Semantic Web

Yue Ma

Laboratoire de Recherche en Informatique (LRI)
Université Paris Sud
ma@lri.fr

More than Horn ?

MOTIVATION :

to have a higher expressivity in representing knowledge

$RB = \{ r_1$: if *GoodWork* and not *LateSubmission*, then *NoPenalty*

r_2 : if *NoPenalty*, then *GoodScore*

$\}$

$FB = \{$ *GoodWork* $\}$

What can we deduce from the knowledge base $RB \cup FB$?

More than Horn ?

MOTIVATION :

to have a higher expressivity in representing knowledge

$RB = \{r_1$: if *GoodWork* and not *LateSubmission*, then *NoPenalty*
r_2 : if *NoPenalty*, then *GoodScore*
r_3 : if *SubmissionAfterOct*, then *LateSubmission*$\}$
$FB = \{$*GoodWork*, *SubmissionAfterOct*$\}$

What can we deduce from the knowledge base $RB \cup FB$?

Negation as Failure

Definition

$match(not\ p, BF, BR) = true$ iff. $BR \cup BF$ does not imply p.

In terms of forward- and backward-chaining, it will be ...

$match(not\ p, BF, BR) = true$

Negation as Failure

Definition

$match(\textit{not}\ p, BF, BR) = \textit{true}$ iff. $BR \cup BF$ does not imply p.

In terms of forward- and backward-chaining, it will be ...

$match(\textit{not}\ p, BF, BR) = \textit{true}$
iff. $p \notin ForwardChaining(BF, BR)$

Negation as Failure

Definition

$match(\textit{not}\ p, BF, BR) = \textit{true}$ iff. $BR \cup BF$ does not imply p.

In terms of forward- and backward-chaining, it will be ...

$match(\textit{not}\ p, BF, BR) = \textit{true}$

iff. $p \notin \textit{ForwardChaining}(BF, BR)$

iff. $\textit{BackwardChaining}(p, BF, BR) = \textit{false}$

Forward-chaining rule v.s. Material Implication

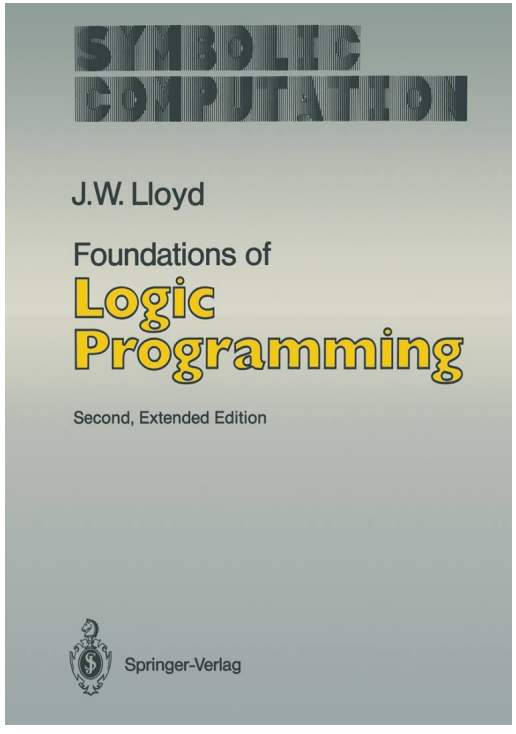
- $RB = \{r_1 : \text{if } p \text{ then } q\}$
- $FB = \{\neg q\}$
- What can we deduce from $RB \cup FB$ by forward-chaining ?
- Is it reasonable to deduce $\neg p$ from $KB = RB \cup FB$?

Forward-chaining rule v.s. Material Implication

- $RB = \{r_1 : \text{if } p \text{ then } q\}$
- $FB = \{\neg q\}$
- What can we deduce from $RB \cup FB$ by forward-chaining ?
- Is it reasonable to deduce $\neg p$ from $KB = RB \cup FB$?

More topics on Horn rule and its extensions (aka logic programming)

The Potassco system for answer set based logic programming (`http://potassco.sourceforge.net`)
And



Limits of propositional logic

- Example 1
  - All man is mortal,
  - Scorate is a man,
  - So Socrate is mortal.
- In propositional logic :
  - p : All man is mortal,
  - q : Scorate is a man,
  - r : Socrate is mortal.
  - : So we have $p \wedge q \rightarrow r$ (What do we lose ?)

Limits of propositional logic

- for all x, if x is a man so x is mortal,
- Scorate is a man,
- So Socrate is mortal.
- Einstein is a man,
- So Einstein is mortal.
- ...

In predicate logic (first-order logic) :

"x is a man" and "x is mortal" are represented by $H(x)$ and $M(x)$

- The piece of knowledge "for all x, if x is a man so x is mortal" can be represented by

$$\forall x(H(x) \rightarrow M(x)).$$

$$\{\underline{\forall x(H(x) \rightarrow M(x))}, H(Socrate)\} \models M(Socrate)$$
$$\{\underline{\forall x(H(x) \rightarrow M(x))}, H(Einstein)\} \models M(Einstein)$$

Limits of propositional logic

Example 2. In predicate logic (first-order logic) :

$capitalOf(x, y)$: x is the capital of y.

$locatedIn(x, y)$: x is located in y.

- $capitalOf(x, y) \wedge locatedIn(y, z) \rightarrow locatedIn(x, z)$
- $capitalOf(Paris, France), locatedIn(France, Europe)$

So, what can we deduce ?

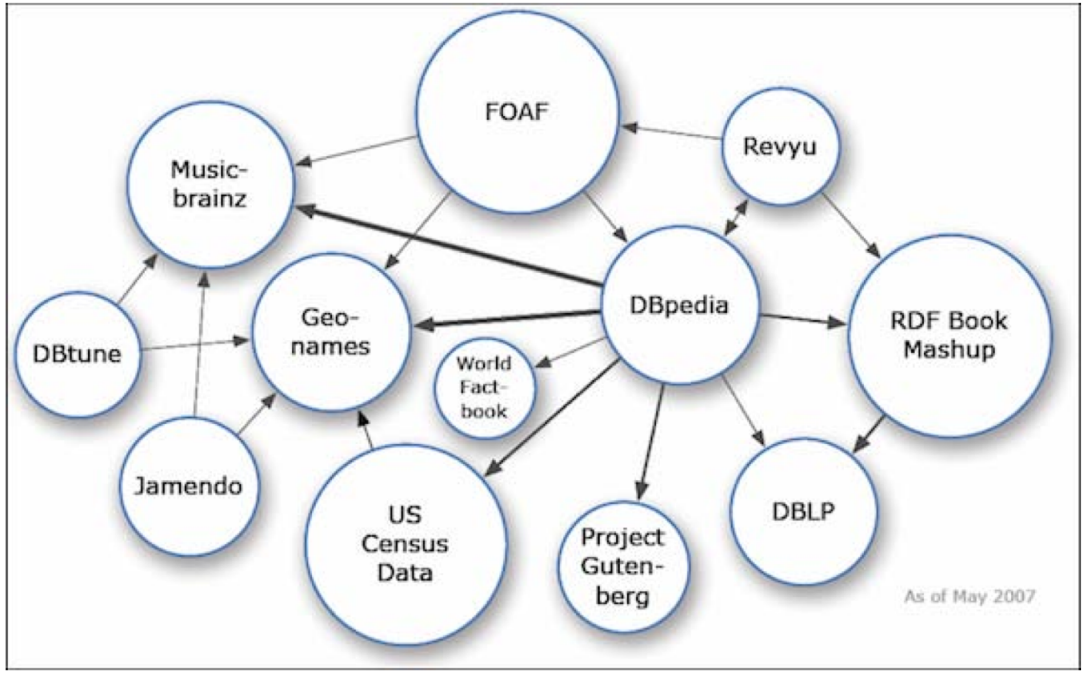
Limits of propositional logic

Example 3 : TED talk of Tim Berners-Lee (2009)

Semantic Web, Linked Data, ...
from http ://www.w3.org/DesignIssues/LinkedData.html (Tim Berners-Lee, 2006)

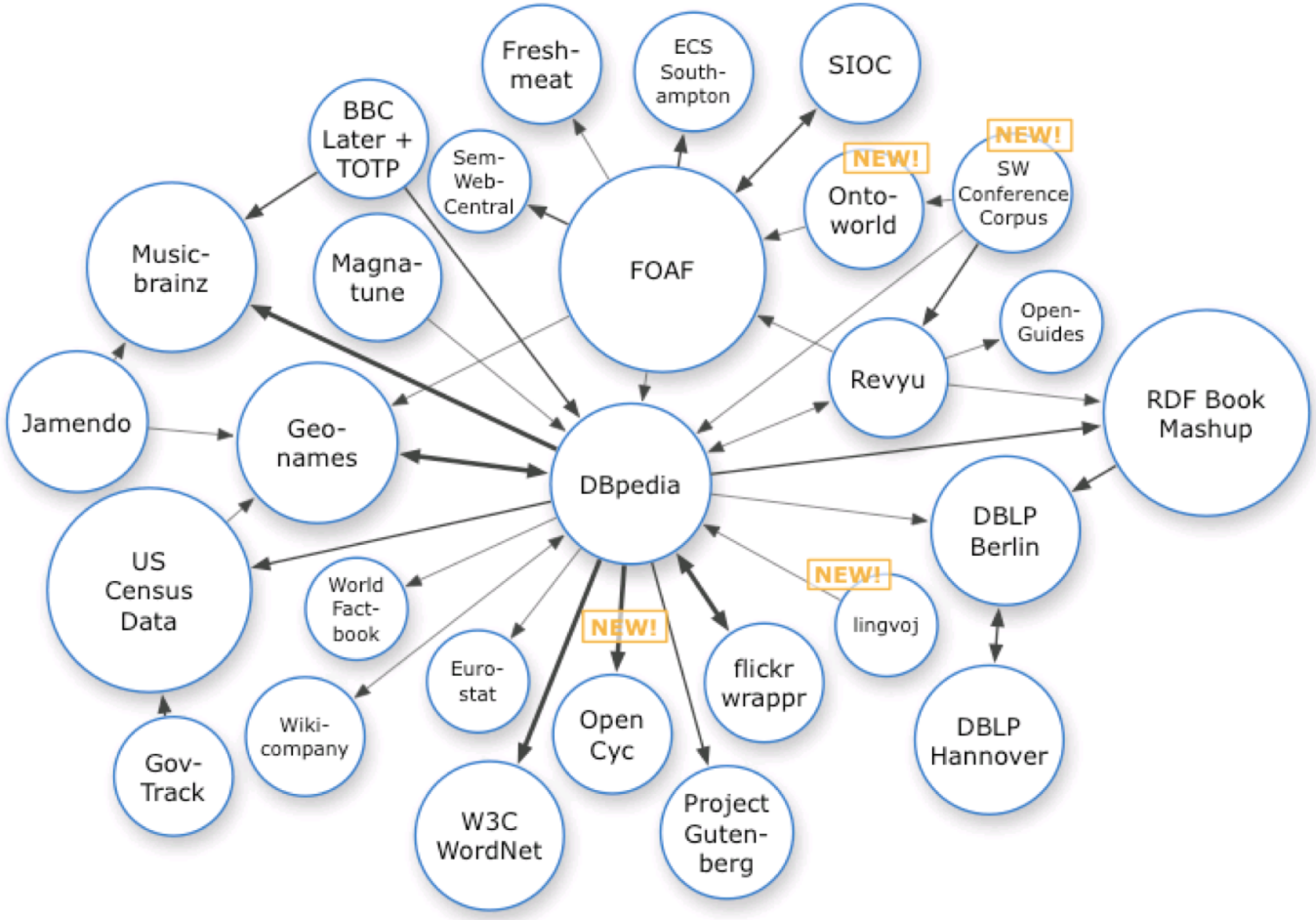
- Use URIs as names for things
- Use HTTP URIs so that people can look up those names.
- When someone looks up a URI, provide useful information, using the standards (RDF*, SPARQL)
- Include links to other URIs. so that they can discover more things.

Linked Open Data 2007

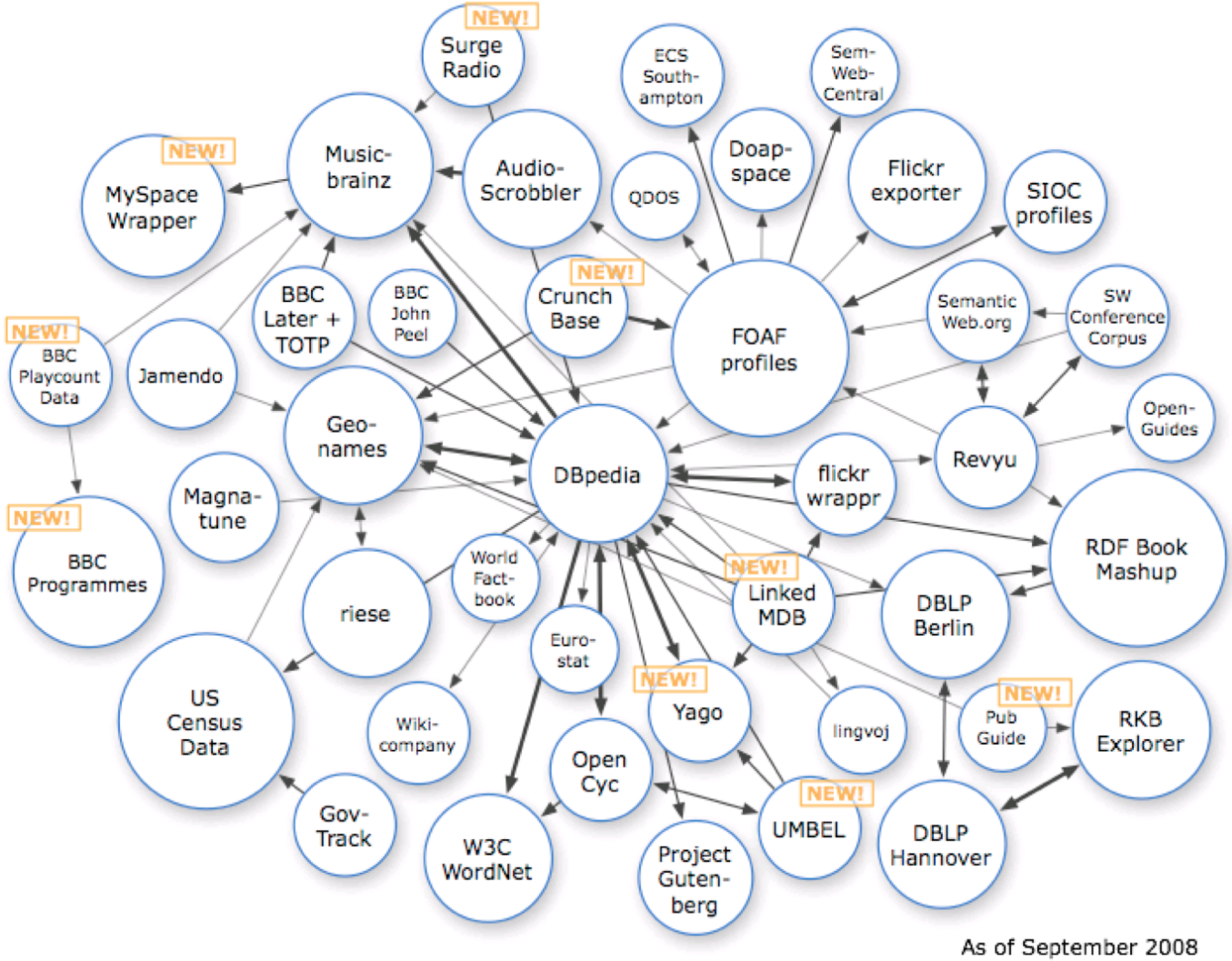


Linking Open Data cloud diagram, this and subsequent pages, by Richard Cyganiak and Anja Jentzsch. http://lod-cloud.net/

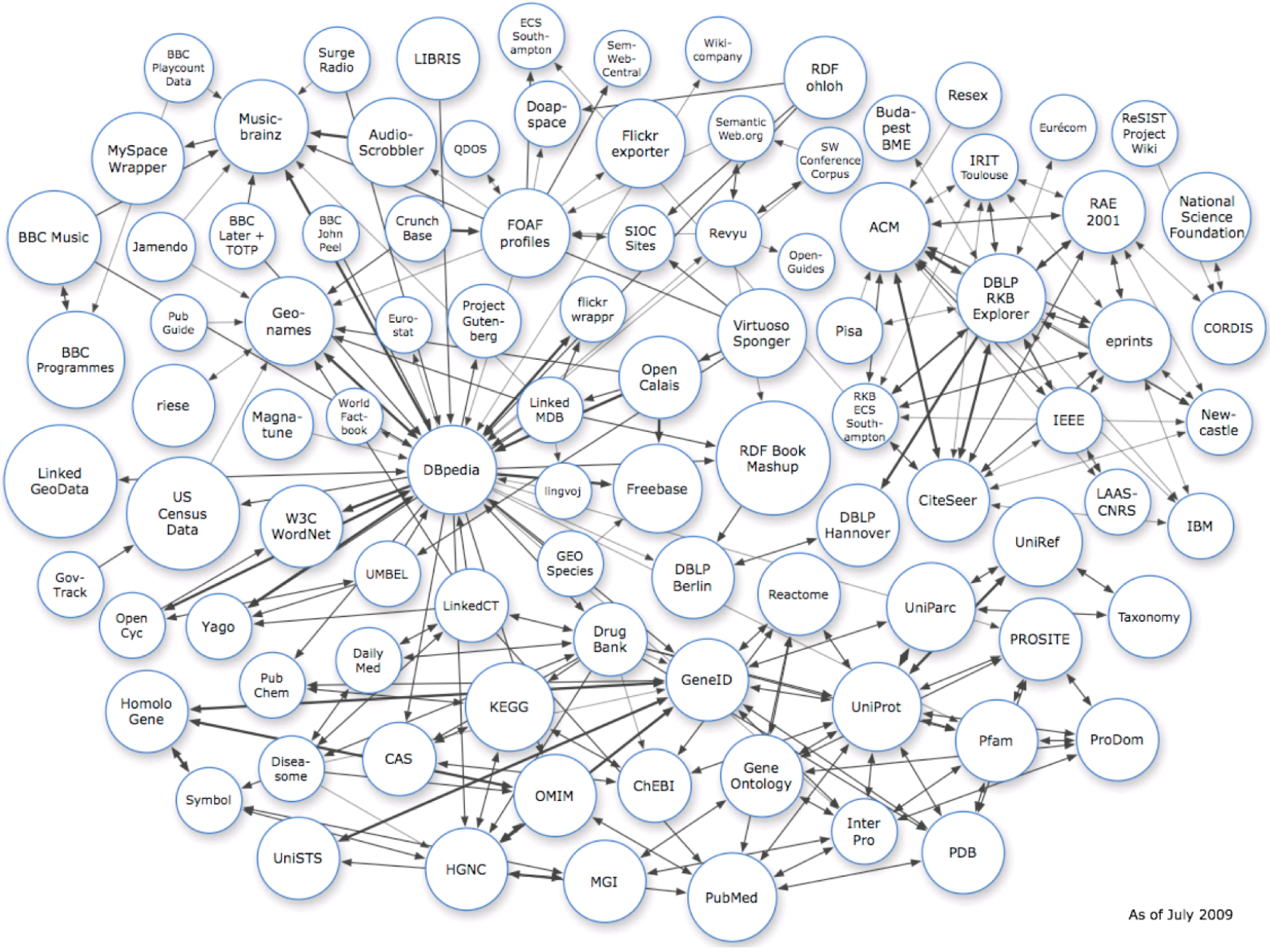
Linked Open Data 2007



Linked Open Data 2008

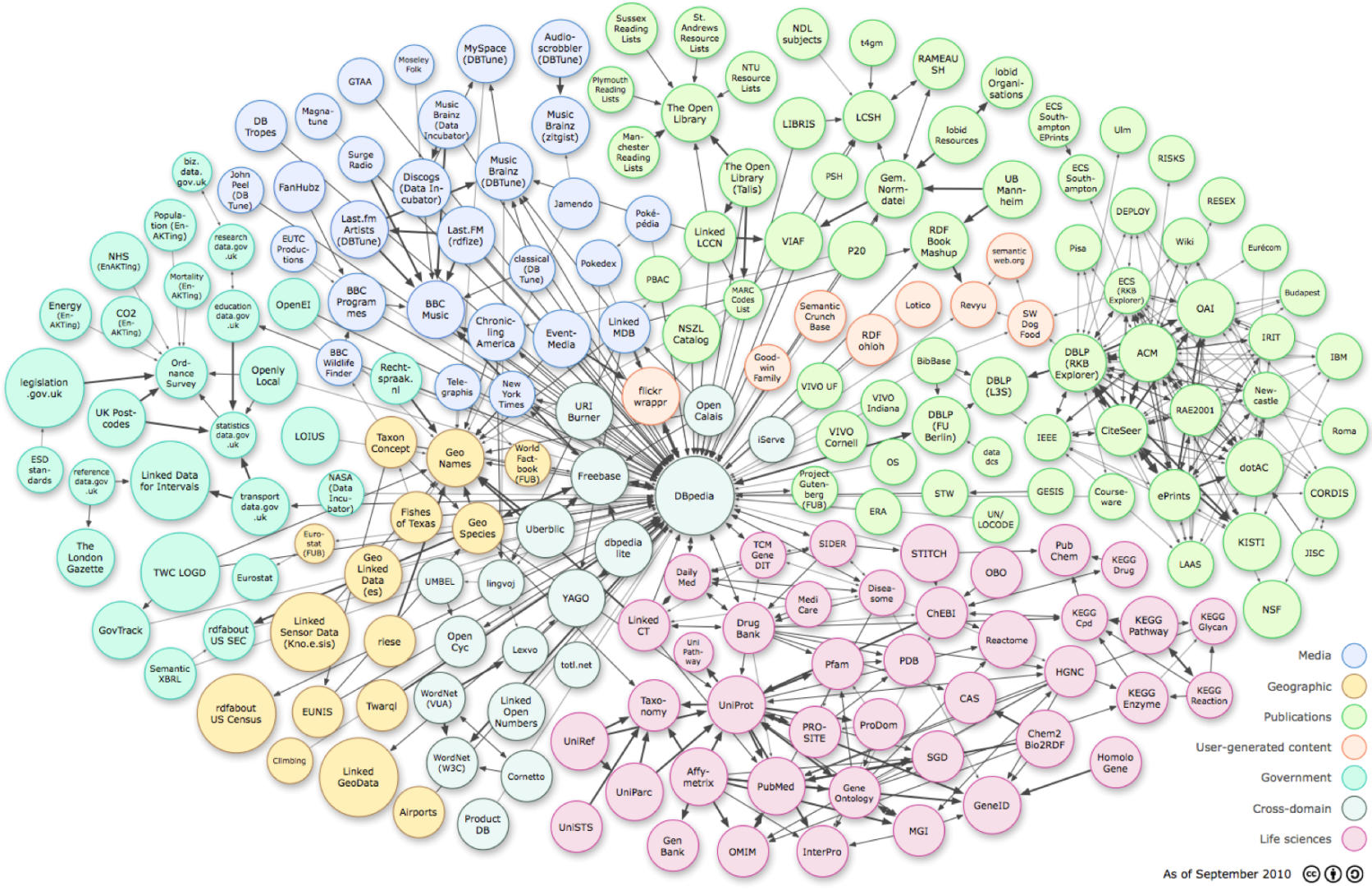


Linked Open Data 2009

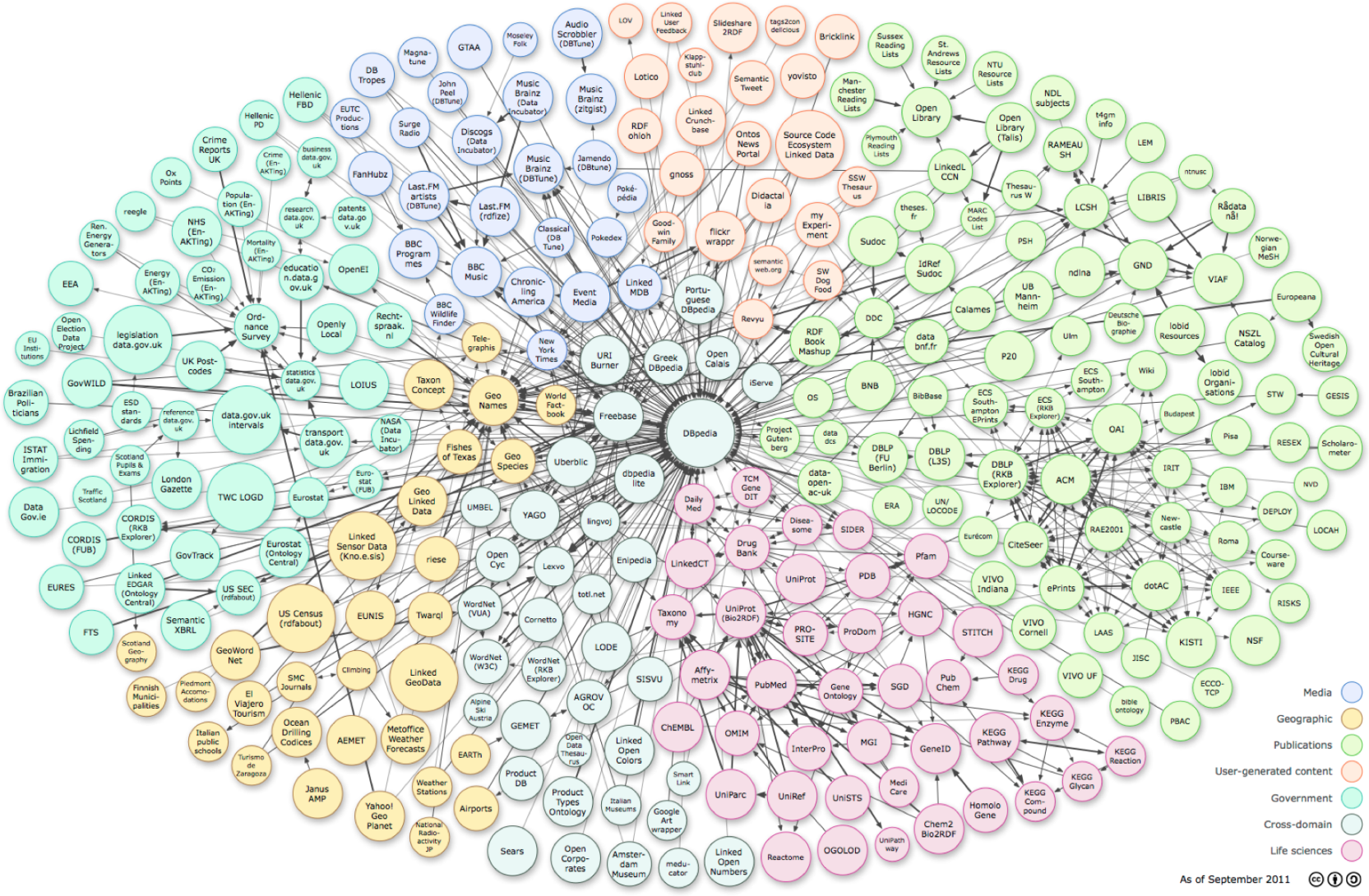


As of July 2009

Linked Open Data 2010



Linked Open Data 2011



Linked Open Data

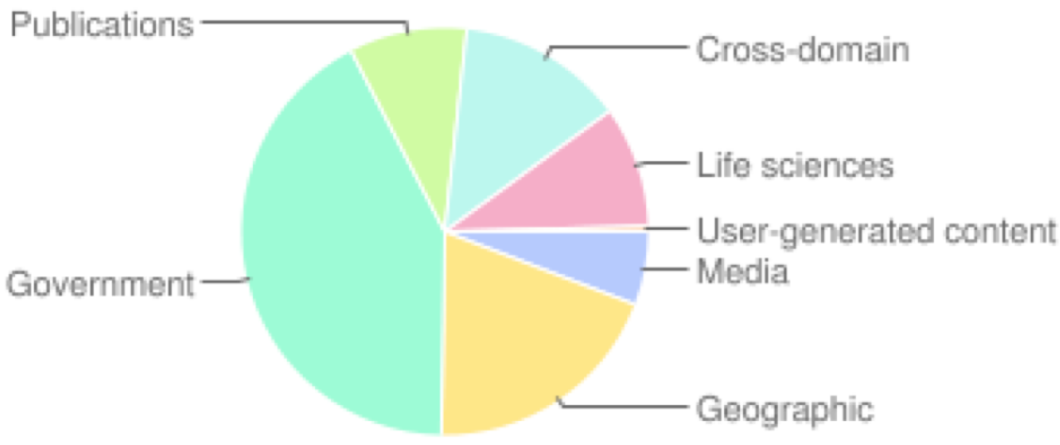
Number of Datasets

2011-09-19	**295**
2010-09-22	**203**
2009-07-14	**95**
2008-09-18	**45**
2007-10-08	**25**
2007-05-01	**12**

Number of triples (Sept 2011)

31,634,213,770

with 503,998,829 out-links



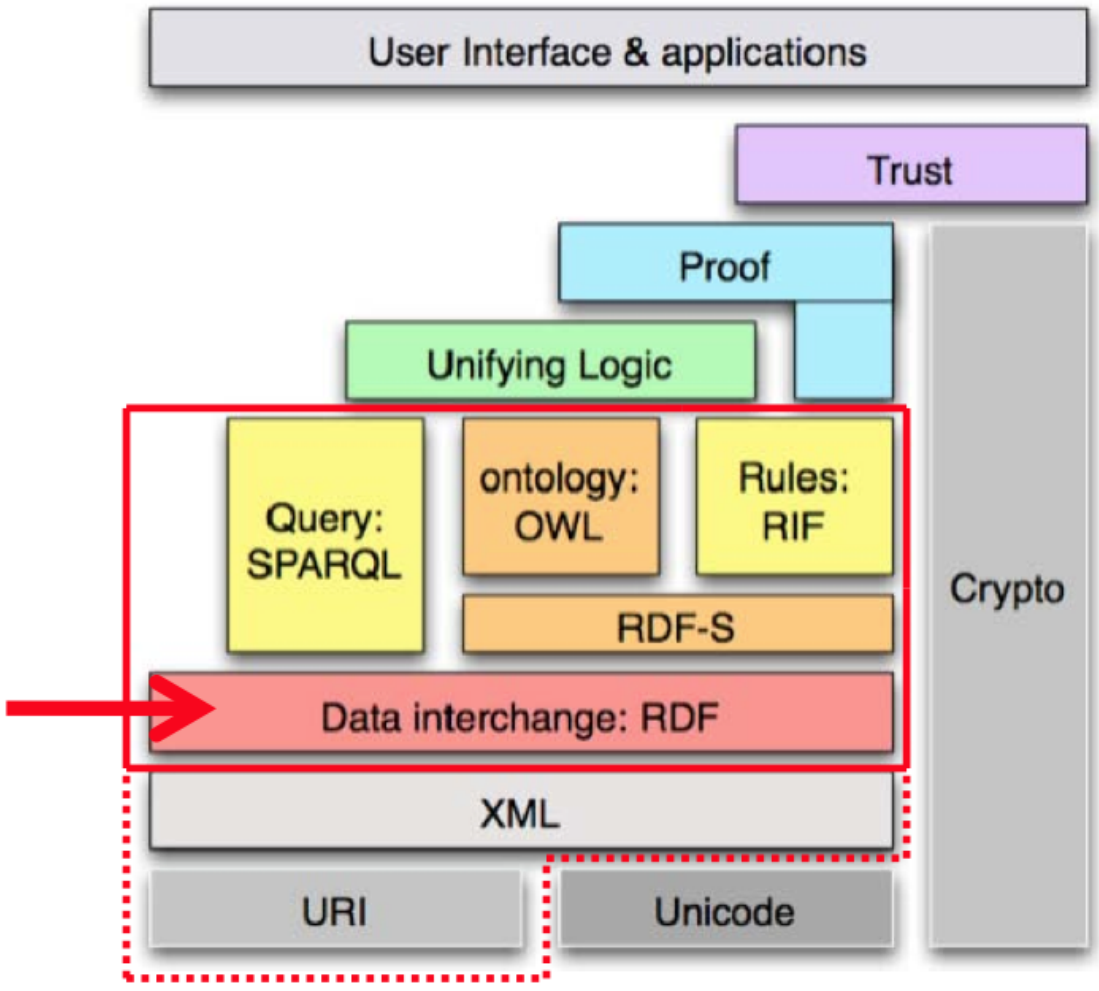
From http://www4.wiwiss.fu-berlin.de/lodcloud/state/

Linked Open Data

Populated Place Features (city, village,...)			
2,518,403	P.PPL	populated place	a city, town, village, or other agglomeration of buildings where people live and work
48,483	P.PPLX	section of populated place	
39,336	P.PPLL	populated locality	an area similar to a locality but with a small group of dwellings or other buildings
13,306	P.PPLQ	abandoned populated place	
2,684	P.PPLA4	seat of a fourth-order administrative division	
2,028	P.PPLA	seat of a first-order administrative division	seat of a first-order administrative division (PPLC takes precedence over PPLA)
1,847	P.PPLW	destroyed populated place	a village, town or city destroyed by a natural disaster, or by war
1,006	P.PPLF	farm village	a populated place where the population is largely engaged in agricultural activities
930	P.PPLA3	seat of a third-order administrative division	
695	P.PPLA2	seat of a second-order administrative division	
253	P.PPLS	populated places	cities, towns, villages, or other agglomerations of buildings where people live and work
249	P.STLMT	israeli settlement	
235	P.PPLC	capital of a political entity	
57	P.		
29	P.PPLR	religious populated place	a populated place whose population is largely engaged in religious occupations
6	P.PPLG	seat of government of a political entity	
2,629,547	Total for P		

rdfs:subClassOf?

Semantic Web Cake Layer



XML Problems I

- **How do you encode the piece of knowledge "The book FOST is published by CRC Press"**

- **<book>**
  <title>FOST</title>
  <publisher>CRC Press</publisher>
  </book>

- **<publisher>**
  <name>CRC Press</name>
  <book><title>FOST</title><book>
  </publisher>

- **etc.**

(From P. Hitzler, 2012)

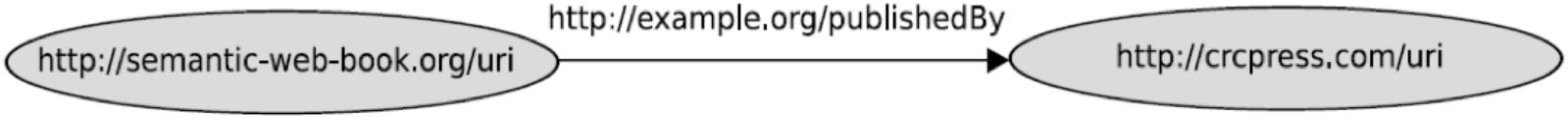
XML Problems II

- **Merging trees is rather cumbersome and the result isn't always clear.**
  - **<publisher>**
    <name>CRC Press</name>
    <book><title>FOST</title><book>
    </publisher>
  - **<book>**
    <title>Semantic Web</title>
    <publisher>Springer</publisher>
    </book>

(From P. Hitzler, 2012)

RDF

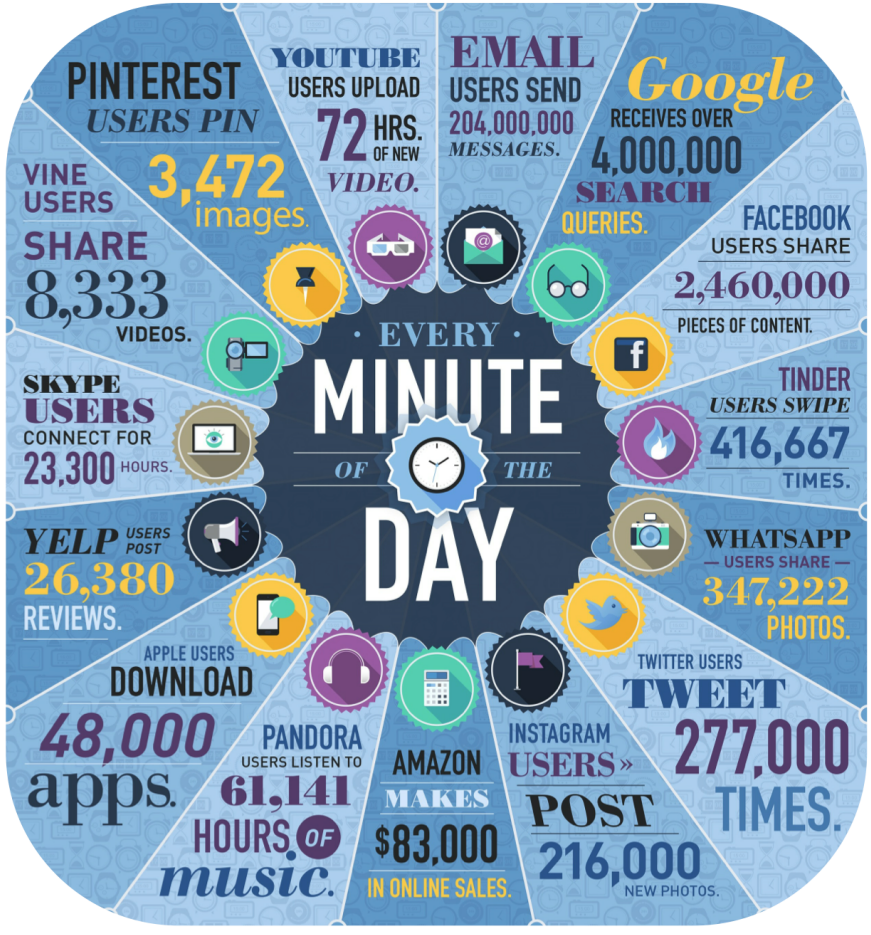
- **Use (directed) graphs as data model**



The data model can be recorded (for sharing) in XML (as it is indeed).

(From P. Hitzler, 2012)

We are Living in the Era of Big Data

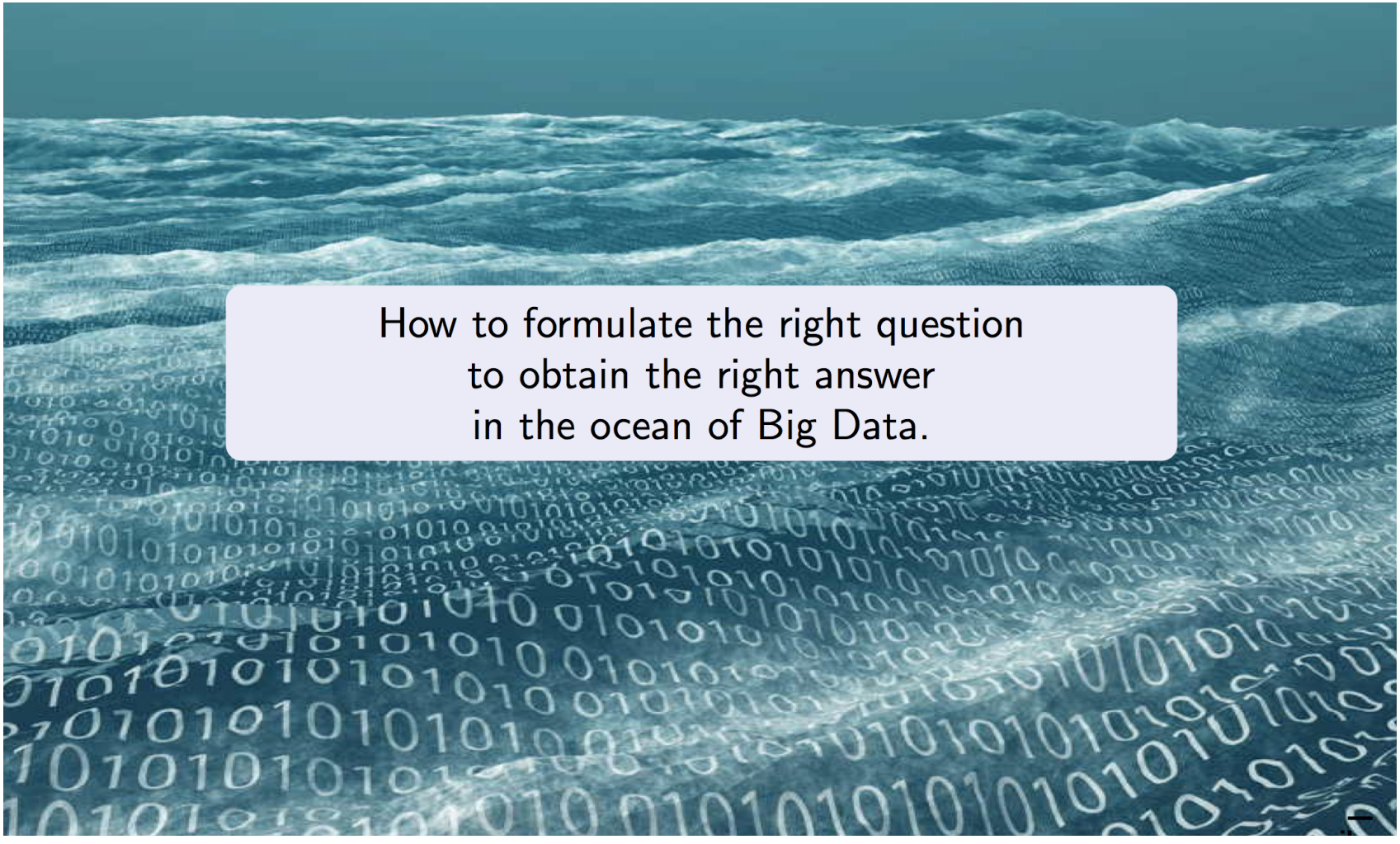


Data Never Sleeps 2.0

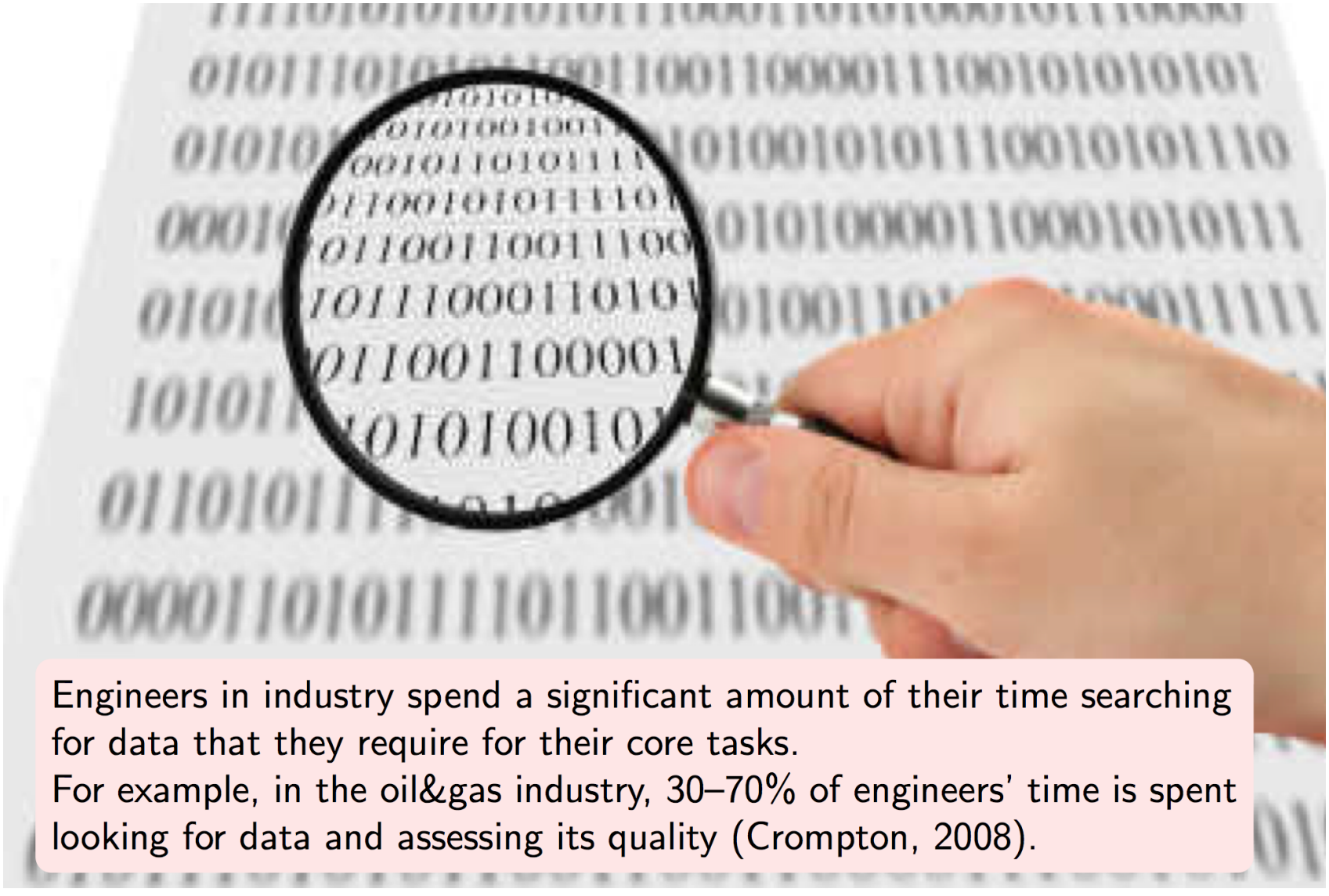
(The following pages from G. Xiao, 2015)

The Problem : information access

The Problem : information access



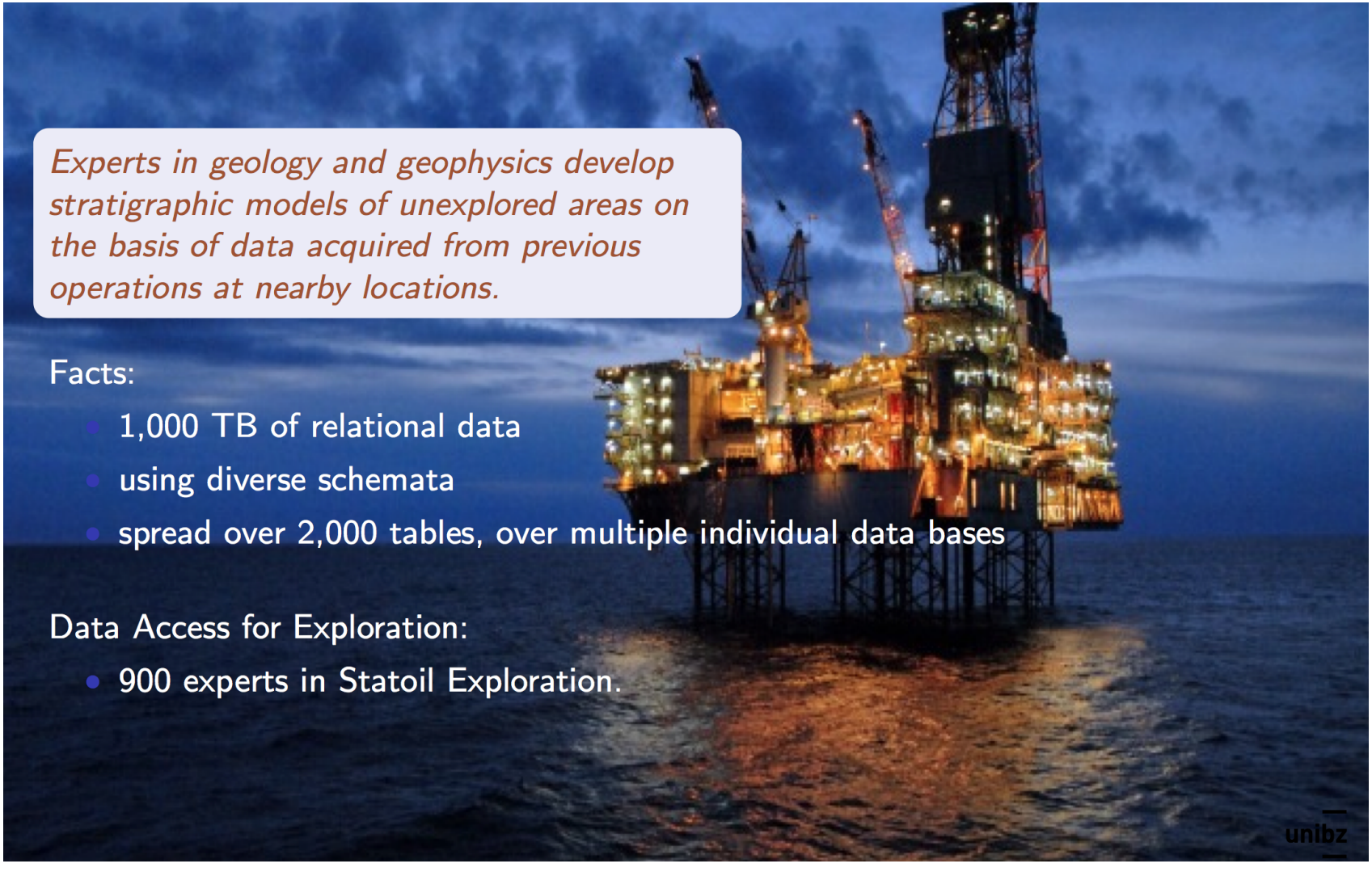
How much time is spent searching for data ?



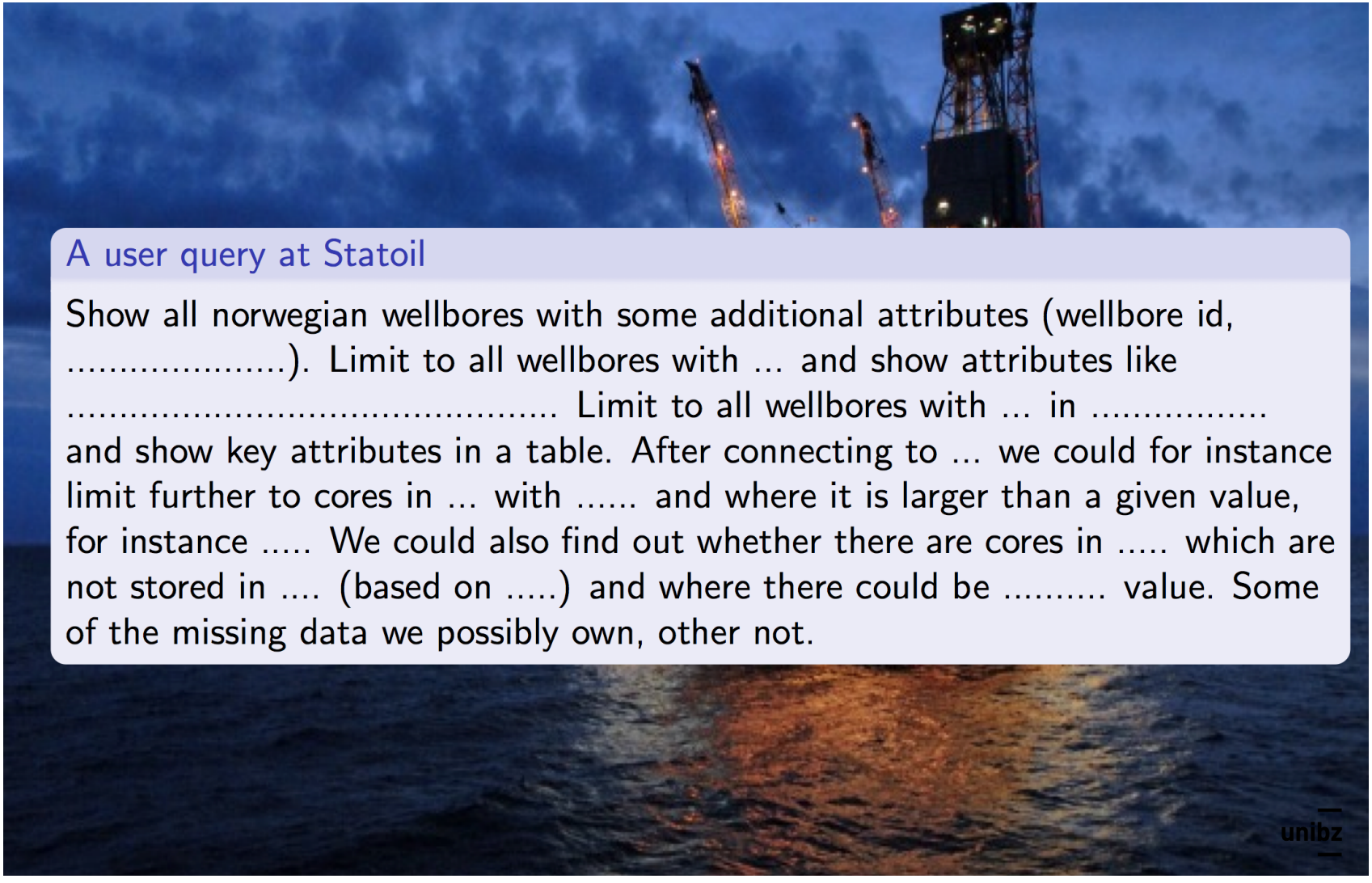
Engineers in industry spend a significant amount of their time searching for data that they require for their core tasks.
For example, in the oil&gas industry, 30–70% of engineers' time is spent looking for data and assessing its quality (Crompton, 2008).

Example : Statoil Exploration

How much time/money is spent searching for data ?



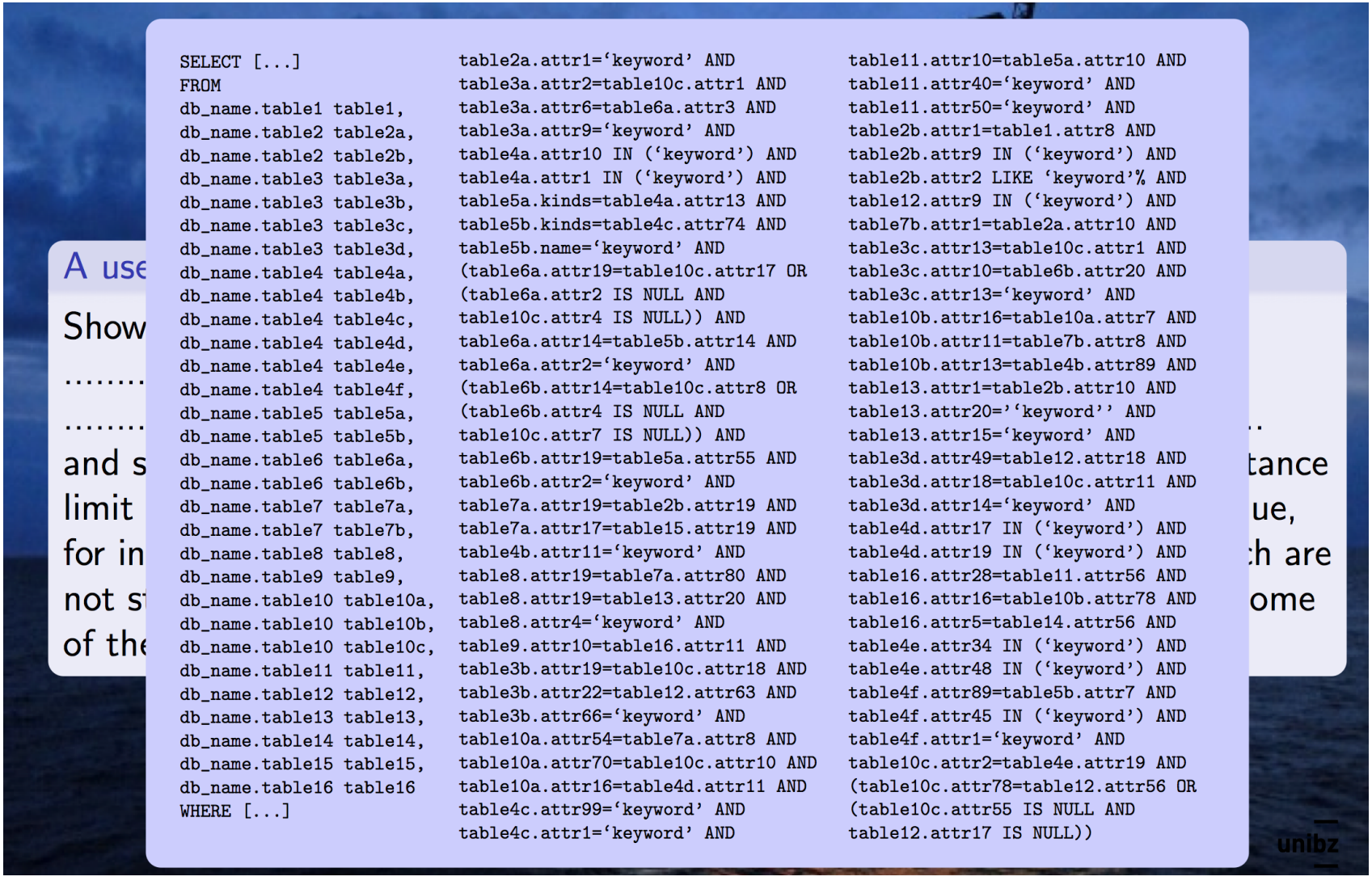
How much time/money is spent searching for data ?



A user query at Statoil

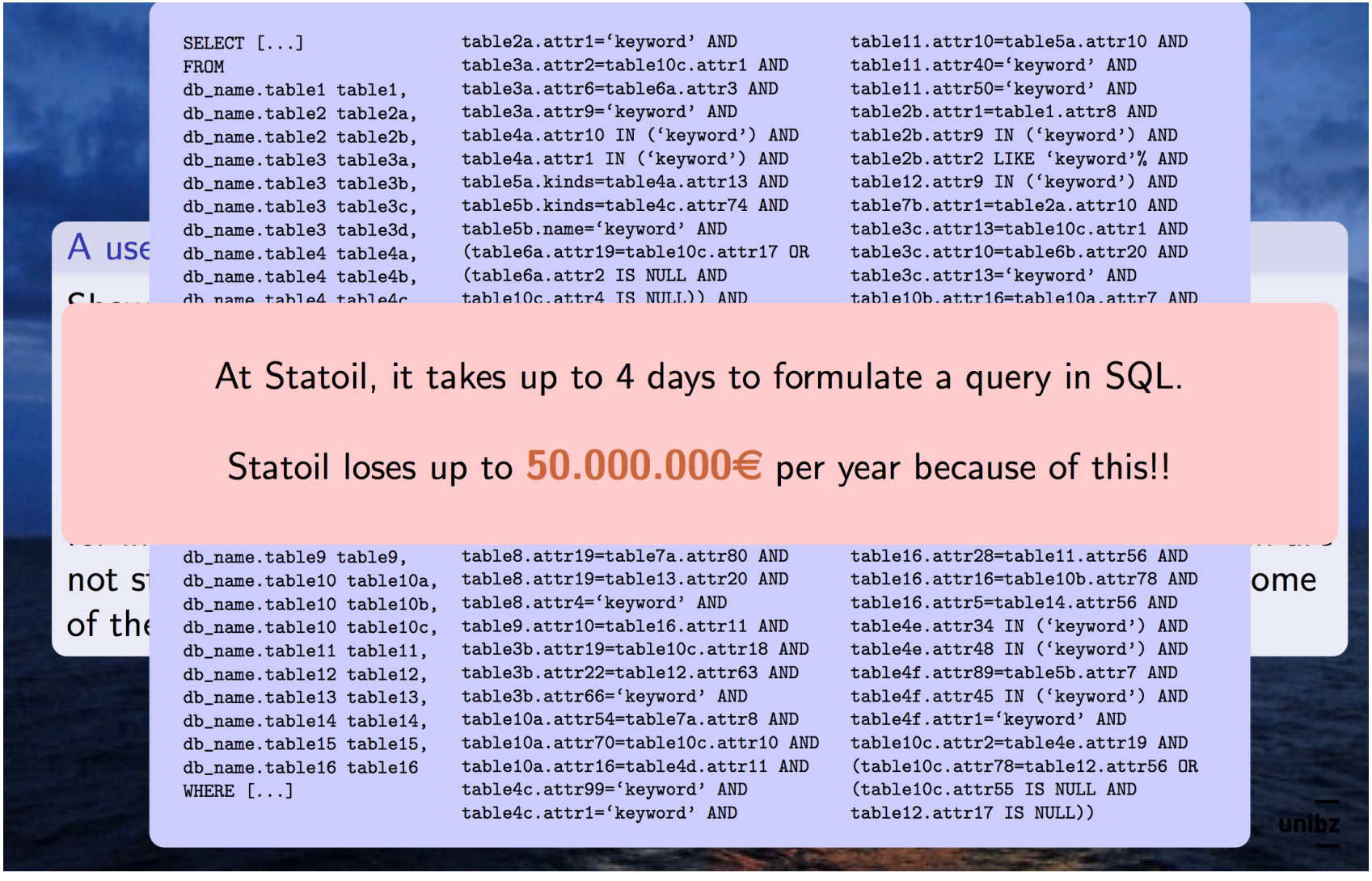
Show all norwegian wellbores with some additional attributes (wellbore id, ...). Limit to all wellbores with ... and show attributes like ..... Limit to all wellbores with ... in ..... and show key attributes in a table. After connecting to ... we could for instance limit further to cores in ... with ..... and where it is larger than a given value, for instance ..... We could also find out whether there are cores in ..... which are not stored in .... (based on .....) and where there could be ..... value. Some of the missing data we possibly own, other not.

How much time/money is spent searching for data ?



```
SELECT [...]
FROM
db_name.table1 table1,
db_name.table2 table2a,
db_name.table2 table2b,
db_name.table3 table3a,
db_name.table3 table3b,
db_name.table3 table3c,
db_name.table3 table3d,
db_name.table4 table4a,
db_name.table4 table4b,
db_name.table4 table4c,
db_name.table4 table4d,
db_name.table4 table4e,
db_name.table4 table4f,
db_name.table5 table5a,
db_name.table5 table5b,
db_name.table6 table6a,
db_name.table6 table6b,
db_name.table7 table7a,
db_name.table7 table7b,
db_name.table8 table8,
db_name.table9 table9,
db_name.table10 table10a,
db_name.table10 table10b,
db_name.table10 table10c,
db_name.table11 table11,
db_name.table12 table12,
db_name.table13 table13,
db_name.table14 table14,
db_name.table15 table15,
db_name.table16 table16
WHERE [...]
table2a.attr1='keyword' AND
table3a.attr2=table10c.attr1 AND
table3a.attr6=table6a.attr3 AND
table3a.attr9='keyword' AND
table4a.attr10 IN ('keyword') AND
table4a.attr1 IN ('keyword') AND
table5a.kinds=table4a.attr13 AND
table5b.kinds=table4c.attr74 AND
table5b.name='keyword' AND
(table6a.attr19=table10c.attr17 OR
(table6a.attr2 IS NULL AND
table10c.attr4 IS NULL)) AND
table6a.attr14=table5b.attr14 AND
table6a.attr2='keyword' AND
(table6b.attr14=table10c.attr8 OR
(table6b.attr4 IS NULL AND
table10c.attr7 IS NULL)) AND
table6b.attr19=table5a.attr55 AND
table6b.attr2='keyword' AND
table7a.attr19=table2b.attr19 AND
table7a.attr17=table15.attr19 AND
table4b.attr11='keyword' AND
table8.attr19=table7a.attr80 AND
table8.attr19=table13.attr20 AND
table8.attr4='keyword' AND
table9.attr10=table16.attr11 AND
table3b.attr19=table10c.attr18 AND
table3b.attr22=table12.attr63 AND
table3b.attr66='keyword' AND
table10a.attr54=table7a.attr8 AND
table10a.attr70=table10c.attr10 AND
table10a.attr16=table4d.attr11 AND
table4c.attr99='keyword' AND
table4c.attr1='keyword' AND
table11.attr10=table5a.attr10 AND
table11.attr40='keyword' AND
table11.attr50='keyword' AND
table2b.attr1=table1.attr8 AND
table2b.attr9 IN ('keyword') AND
table2b.attr2 LIKE 'keyword'% AND
table12.attr9 IN ('keyword') AND
table7b.attr1=table2a.attr10 AND
table3c.attr13=table10c.attr1 AND
table3c.attr10=table6b.attr20 AND
table3c.attr13='keyword' AND
table10b.attr16=table10a.attr7 AND
table10b.attr11=table7b.attr8 AND
table10b.attr13=table4b.attr89 AND
table13.attr1=table2b.attr10 AND
table13.attr20=''keyword'' AND
table13.attr15='keyword' AND
table3d.attr49=table12.attr18 AND
table3d.attr18=table10c.attr11 AND
table3d.attr14='keyword' AND
table4d.attr17 IN ('keyword') AND
table4d.attr19 IN ('keyword') AND
table16.attr28=table11.attr56 AND
table16.attr16=table10b.attr78 AND
table16.attr5=table14.attr56 AND
table4e.attr34 IN ('keyword') AND
table4e.attr48 IN ('keyword') AND
table4f.attr89=table5b.attr7 AND
table4f.attr45 IN ('keyword') AND
table4f.attr1='keyword' AND
table10c.attr2=table4e.attr19 AND
(table10c.attr78=table12.attr56 OR
(table10c.attr55 IS NULL AND
table12.attr17 IS NULL))
```

Challenges Accessing Big Data



```
SELECT [...]
FROM
db_name.table1 table1,
db_name.table2 table2a,
db_name.table2 table2b,
db_name.table3 table3a,
db_name.table3 table3b,
db_name.table3 table3c,
db_name.table3 table3d,
db_name.table4 table4a,
db_name.table4 table4b,
db_name.table4 table4c,
db_name.table9 table9,
db_name.table10 table10a,
db_name.table10 table10b,
db_name.table10 table10c,
db_name.table11 table11,
db_name.table12 table12,
db_name.table13 table13,
db_name.table14 table14,
db_name.table15 table15,
db_name.table16 table16
WHERE [...]
```

```
table2a.attr1='keyword' AND
table3a.attr2=table10c.attr1 AND
table3a.attr6=table6a.attr3 AND
table3a.attr9='keyword' AND
table4a.attr10 IN ('keyword') AND
table4a.attr1 IN ('keyword') AND
table5a.kinds=table4a.attr13 AND
table5b.kinds=table4c.attr74 AND
table5b.name='keyword' AND
(table6a.attr19=table10c.attr17 OR
(table6a.attr2 IS NULL AND
table10c.attr4 IS NULL)) AND
table8.attr19=table7a.attr80 AND
table8.attr19=table13.attr20 AND
table8.attr4='keyword' AND
table9.attr10=table16.attr11 AND
table3b.attr19=table10c.attr18 AND
table3b.attr22=table12.attr63 AND
table3b.attr66='keyword' AND
table10a.attr54=table7a.attr8 AND
table10a.attr70=table10c.attr10 AND
table10a.attr16=table4d.attr11 AND
table4c.attr99='keyword' AND
table4c.attr1='keyword' AND
```

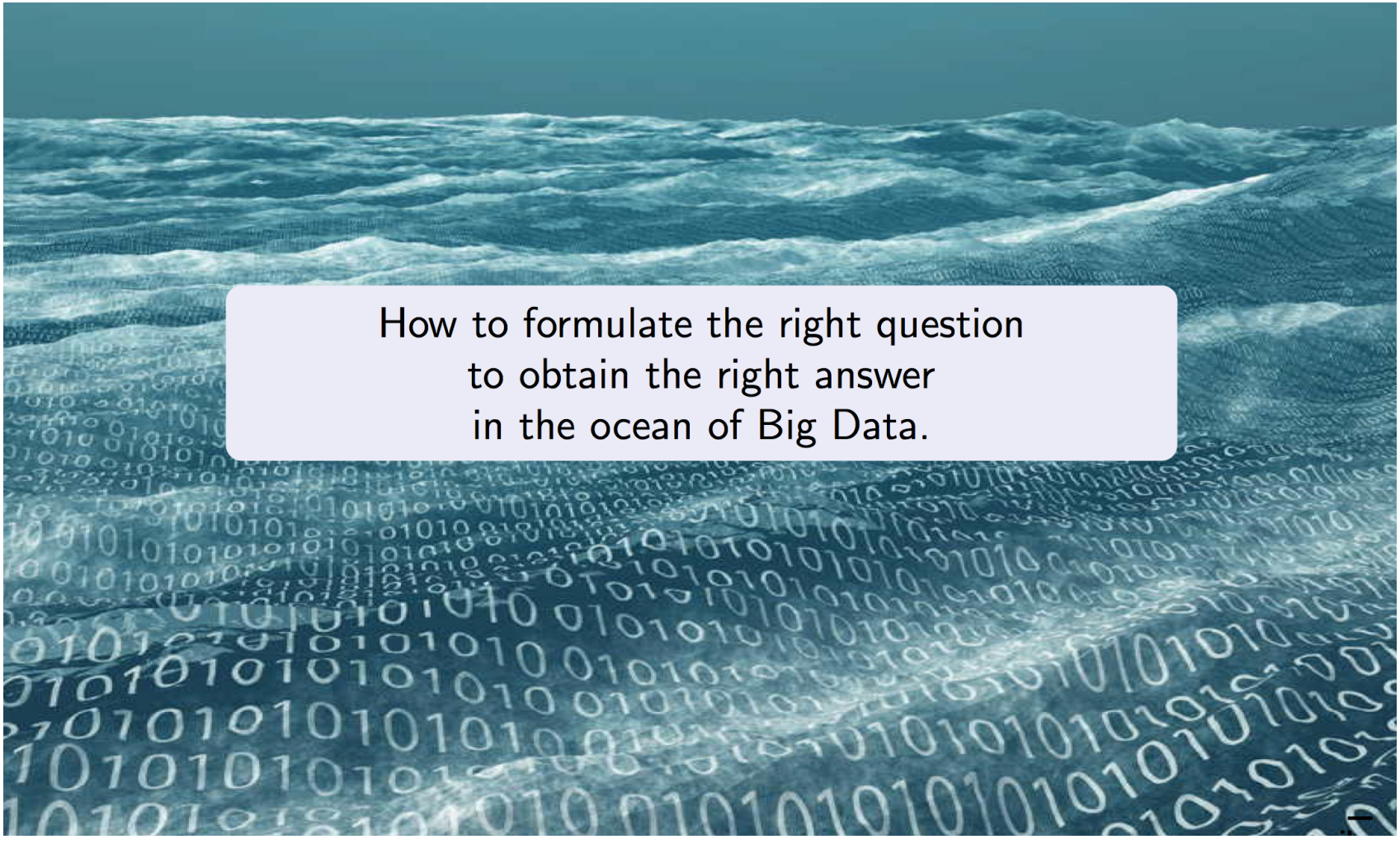
```
table11.attr10=table5a.attr10 AND
table11.attr40='keyword' AND
table11.attr50='keyword' AND
table2b.attr1=table1.attr8 AND
table2b.attr9 IN ('keyword') AND
table2b.attr2 LIKE 'keyword'% AND
table12.attr9 IN ('keyword') AND
table7b.attr1=table2a.attr10 AND
table3c.attr13=table10c.attr1 AND
table3c.attr10=table6b.attr20 AND
table3c.attr13='keyword' AND
table10b.attr16=table10a.attr7 AND
table16.attr28=table11.attr56 AND
table16.attr16=table10b.attr78 AND
table16.attr5=table14.attr56 AND
table4e.attr34 IN ('keyword') AND
table4e.attr48 IN ('keyword') AND
table4f.attr89=table5b.attr7 AND
table4f.attr45 IN ('keyword') AND
table4f.attr1='keyword' AND
table10c.attr2=table4e.attr19 AND
(table10c.attr78=table12.attr56 OR
(table10c.attr55 IS NULL AND
table12.attr17 IS NULL))
```

At Statoil, it takes up to 4 days to formulate a query in SQL.

Statoil loses up to **50.000.000€** per year because of this!!

unibz

The Problem : information access



Need for Abstraction

We need to facilitate access to Data

- by abstracting away from how the data is stored, and
- by making use of high level views on the data, so called ontologies.

We will see the theoretical foundations of ontologies in today's course (and some related W3C standards if time permits)