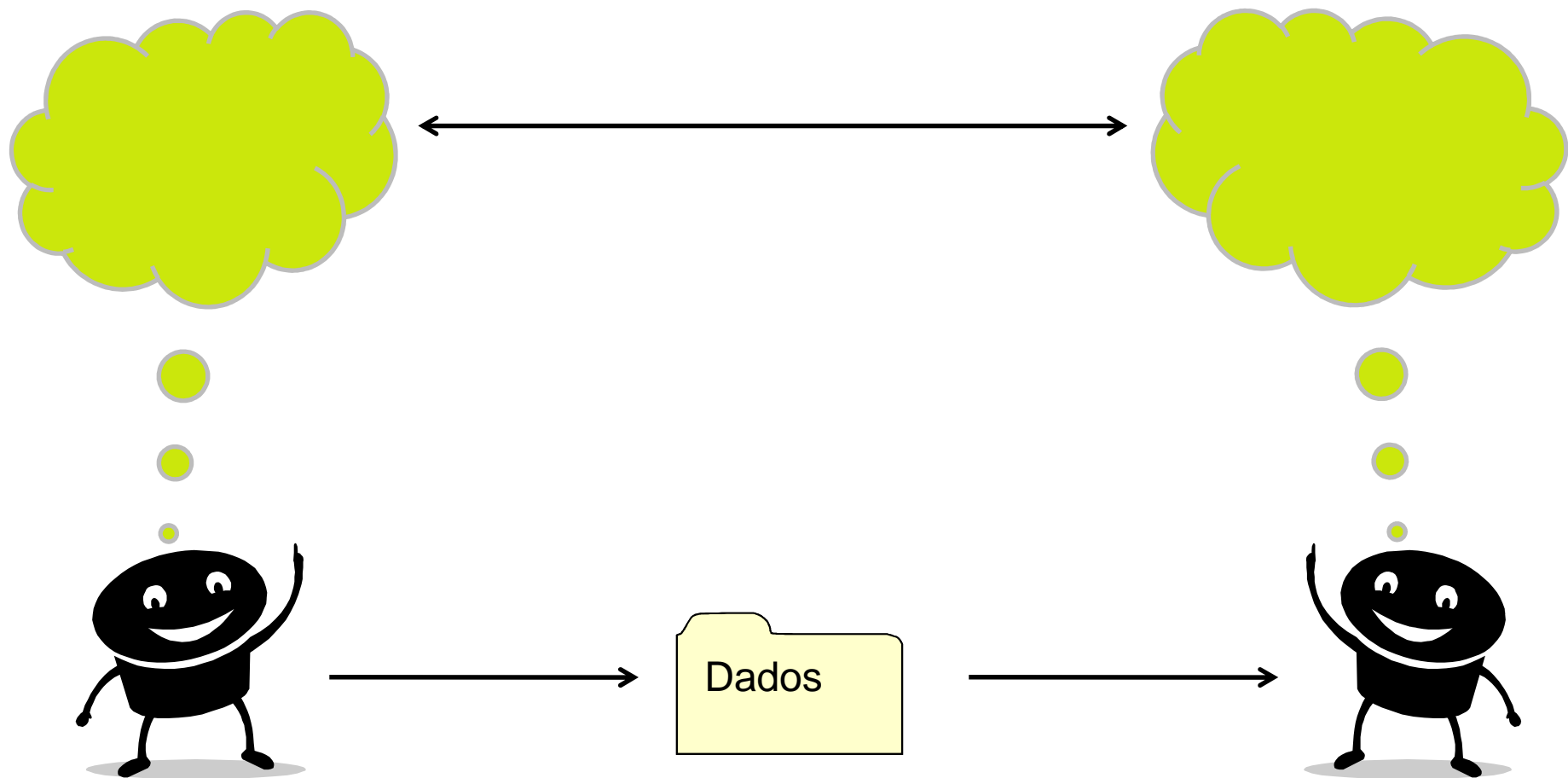
**Na essência um problema de comunicação**



Premissa essencial: Não há telepatia...

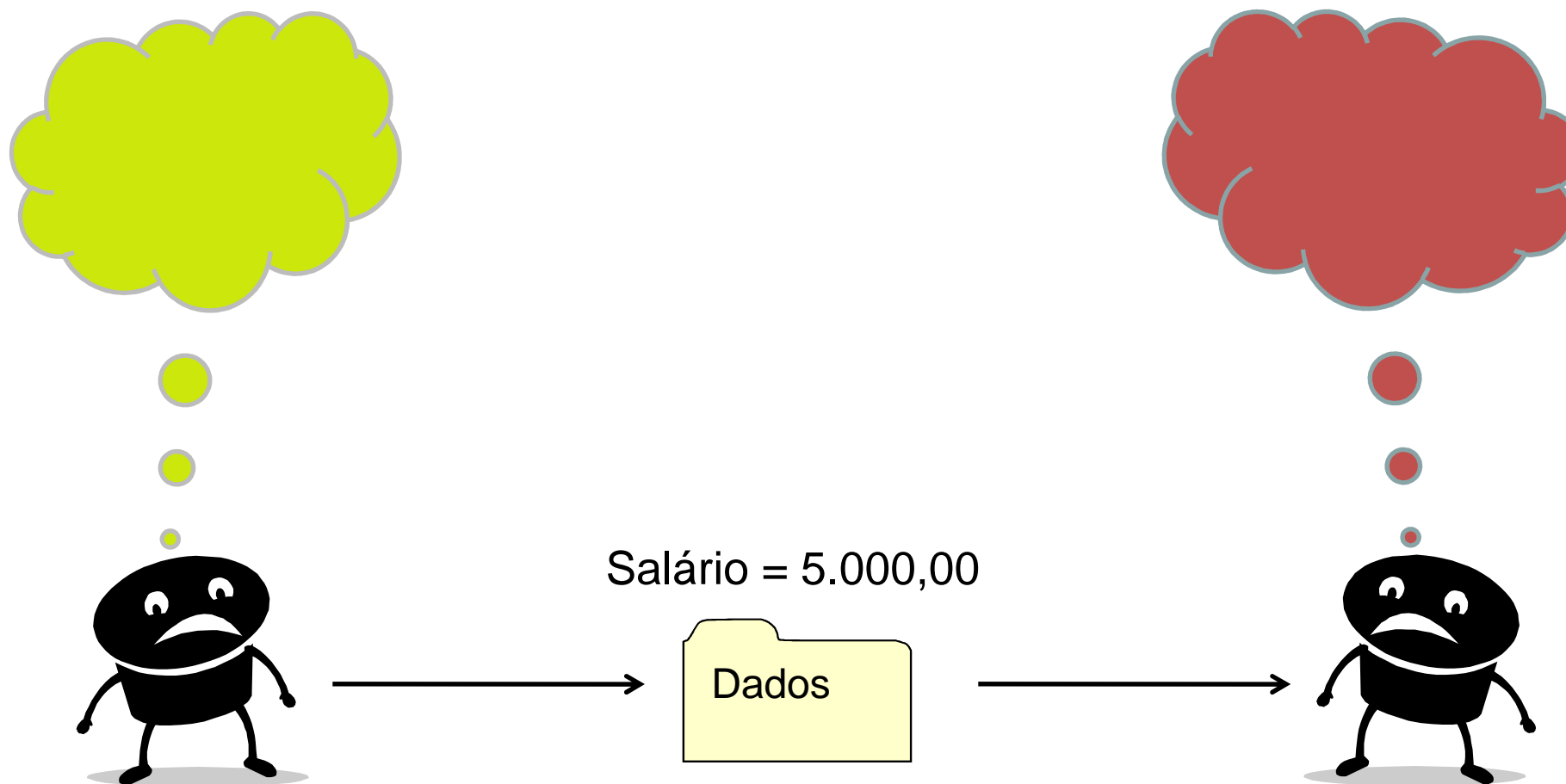


# Dados na web: comunicação entre pessoas



# Desafio

Dados podem ser interpretados de diferentes formas



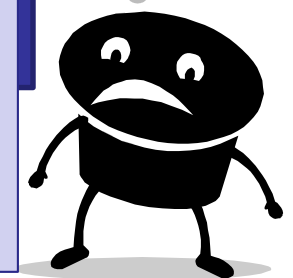
# O Problema: Interoperabilidade Semântica



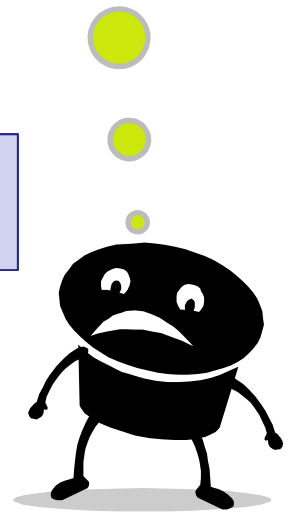
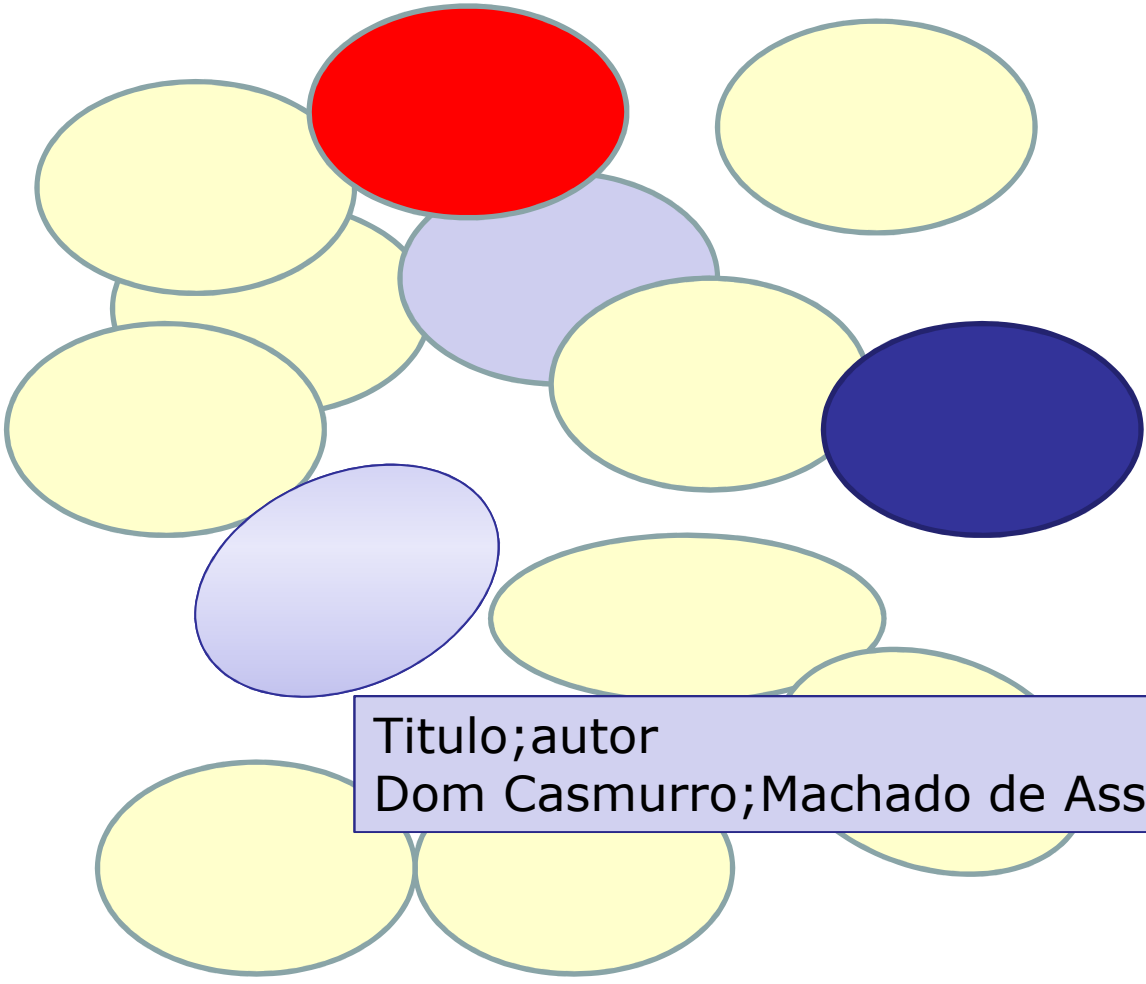
```
<xsd:element name="catalogo">
  <xsd:complexType>
    <xsd:sequence>
      <xsd:element name="livro"
        maxOccurs="unbounded">
        <xsd:complexType>
          <xsd:sequence>
            <xsd:element
              name="titulo"
              type="xsd:string"/>
            <xsd:element
              name="autor"
              type="xsd:string"/>
          </xsd:sequence>
        </xsd:complexType>
      </xsd:element>
    </xsd:sequence>
  </xsd:complexType>
</xsd:element>
```



```
<catalogo>
  <livro>
    <titulo>Dom Casmurro</titulo>
    <autor>Machado de Assis</autor>
  </livro>
</catalogo>
```



# O Problema: Interoperabilidade Semântica



# O Problema: Interoperabilidade Semântica

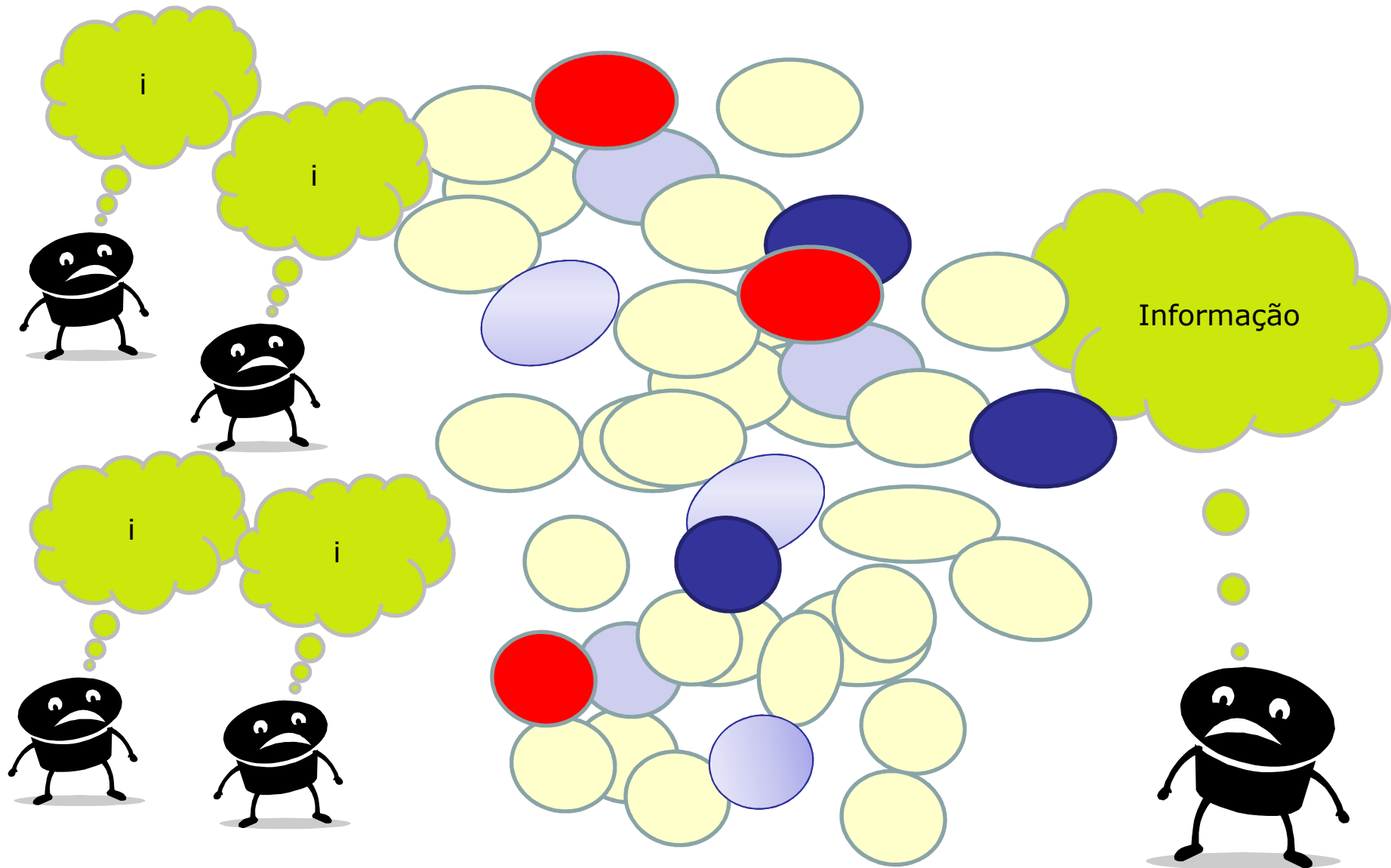


- Problema de interoperabilidade não é resolvido pela interoperabilidade sintática

<livro> representa:

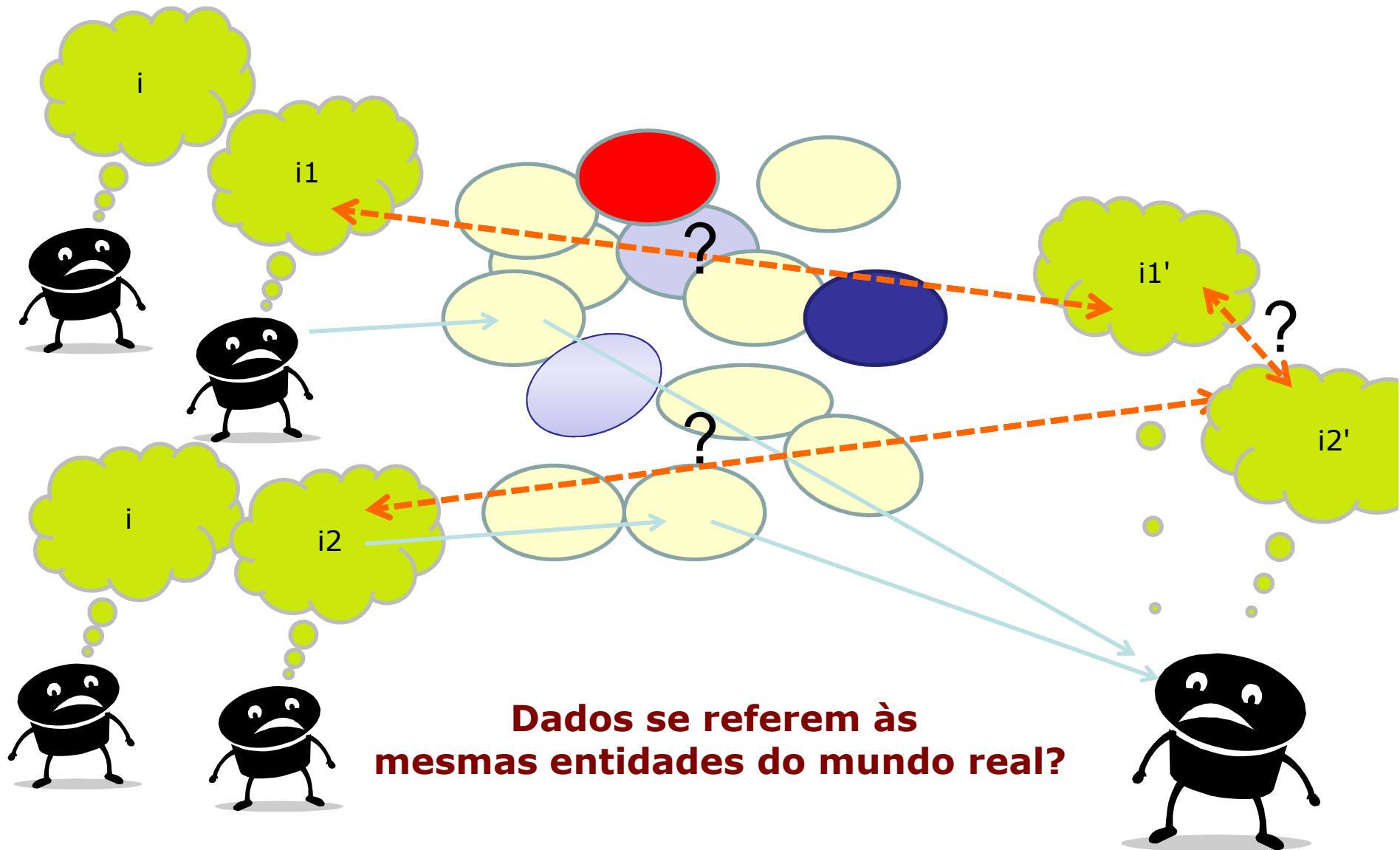
- Cópia (manifestação) específica do livro à disposição em uma acervo/biblioteca?
    - se houvesse 2 cópias teríamos 2 livros no catálogo?
  - Cópia do livro à venda em um catálogo de loja online?
    - uma entrada de <livro> representa muitas cópias
  - Edição do livro (não a cópia física)?
  - Livro como obra abstrata (aquilo que é preservado entre as várias edições?)
-

# Um problema sério de **escala**



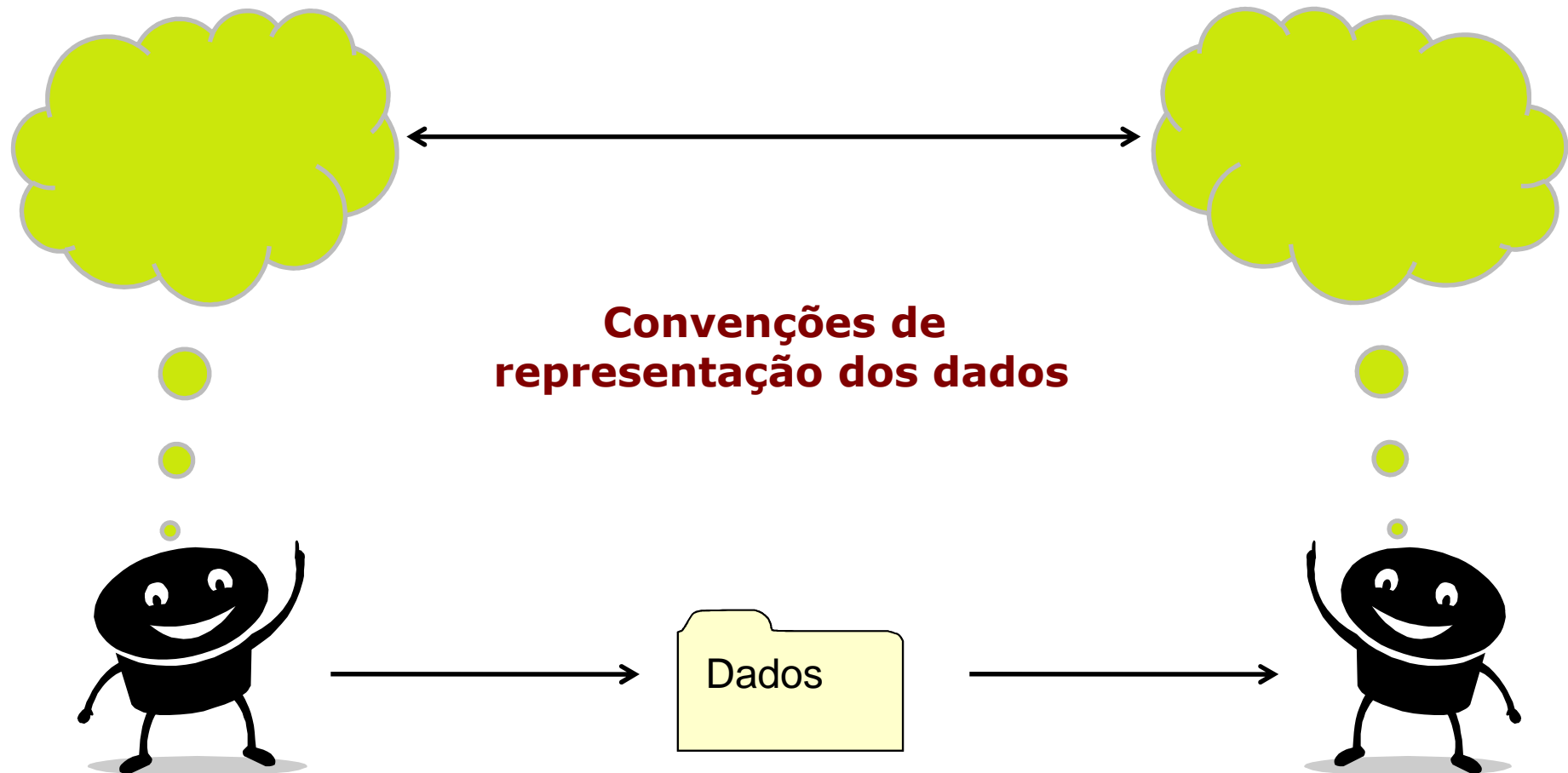


# Exacerbado na **ligação**

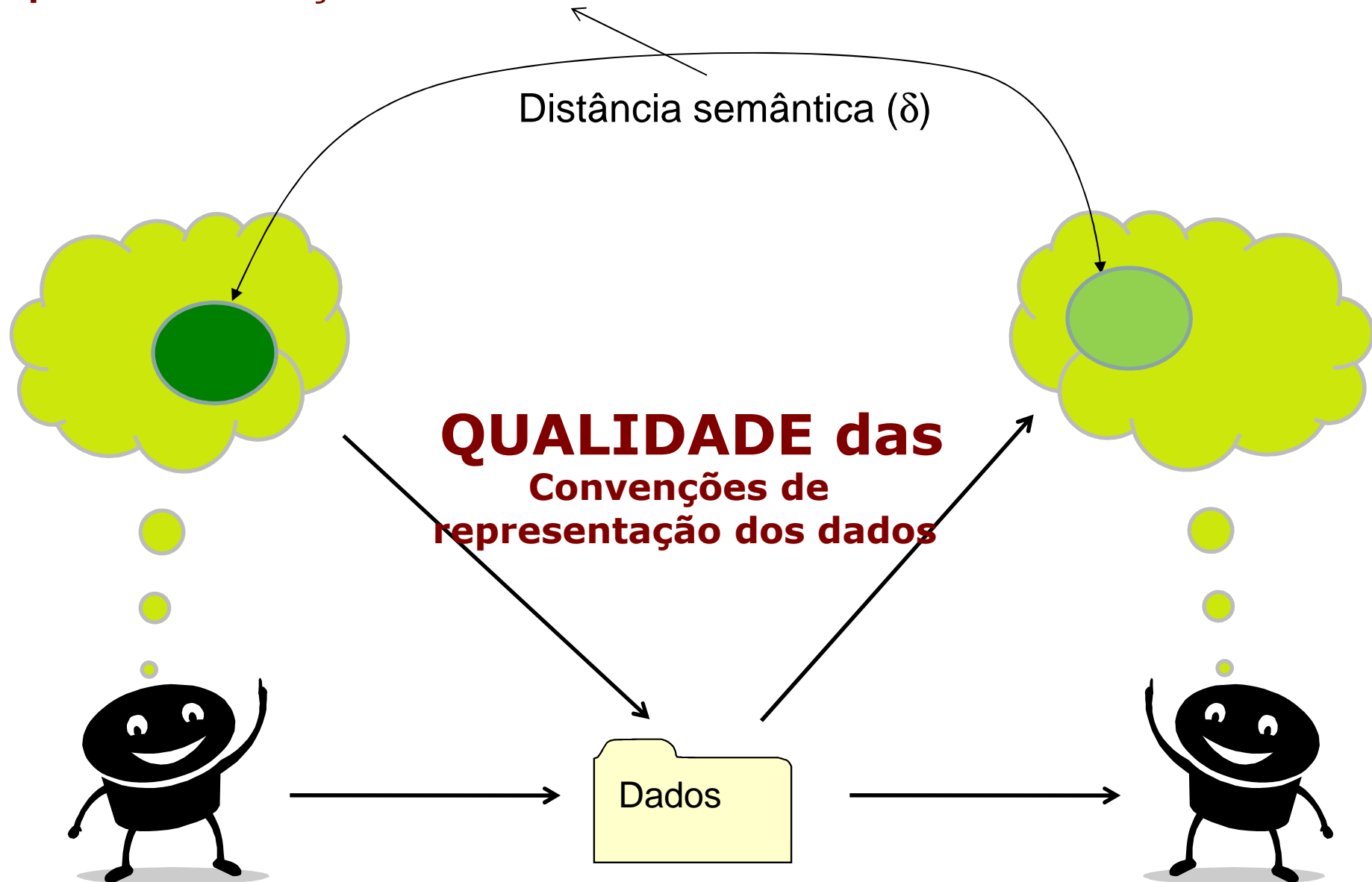


**Dados se referem às  
mesmas entidades do mundo real?**

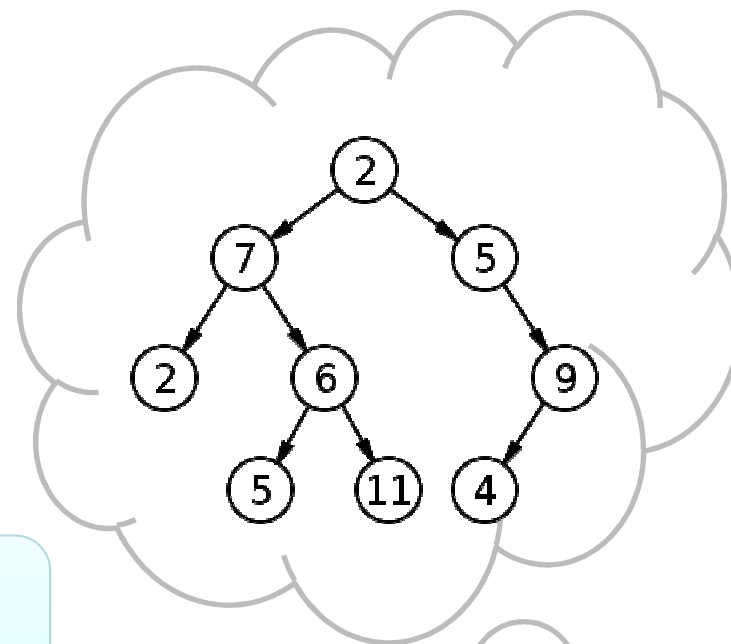
# Papel dos “formatos” de dados



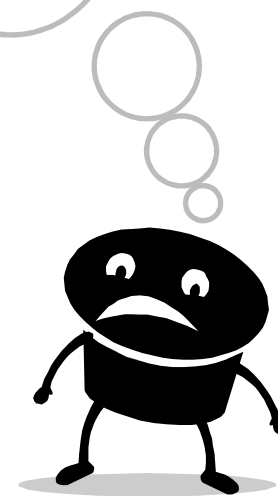
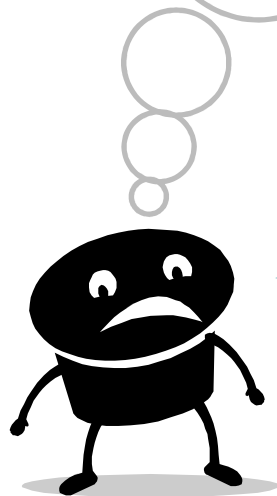
**Distância semântica tem que controlada para comunicação eficaz**



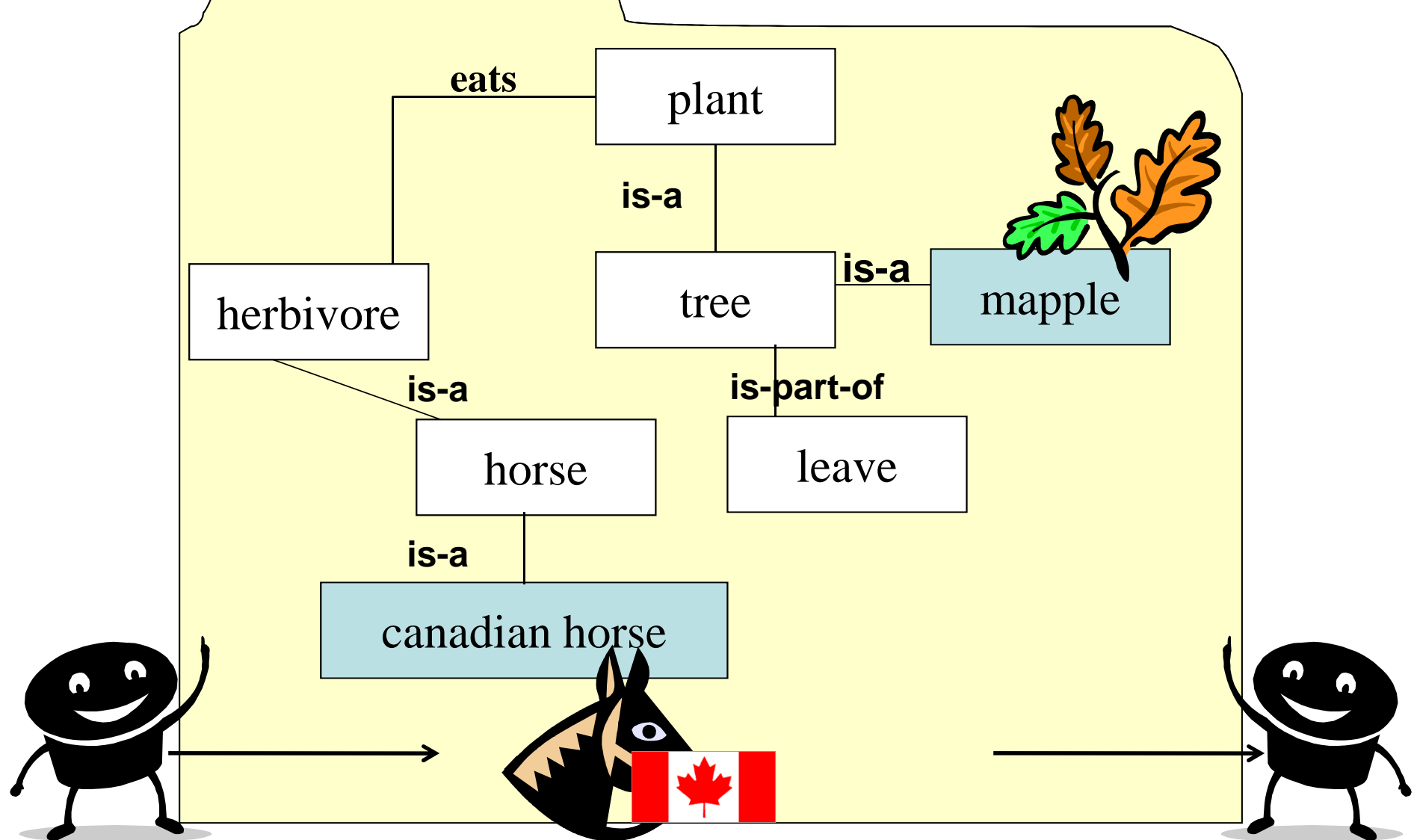
# Terminologia em Linguagem Natural Não é Solução Suficiente



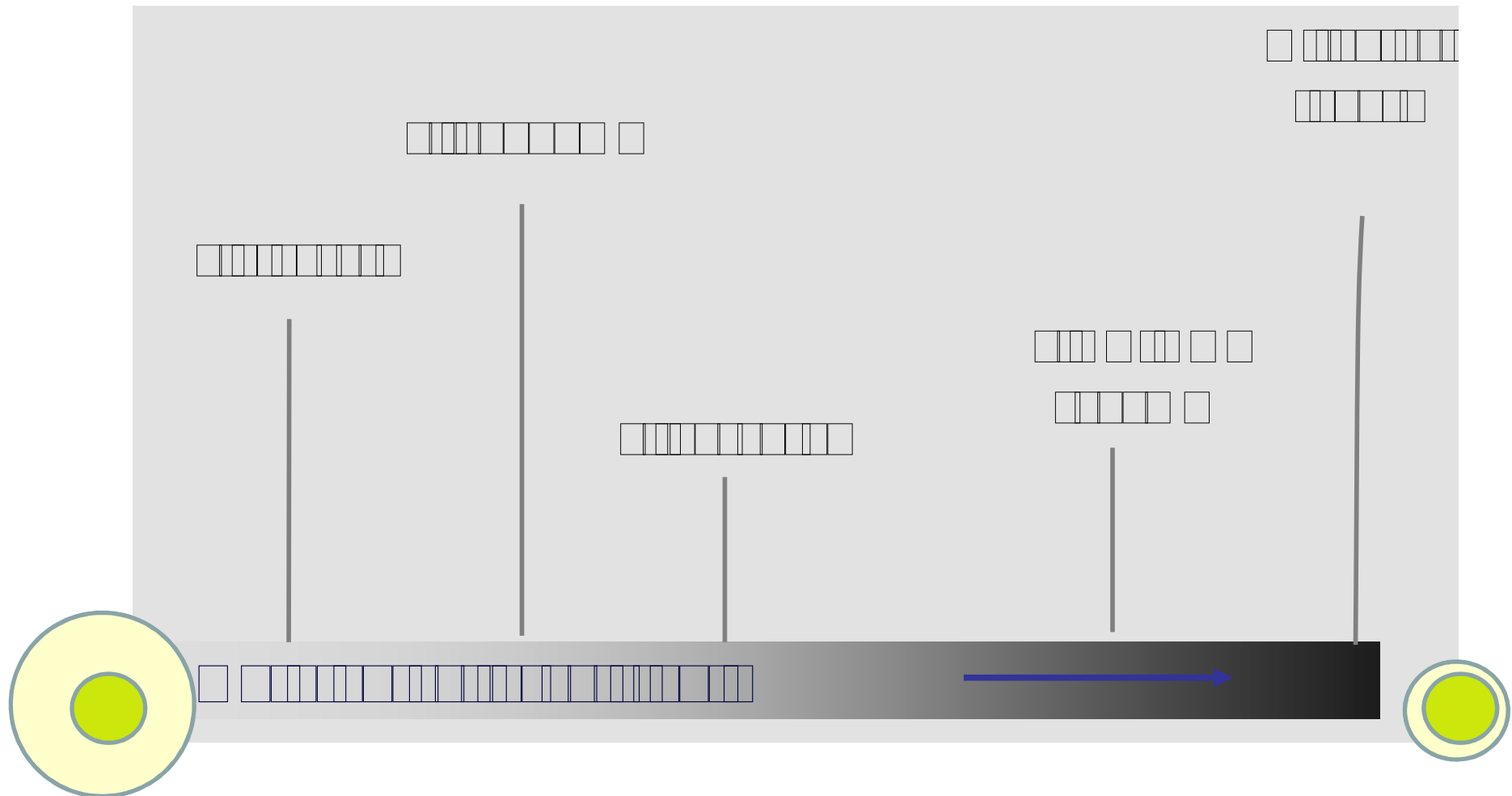
**ÁRVORE**



# Uso de modelos conceituais para comunicação



# Diferentes níveis de “precisão”, “rigor”, “estruturação”, “expressividade”

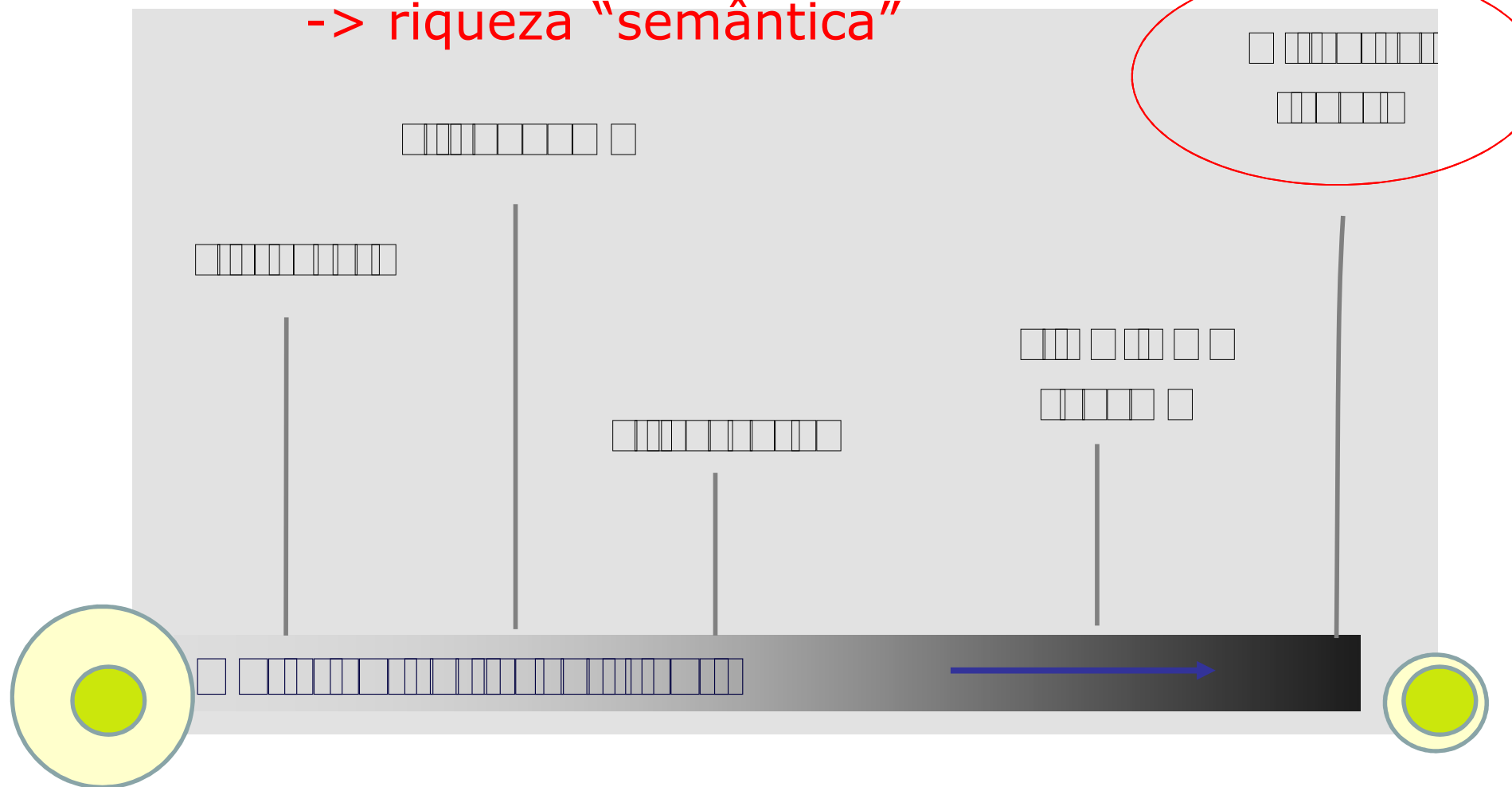


Adaptado de Alessandro Oltramari

# Diferentes níveis de “precisão”, “rigor”, “estruturação”, “expressividade”



Ainda diferentes níveis de “precisão”  
-> riqueza “semântica”

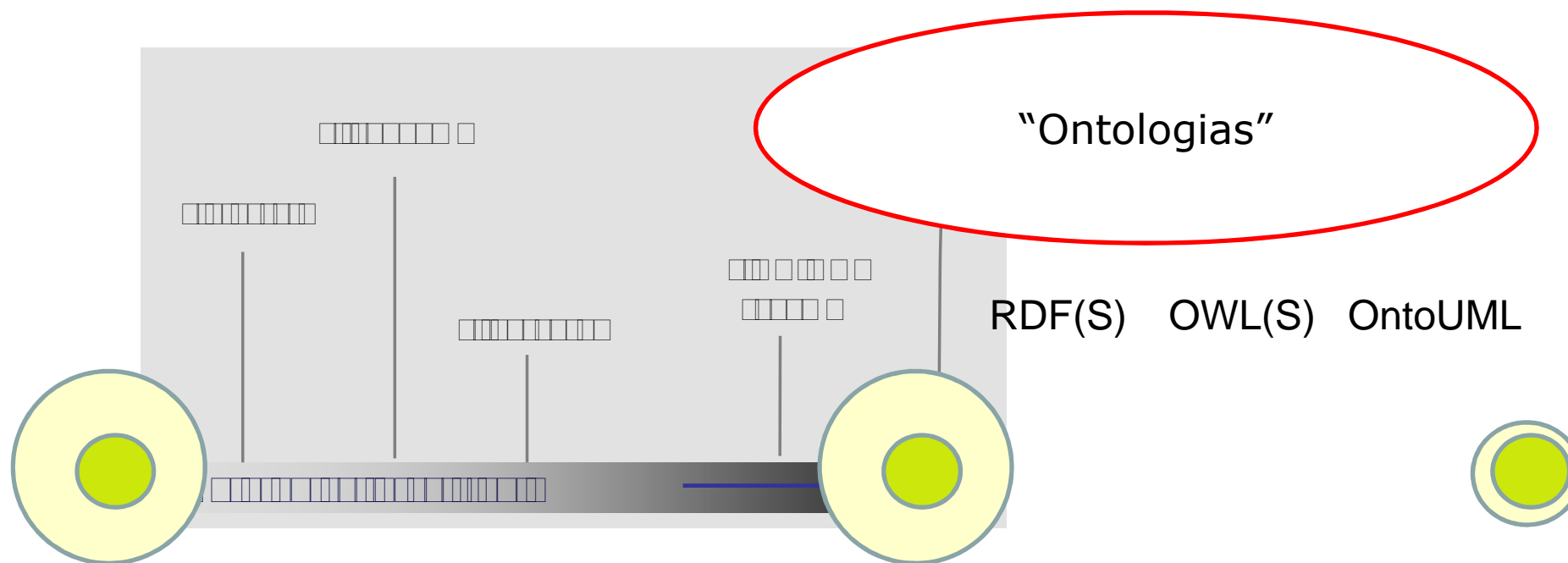


Adaptado de Alessandro Oltramari

Diferentes níveis de “precisão”, “rigor”,  
“estruturação”, “expressividade”

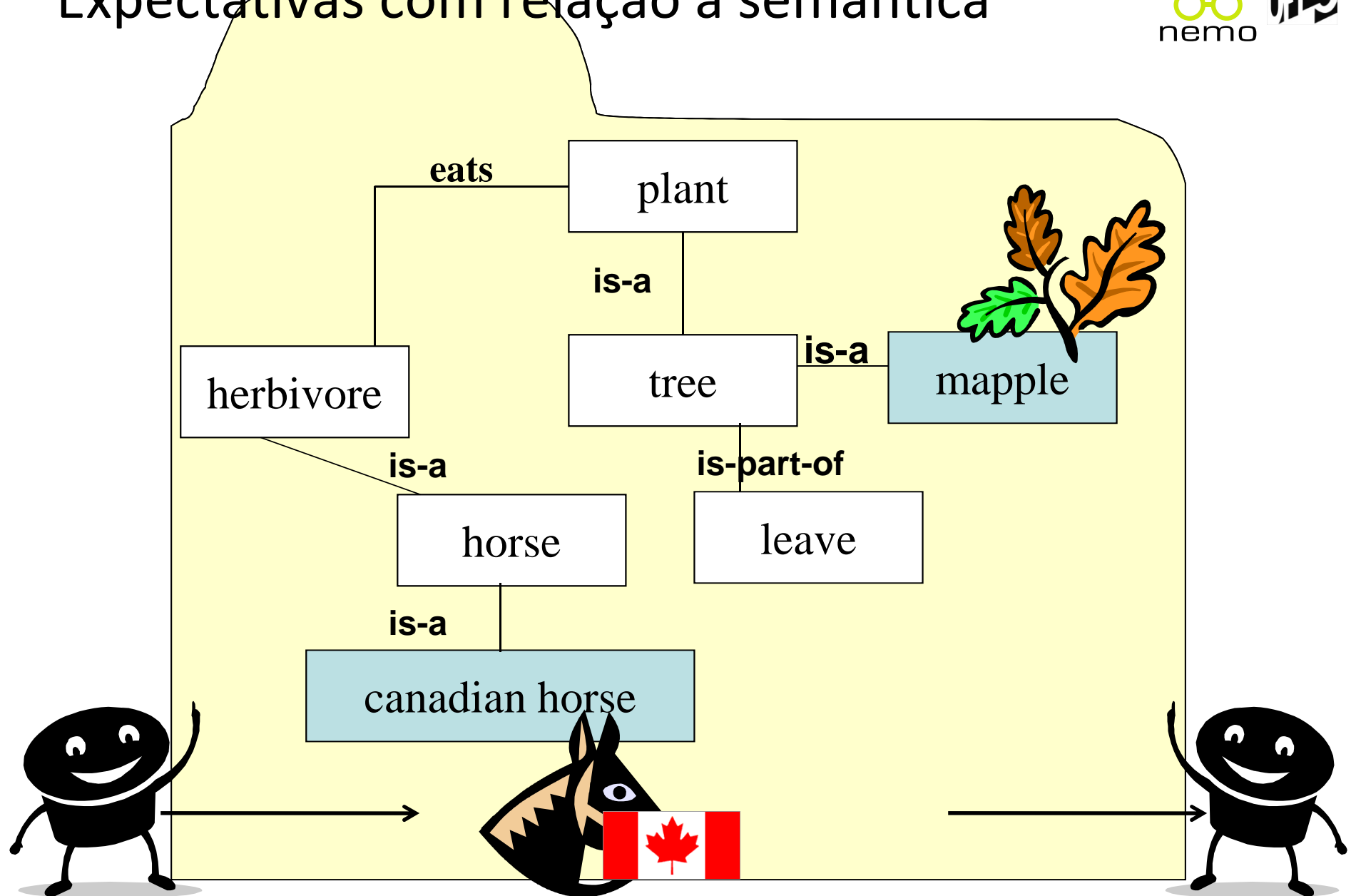


Ainda diferentes níveis de “precisão”  
-> riqueza “semântica”

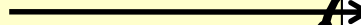
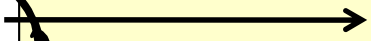
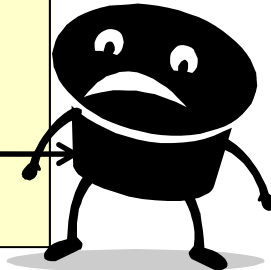
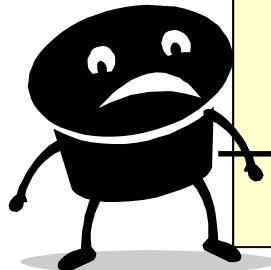
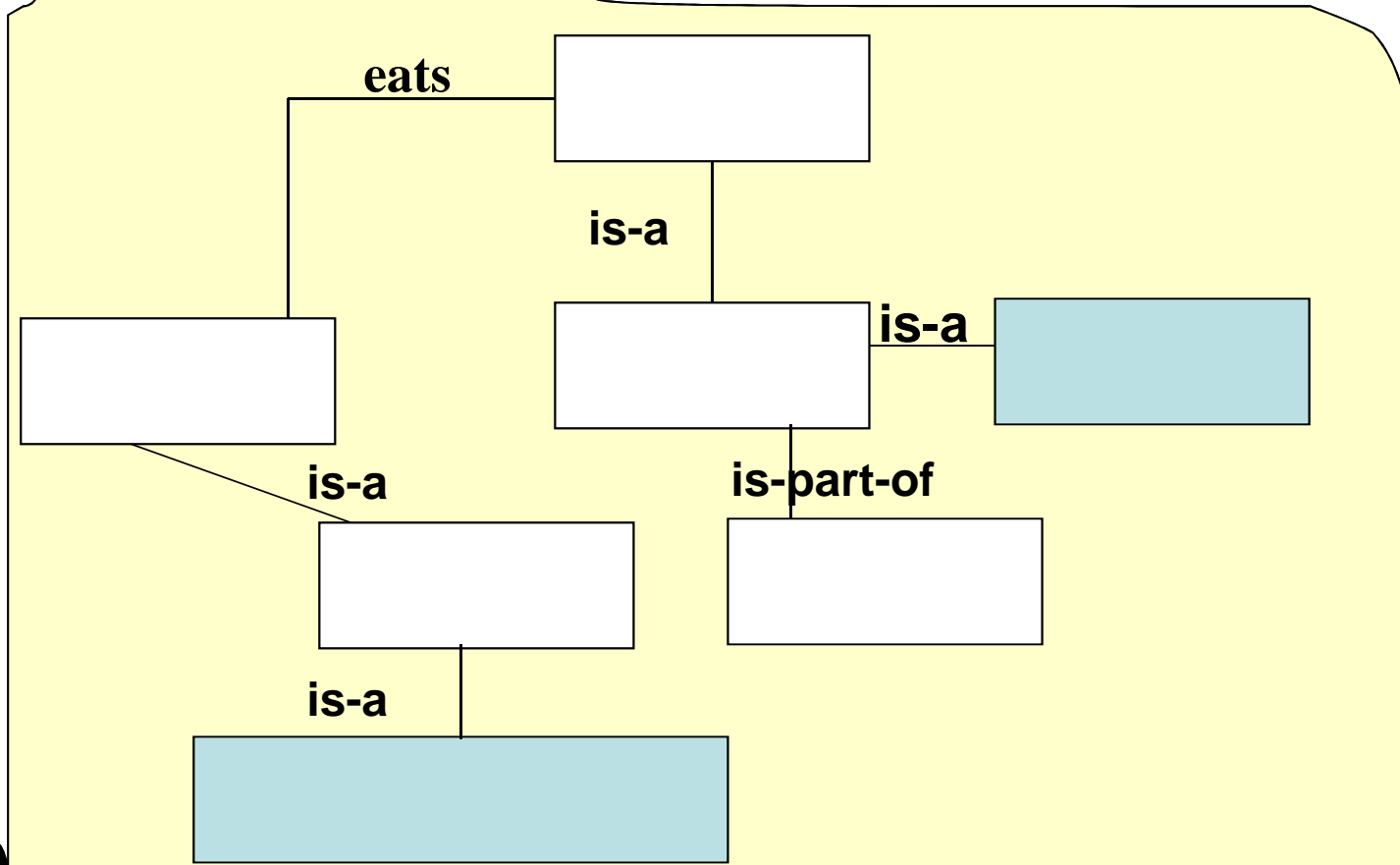




# Expectativas com relação à semântica



# Expectativas com relação à semântica



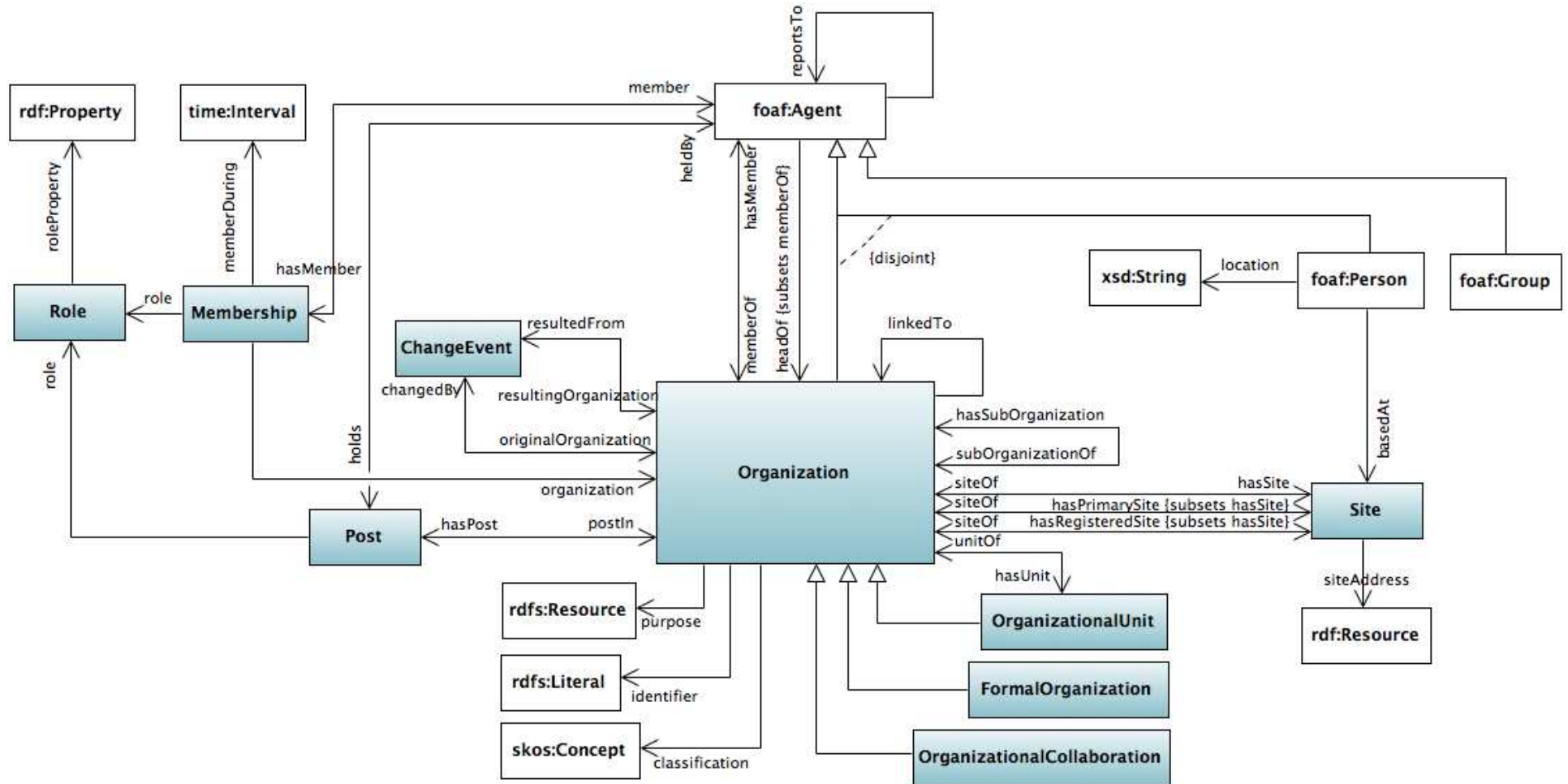
# Expectativas com relação à semântica

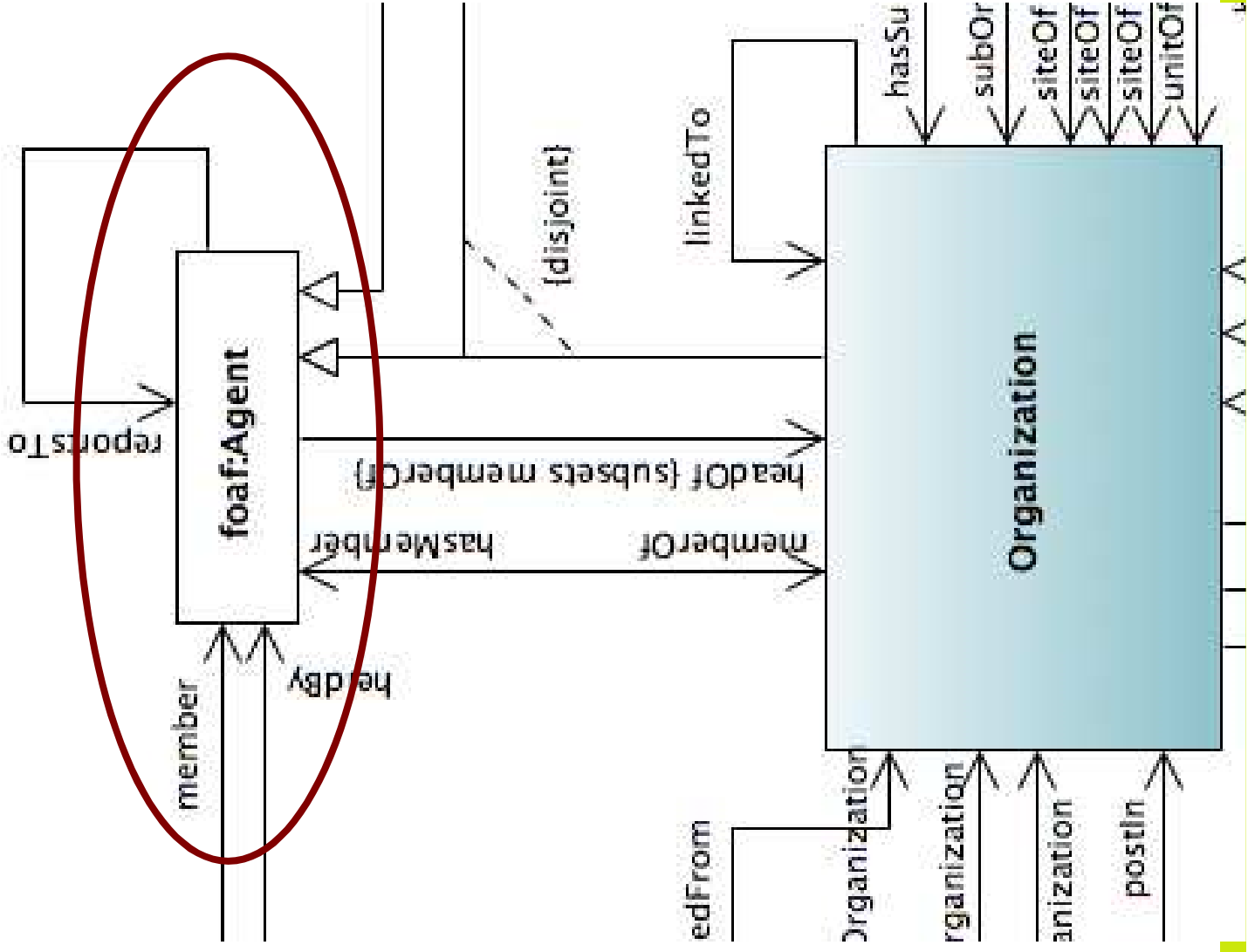


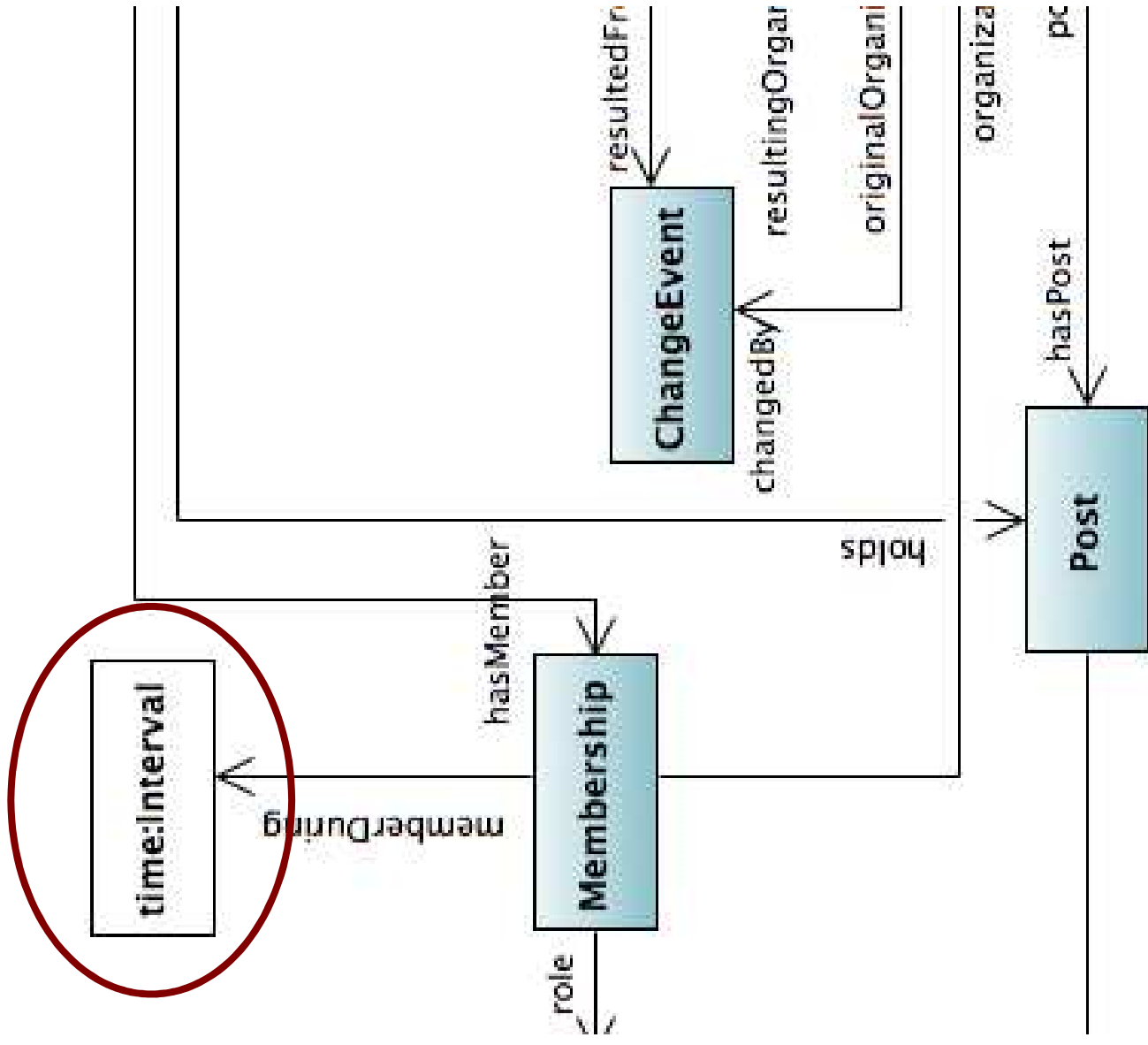
- De qualquer forma, nós ainda estamos dependendo da construção de consenso negociado através da linguagem natural
- O objetivo **é reduzir nossa dependência** da informalidade
- E construir uma ferramenta que permita que:
  - Aquele que crie o modelo conceitual expresse as distinções conceituais relevantes para o domínio
  - E aquele que leia o modelo consiga extrair as distinções conceituais relevantes para o domínio
- **Essencial para Dados Abertos**



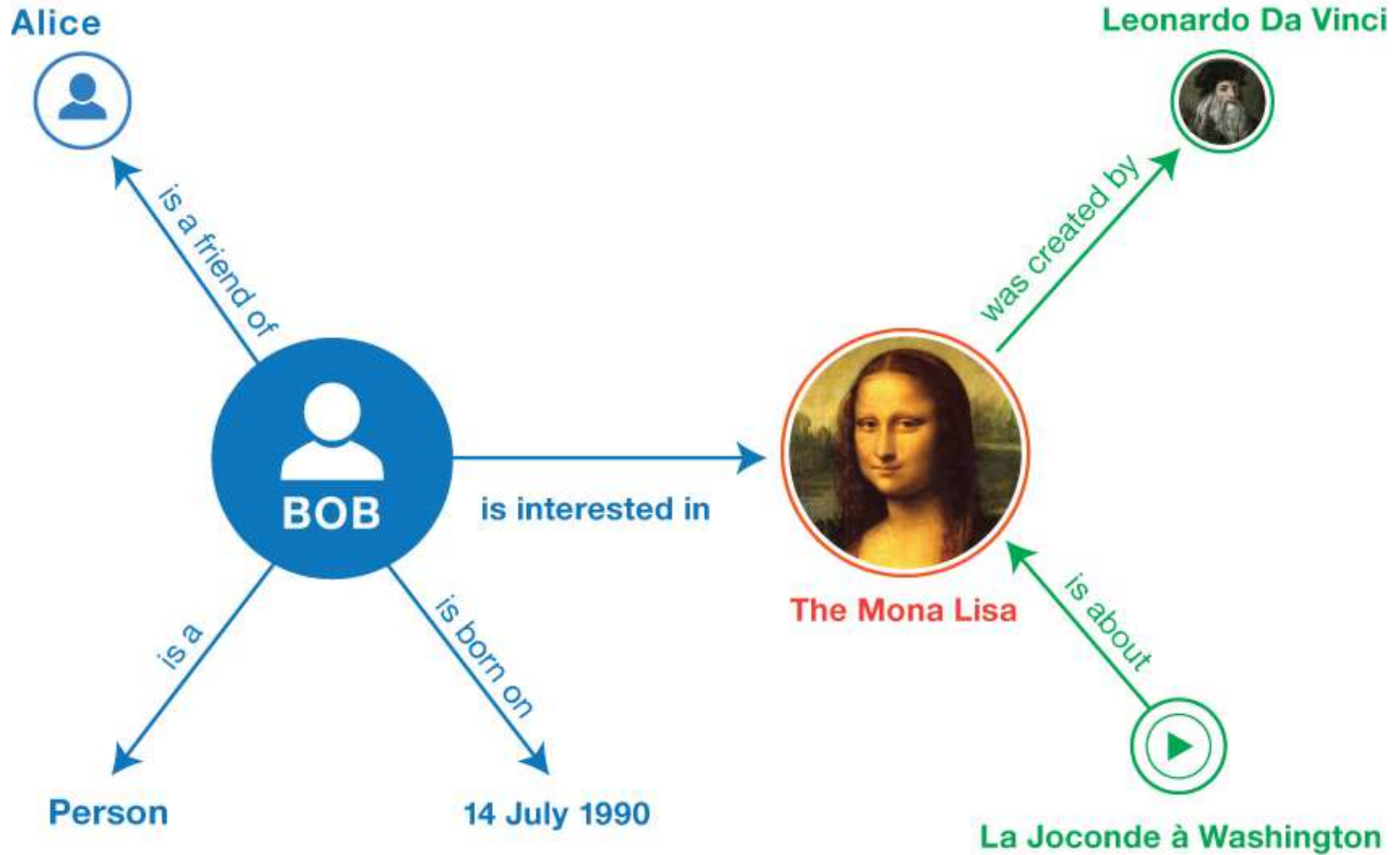
# W3C ORG Ontology



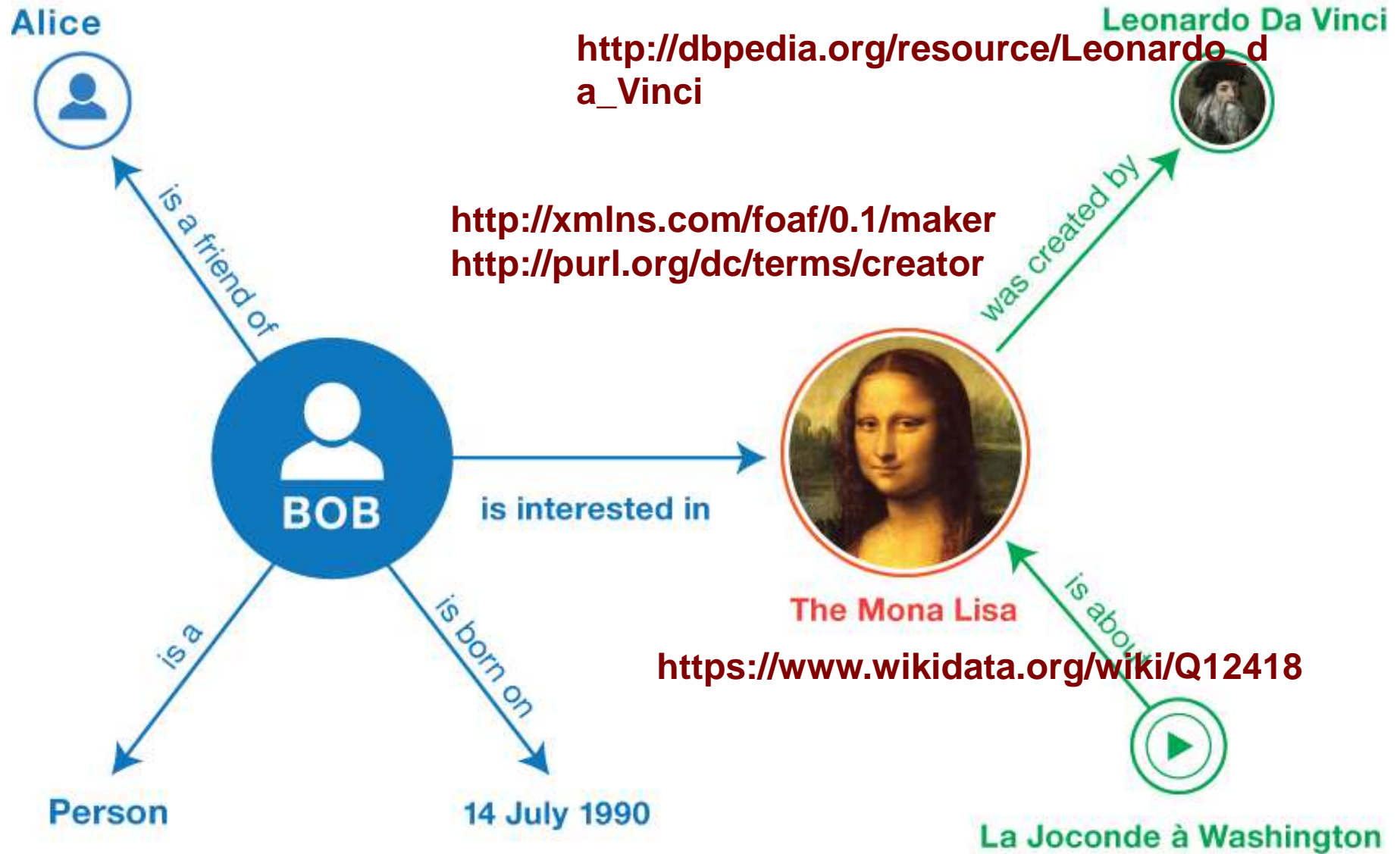




# Triplas e Grafos RDF na Web



# Triplas e Grafos RDF na Web - IRIs





Estudo (Bassetti, Azevedo, Almeida, 2014)



- **Servidores Civis (Servidores e Remuneração - Novembro/2013)**

- Autor: Ministério do Planejamento, Banco Central, Comandos Militares, MRE, DPF, RFB, MAPA, DEST e Ministérios [Honorários (Jetons)]

- Disponível em:

- <http://www.portaltransparencia.gov.br/downloads/servidores.asp>

- **Sistema de Informações Organizacionais do Governo Federal (SIORG)**

- Autor: Ministério do Planejamento, Orçamento e Gestão (MP)

- Disponível em: <http://dados.gov.br/dataset/siorg>

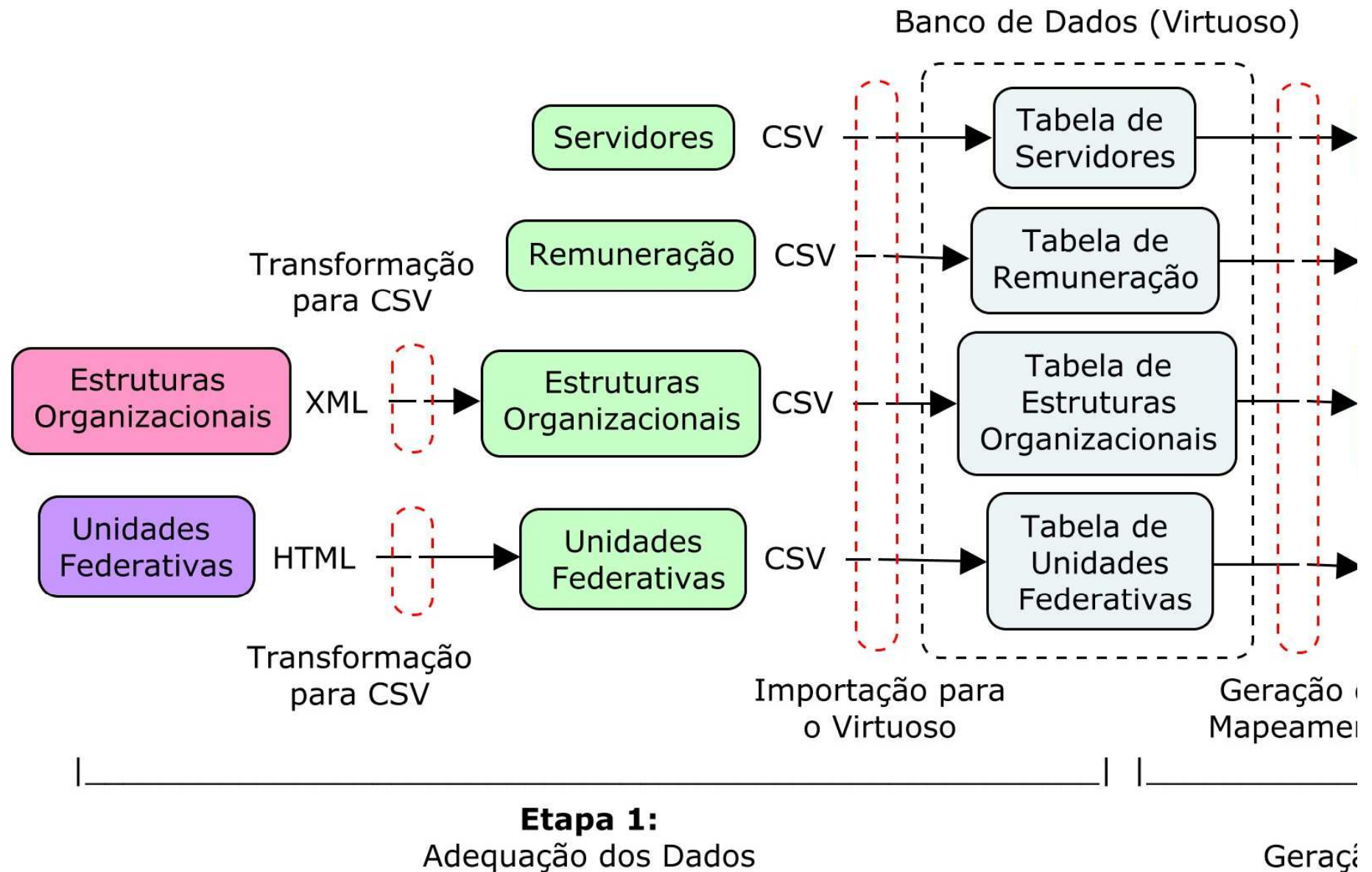
- **Unidade Federativas**

- Autor: Instituto Brasileiro de Geografia e Estatística (IBGE)

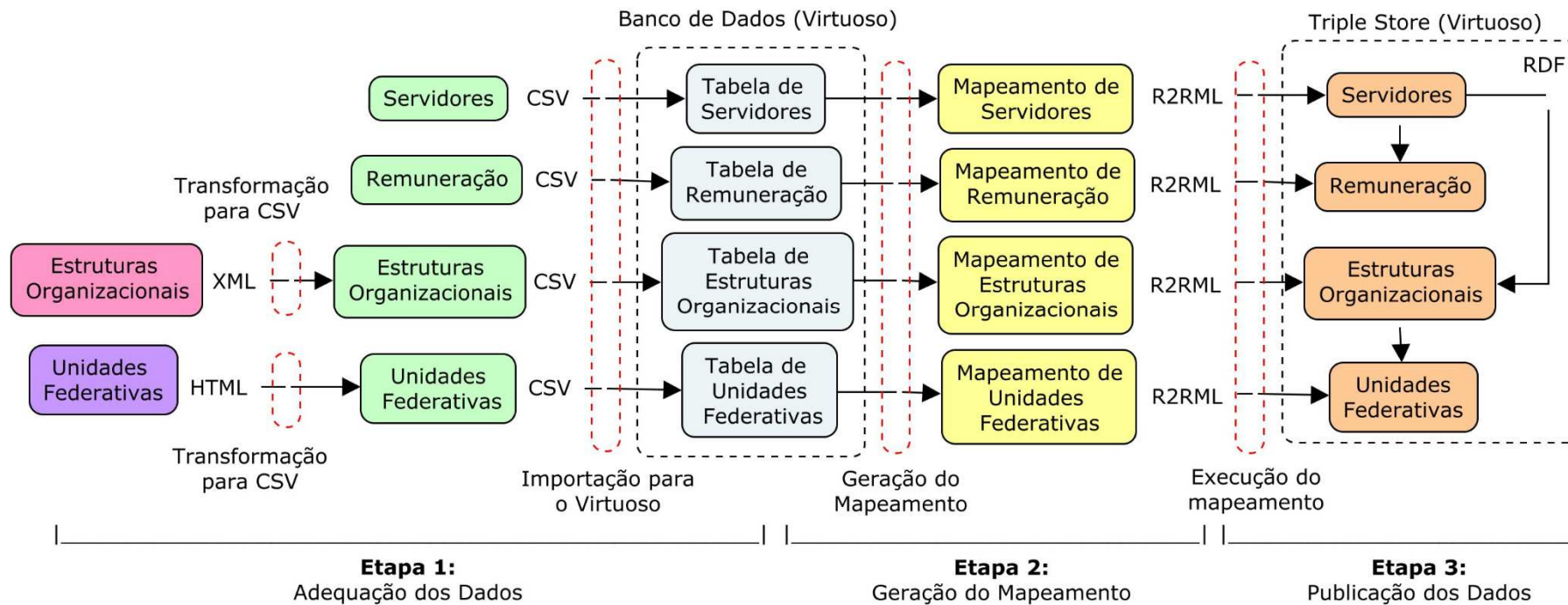
- Disponível em:

- <http://www.ibge.gov.br/home/geociencias/areaterritorial/principal.shtm>

# Fontes em Formatos Diferentes



# Abordagem



# Representação Inconsistente



- Representação para Presidência da República na base de dados de Servidores Civis
- 20101 – PRESIDENCIA DA REPUBLICA
- Representação para Presidência da República na base de dados de Estruturas Organizacionais
- 26 – Presidência da República



# Descrições Inadequadas



DADOS_CADASTRO_NOME	DADOS_CADASTRO_CODIGO	DADOS_CADASTRO_CODIGO_PAI
MINISTERIO DA FAZENDA	1929	26
MINISTERIO DA FAZENDA	118699	1929
MINISTERIO DA FAZENDA	118700	1929
MINISTERIO DA FAZENDA	118701	1929
MINISTERIO DA FAZENDA	118702	1929
MINISTERIO DA FAZENDA	118703	1929
MINISTERIO DA FAZENDA	118704	1929
MINISTERIO DA FAZENDA	118708	1929
MINISTERIO DA FAZENDA	118709	1929

# Falta de Identificação Única



<b>CARGO</b>
AGENTE DE SEGURANCA FERROVIARIA
AGENTE DE SEGURANCA FERROVIARIO
AGENTE DE TELEC E ELETRICIDADE
AGENTE DE TELEC ELETRICIDADE
AGENTE DE TELECOMUNI E ELETRICIDADE
AGENTE DE TELECOMUNIC E ELETRICIDADE
AGENTE DE TRANSP MARITIMO E FLUVIAL
AGENTE DE TRANSP MARITIMO FLUVIAL



# Impacto dos Problemas das Bases de Dados no Consumo



Property	Value
<a href="http://www.w3.org/1999/02/22-rdf-syntax-ns#type">http://www.w3.org/1999/02/22-rdf-syntax-ns#type</a>	<a href="http://www.w3.org/ns/org#OrganizationalUnit">http://www.w3.org/ns/org#OrganizationalUnit</a>
<a href="http://www.w3.org/2004/02/skos/core#prefLabel">http://www.w3.org/2004/02/skos/core#prefLabel</a>	COORDENACAO DE PLANEJAMENTO
<a href="http://www.w3.org/ns/org#unitOf">http://www.w3.org/ns/org#unitOf</a>	<a href="http://localhost:8080/NOM/resource/organization/36205">http://localhost:8080/NOM/resource/organization/36205</a>
<a href="http://www.w3.org/ns/org#unitOf">http://www.w3.org/ns/org#unitOf</a>	<a href="http://localhost:8080/NOM/resource/organization/25000">http://localhost:8080/NOM/resource/organization/25000</a>
<a href="http://www.w3.org/ns/org#unitOf">http://www.w3.org/ns/org#unitOf</a>	<a href="http://localhost:8080/NOM/resource/organization/40112">http://localhost:8080/NOM/resource/organization/40112</a>
<a href="http://www.w3.org/ns/org#unitOf">http://www.w3.org/ns/org#unitOf</a>	<a href="http://localhost:8080/NOM/resource/organization/45203">http://localhost:8080/NOM/resource/organization/45203</a>
<a href="http://www.w3.org/ns/org#unitOf">http://www.w3.org/ns/org#unitOf</a>	<a href="http://localhost:8080/NOM/resource/organization/20224">http://localhost:8080/NOM/resource/organization/20224</a>
<a href="http://www.w3.org/ns/org#unitOf">http://www.w3.org/ns/org#unitOf</a>	<a href="http://localhost:8080/NOM/resource/organization/28000">http://localhost:8080/NOM/resource/organization/28000</a>
<a href="http://www.w3.org/ns/org#unitOf">http://www.w3.org/ns/org#unitOf</a>	<a href="http://localhost:8080/NOM/resource/organization/26442">http://localhost:8080/NOM/resource/organization/26442</a>





http://sws.geonames.org/3462386



GeoNames

About ▾

Browse ▾

Download ▾

API ▾

Help ▾

Paris, Mount Everest, New York



Found 3 items in this area ▾



Layers



**Goiabeiras** ca. 4 m

**P PPLL** populated locality

3462386

Brazil BR » Espírito Santo 08 » Vitória 3205309

-20.26444, -40.2975

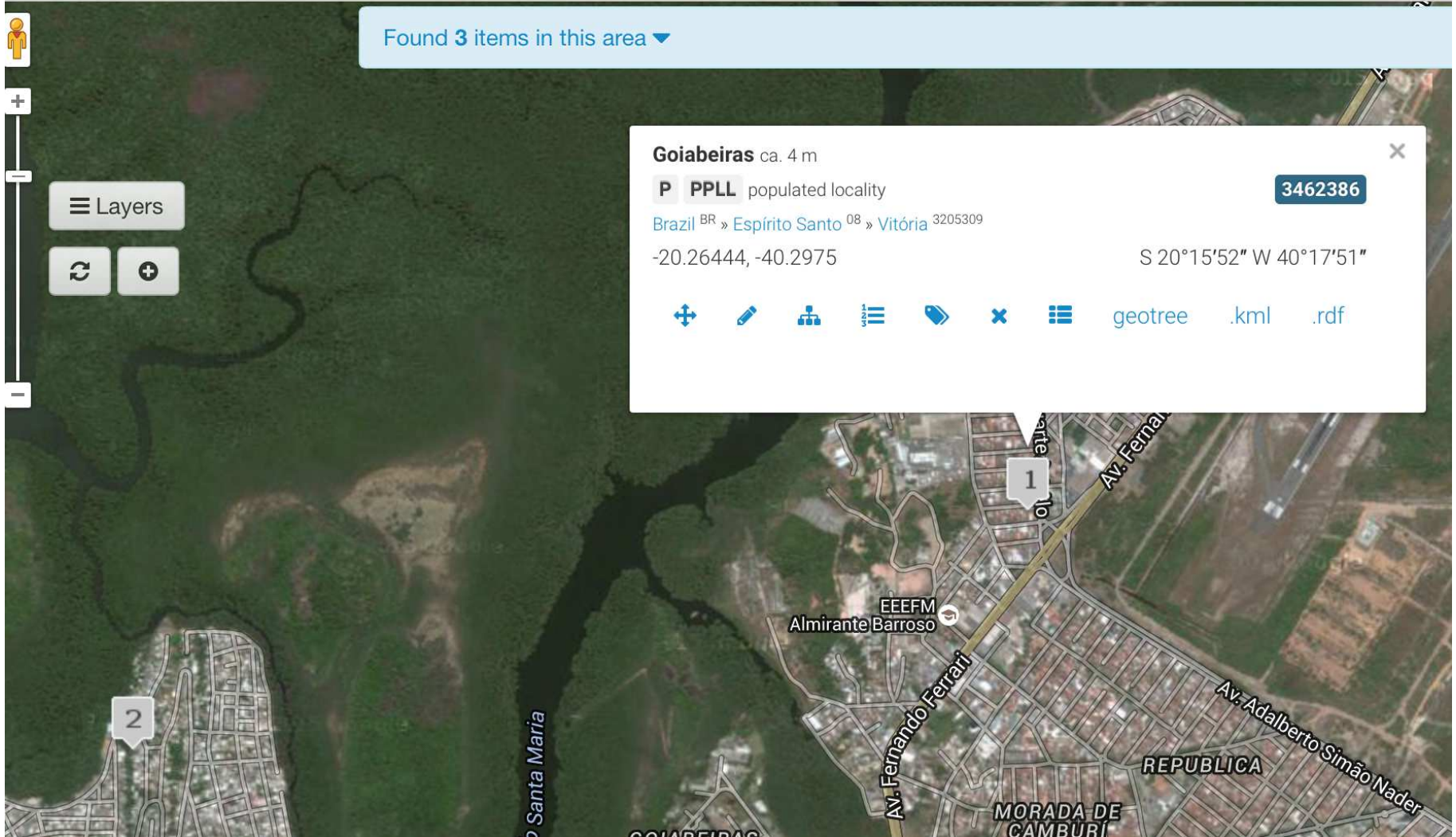
S 20°15'52" W 40°17'51"



geotree

.kml

.rdf







Layers



## 3462386 Goiabeiras

Feature

Hierarchy

History

Tags

Alternate names

Administrative hierarchy

Parents

Children

country

Brazil (BR)

adm1

Espírito Santo (08)

adm2

Vitória (3205309)

adm3

no admin3

adm4

no admin4

adm5

no admin5



2

a Maria

Fernando

Adalberto

REPUBLICA

http://sws.geonames.org/3462386/about.rdf

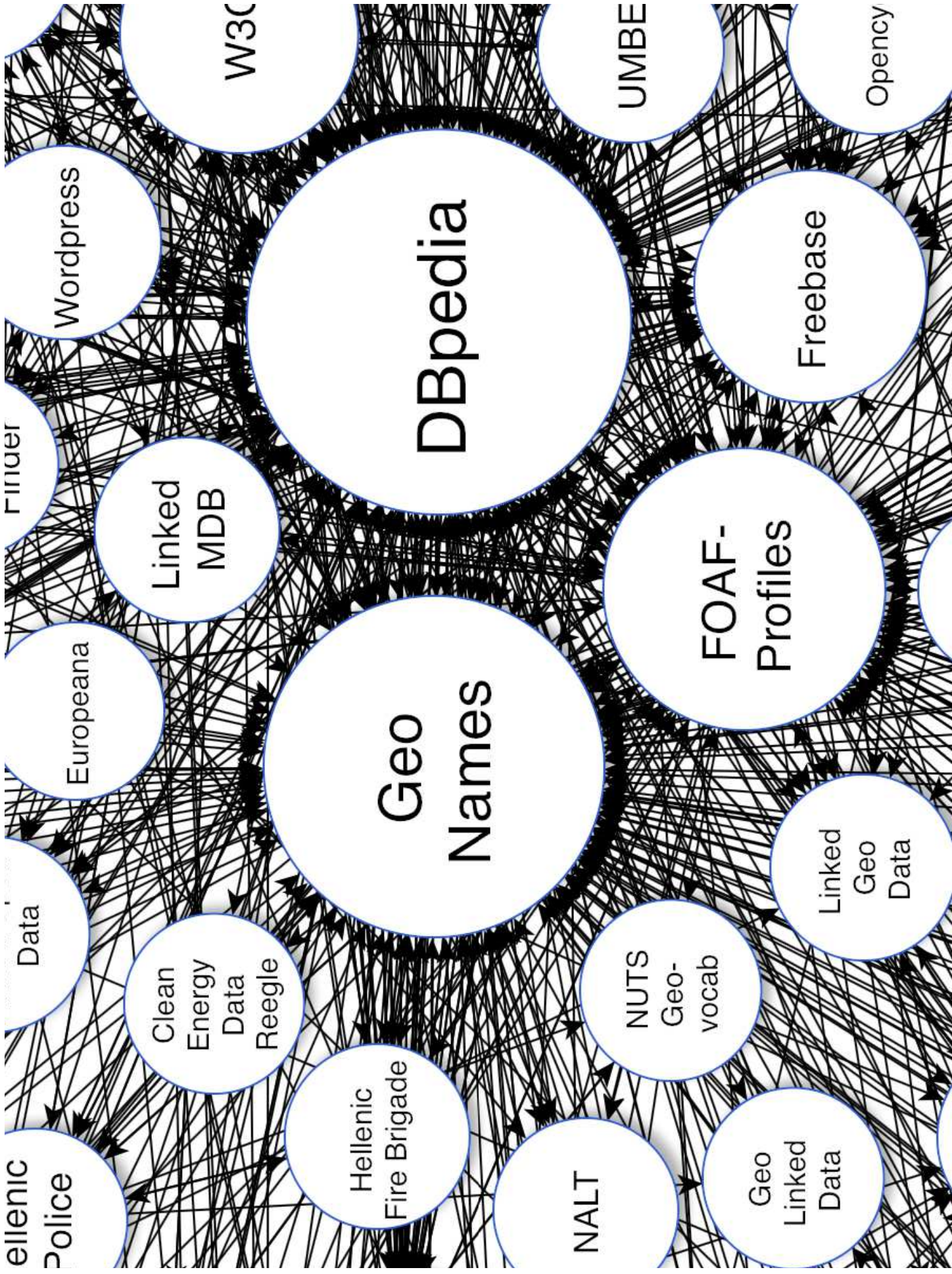


```
- <rdf:RDF>
- <gn:Feature rdf:about="http://sws.geonames.org/3462386/">
  <rdfs:isDefinedBy rdf:resource="http://sws.geonames.org/3462386/about.rdf"/>
  <gn:name>Goiabeiras</gn:name>
  <gn:featureClass rdf:resource="http://www.geonames.org/ontology#P"/>
  <gn:featureCode rdf:resource="http://www.geonames.org/ontology#P.PPLL"/>
  <gn:countryCode>BR</gn:countryCode>
  <wgs84_pos:lat>-20.26444</wgs84_pos:lat>
  <wgs84_pos:long>-40.2975</wgs84_pos:long>
  <gn:parentFeature rdf:resource="http://sws.geonames.org/6322000/">
  <gn:parentCountry rdf:resource="http://sws.geonames.org/3469034/">
  <gn:parentADM1 rdf:resource="http://sws.geonames.org/3463930/">
  <gn:parentADM2 rdf:resource="http://sws.geonames.org/6322000/">
  <gn:nearbyFeatures rdf:resource="http://sws.geonames.org/3462386/nearby.rdf"/>
  <gn:locationMap rdf:resource="http://www.geonames.org/3462386/goiabeiras.html"/>
</gn:Feature>
- <foaf:Document rdf:about="http://sws.geonames.org/3462386/about.rdf">
  <foaf:primaryTopic rdf:resource="http://sws.geonames.org/3462386/">
  <cc:license rdf:resource="http://creativecommons.org/licenses/by/3.0/">
  <cc:attributionURL rdf:resource="http://sws.geonames.org/3462386/">
  <cc:attributionName rdf:datatype="http://www.w3.org/2001/XMLSchema#string">GeoNames</cc:attributionName>
  <dcterms:created rdf:datatype="http://www.w3.org/2001/XMLSchema#date">2006-01-15</dcterms:created>
  <dcterms:modified rdf:datatype="http://www.w3.org/2001/XMLSchema#date">2012-08-03</dcterms:modified>
</foaf:Document>
</rdf:RDF>
```









# Outras informações



- W3C Data on the Web Best Practices Working Group
- W3C Government Linked Data Working Group
- <http://nemo.inf.ufes.br>
- <http://vocab.e.gov.br>
- <http://eping.governoeletronico.gov.br>
- Linked Open Data
- Open Government Data



# Conclusão



- Principais **desafios são sócio-técnicos**
- Coordenação de formatos, modelos conceituais subjacentes
  - Técnicas “semânticas” ou baseadas em ontologias
  - Processo aberto e colaborativo para construção destes formatos
- Investimento na **qualidade destes formatos**
- Investimento na **qualidade dos dados**
  
- Esforços significativos são necessários por parte do governo
  - Envolvendo necessariamente investimento em TI pelo seu papel chave neste processo