

# The controversies within Open Data publishing

Pieter Colpaert  
@pietercolpaert



# I am

Researching Linked Open Data publishing at



I've co-founded

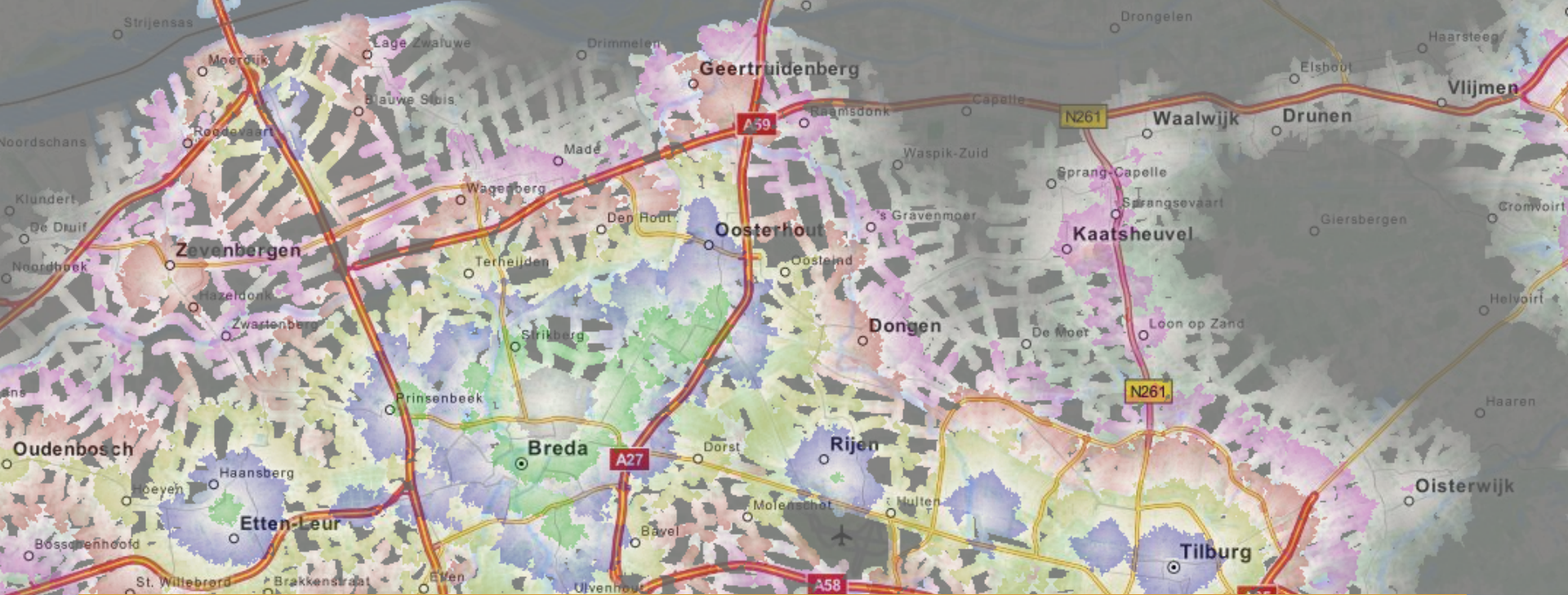


and I've started



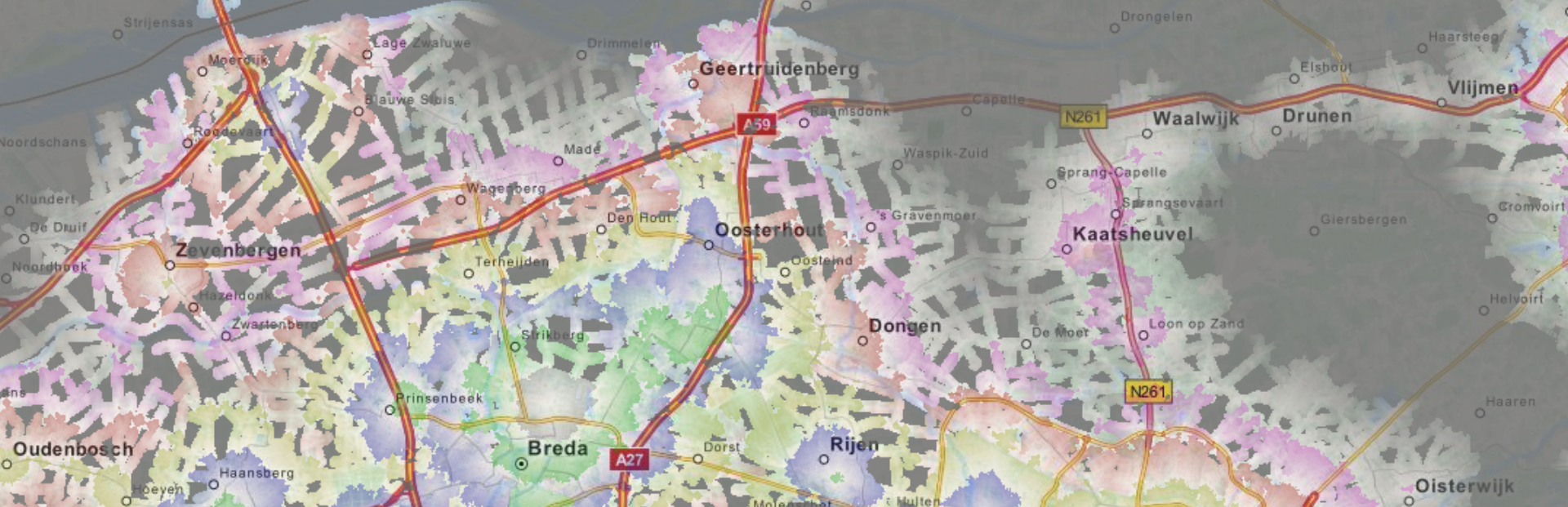
More info: <http://pieter.pm>

@pietercolpaert



Personal goal: journey planning over the Web of data

Open Trip Planner  
with data by @openov



For instance, the Web should be able to answer:

**What's the route to work  
with the highest chance to see a waterbird?**

Open Trip Planner  
with data by @openov



**peterdesmet** on Mar 12 Migrate files from existing dataset repos

1 contributor

1378.167 kb



Raw

History



It's open!

<https://github.com/LifeWatchINBO/data-publication/tree/master/datasets/watervogels-occurrences>

@pietercolpaert

# The Open Definition

---

The **Open Definition** sets out principles that define “openness” in relation to **data and content**.

It makes **precise** the meaning of “open” in the terms “**open data**” and “**open content**” and thereby ensures **quality** and encourages **compatibility** between different pools of open material.

It can be summed up in the statement that:

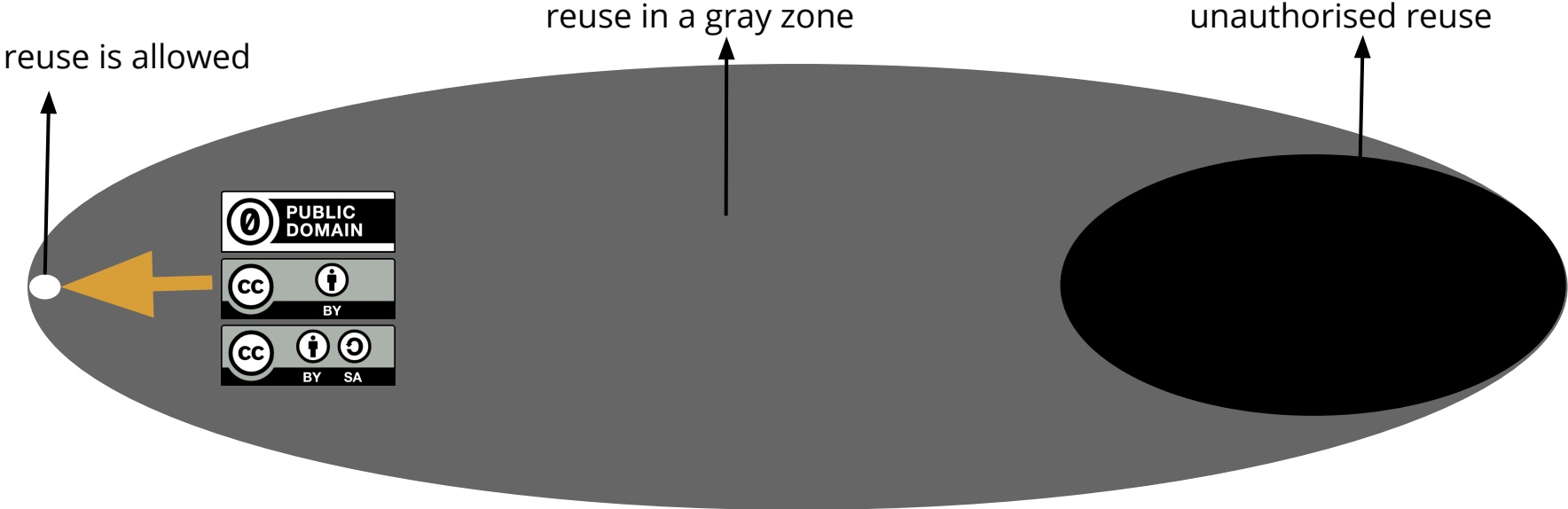
*“Open means **anyone** can **freely access, use, modify, and share** for **any purpose** (subject, at most, to requirements that preserve provenance and openness).”*

Put most succinctly:

OpenDefinition.org

*“Open data and content can be **freely used, modified, and shared** by **anyone** for **any purpose**”*

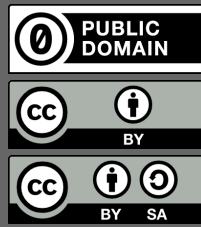
# Documents on the web



reuse is allowed

reuse in a gray zone

unauthorised reuse




# Open Data

Because public data increases its value  
when it gets reused



# 3 controversies

1. Linked Data
2. Maximise reuse, or functionalities?
  3. Why? Where's the money?

A wide-angle shot of a TED talk stage. A large blue screen at the top center displays the text 'Raw Data Now!' in white. Below the screen, the red 'TED' logo is visible on the left. A speaker in a dark shirt and pants stands in the center of the stage, gesturing with his arms. The stage is lit with blue light, and the background features abstract, glowing patterns. The audience is visible in the foreground, seated and facing the stage.

Raw  
Data  
Now!

# 1. Linked Data

# Web of Documents

## Chordate

From Wikipedia, the free encyclopedia

hypertext

Designed for humans

*This article is about the animal phylum. For the tear shape, see [Cordate](#).*

**Chordates** (/ˈkɔːrdzɪts/) are **animals** possessing a **notochord**, a hollow **dorsal nerve cord**, **pharyngeal slits**, an **endostyle**, and a post-anal **tail** for at least some period of their life cycles. Taxonomically, the phylum includes the subphyla **Vertebrata**, including **mammals**, **fish**, **amphibians**, **reptiles**, **birds**; **Tunicata**, including **salps** and **sea squirts**; and **Cephalochordata**, comprising the **lancelets**. Members of the phylum Chordata are **bilaterally symmetric**, **deuterostome coelomates**, and the vertebrate chordates display **segmentation**.

The phylum **Hemichordata** including the **acorn worms** has been presented as a fourth chordate subphylum, but it now is usually treated as a separate phylum. It, along with the phylum **Echinodermata**, including **starfish**, **sea urchins**, and **sea cucumbers** and their kin, are the chordates' closest relatives. Primitive chordates are known from at least as early as the **Cambrian explosion**.

Of the more than 65,000 living species of chordates, about half are bony fish of the class **Osteichthyes**. The world's largest and fastest animals, the **blue whale** and **peregrine falcon** respectively, are chordates, as are **humans**.<sup>[3]</sup>

### Contents [hide]

- Overview of affinities
  - Origin of name
- Definition
- Subdivisions
  - Craniata
  - Tunicata (tunicates, or urochordates)
  - Cephalochordata: lancelets
- Origins
- Classification
  - Taxonomy
  - Phylogeny
- Closest nonchordate relatives
  - Hemichordates
  - Echinoderms
- See also

Reusable by anyone



### Chordata

Temporal range: **Terreneuvian – Present, 540–0Ma**

PreC C O S D C P T J K Pg N



The X-ray tetra (*Pristella maxillaris*) is one of the few chordates with a visible **backbone**. The **spinal cord** is housed within its backbone.

#### Scientific classification

Kingdom: **Animalia**

Superphylum: **Deuterostomia**

Phylum: **Chordata**

Haeckel, 1874<sup>[1][2]</sup>

@pietercolpaert

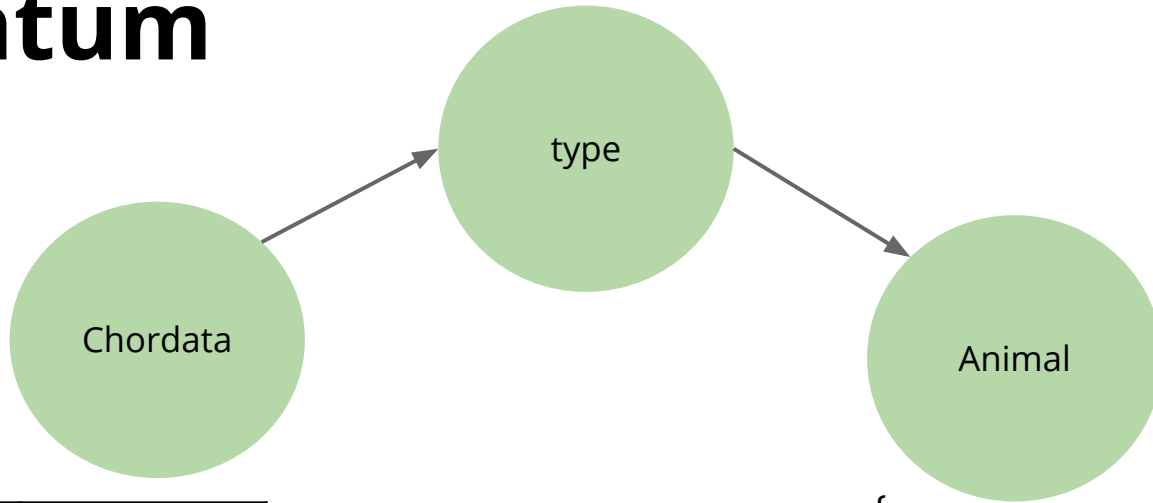
# Web of Documents



*... it was mainly hard for people to imagine that this hypertext system would scale up to where it is today*

*It's the community that made it happen:  
→ we're all doing our part*

# A datum



|          |             |
|----------|-------------|
|          | <b>type</b> |
| Chordata | Animal      |

```
<Chordata>  
  <type>Animal</type>  
</Chordata>
```

```
{  
  "Chordata": {  
    "@type": "Animal"  
  }  
}
```

# Linked Data

Uses global identifiers (URIs)

Data can still reside in any form

Chordata

|          | <b>type</b> |
|----------|-------------|
| Chordata | Animal      |

```
<Chordata>  
  <type>Animal</type>  
</Chordata>
```

Provide a mapping:

Chordata → <http://dbpedia.org/resource/Chordate>

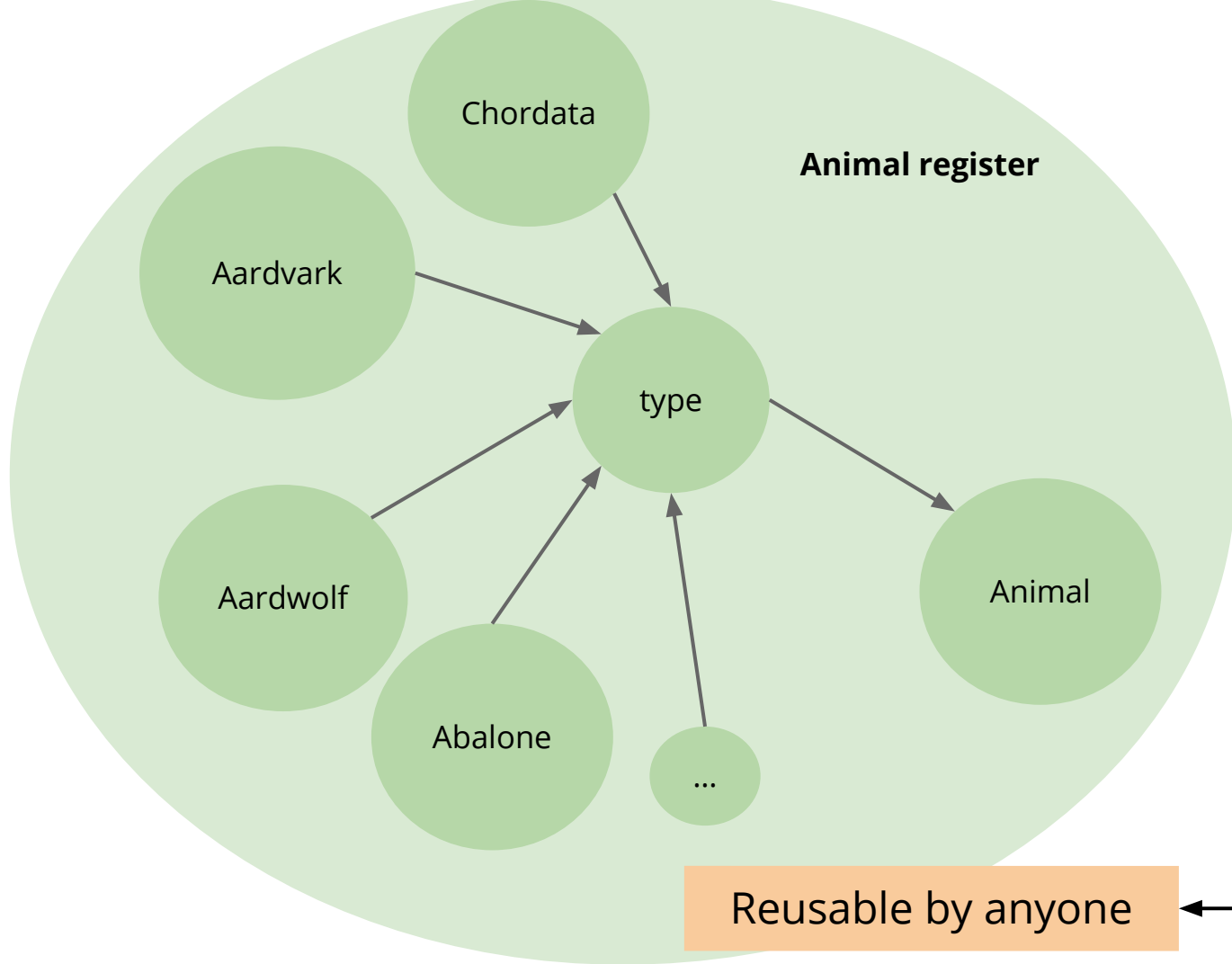
type → <http://www.w3.org/1999/02/22-rdf-syntax-ns#type>

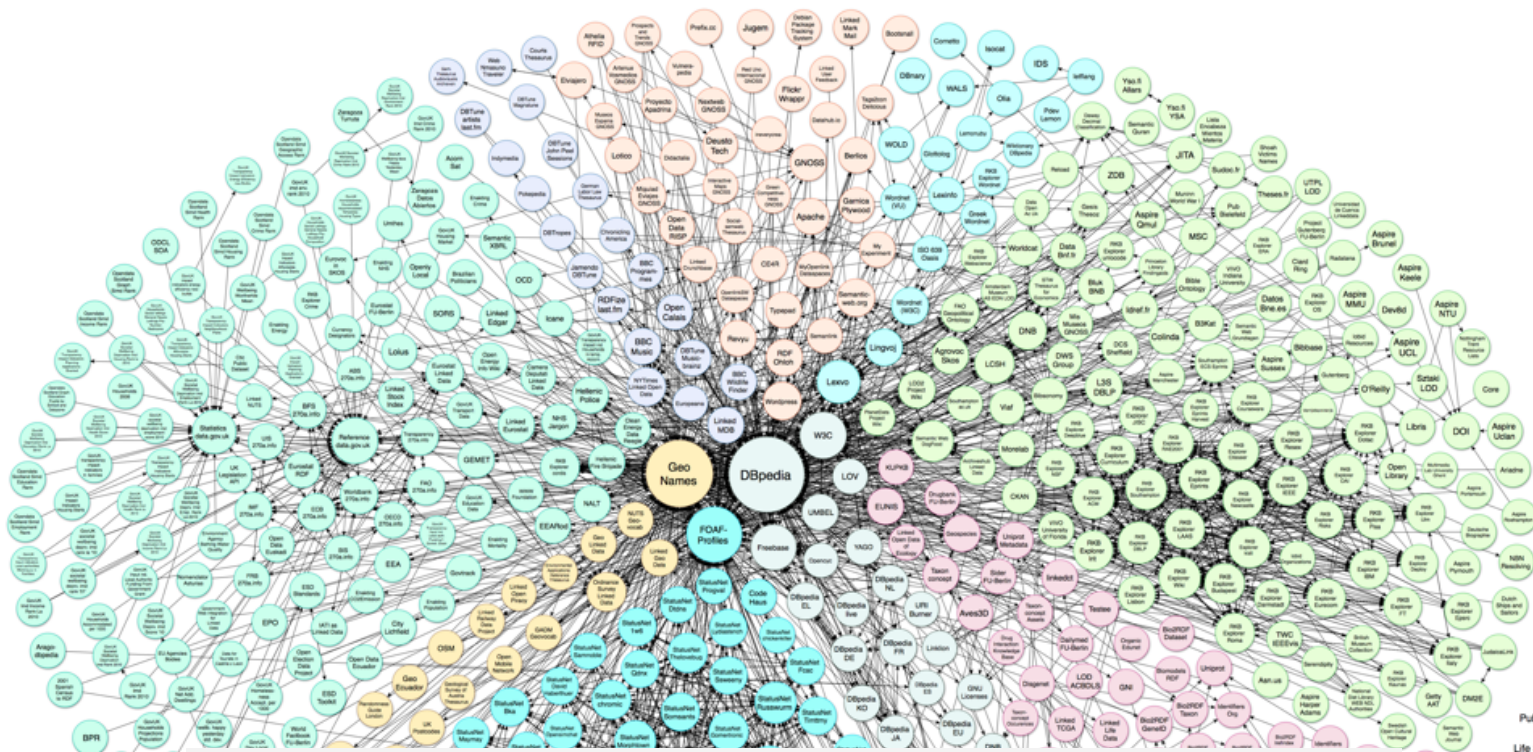
Animal → <http://umbel.org/umbel/rc/Animal>

Animal

```
{  
  "Chordata" : {  
    "@type" : "Animal"  
  }  
}
```

Data can still reside in any form





And if a lot more links and data are added, we get  
The Linked Open Data cloud





# DBpedia – Linked Data Fragments



## DBpedia 2014

Query DBpedia 2014 by triple pattern

subject: \_\_\_\_\_  
predicate: <http://dbpedia.org/ontology/phylum>  
object: <http://dbpedia.org/resource/Chordata>

Find matching triples

Matches in DBpedia 2014 for { ?s <<http://dbpedia.org/ontology/phylum>> <<http://dbpedia.org/re...>

Showing triples 1 to 100 of ±50,737 with 100 triples per page. [next](#)

%3F0ryzomys\_pliocaenicus phylum Chordate.  
%3F\_Nycticebus\_linglom phylum Chordate.  
%C5%BBubro%C5%84 phylum Chordate.  
%CA%BB%C4%80maui phylum Chordate.  
%CA%BB%C5%8C%CA%BB%C5%AB phylum Chordate.  
%CA%BBakepa phylum Chordate.  
%CA%BBakiapola%CA%BBau phylum Chordate.  
%CA%BBakikiki phylum Chordate.  
%CA%BBakohekohe phylum Chordate.  
%CA%BBAnianiau phylum Chordate.  
%CA%BBapapane phylum Chordate.  
%CA%BBi%CA%BBiwi phylum Chordate.  
%CA%BBUla-%CA%BBai-Hawane phylum Chordate.  
%E2%80%98%C5%8Cma%E2%80%99o phylum Chordate.  
%E2%80%98Akeke%E2%80%98e phylum Chordate.  
%E2%80%98Elepai phylum Chordate.  
%E2%80%98Eua\_Rail phylum Chordate.  
10-spine\_grouper phylum Chordate.  
AEECL's\_sportive\_lemur phylum Chordate.

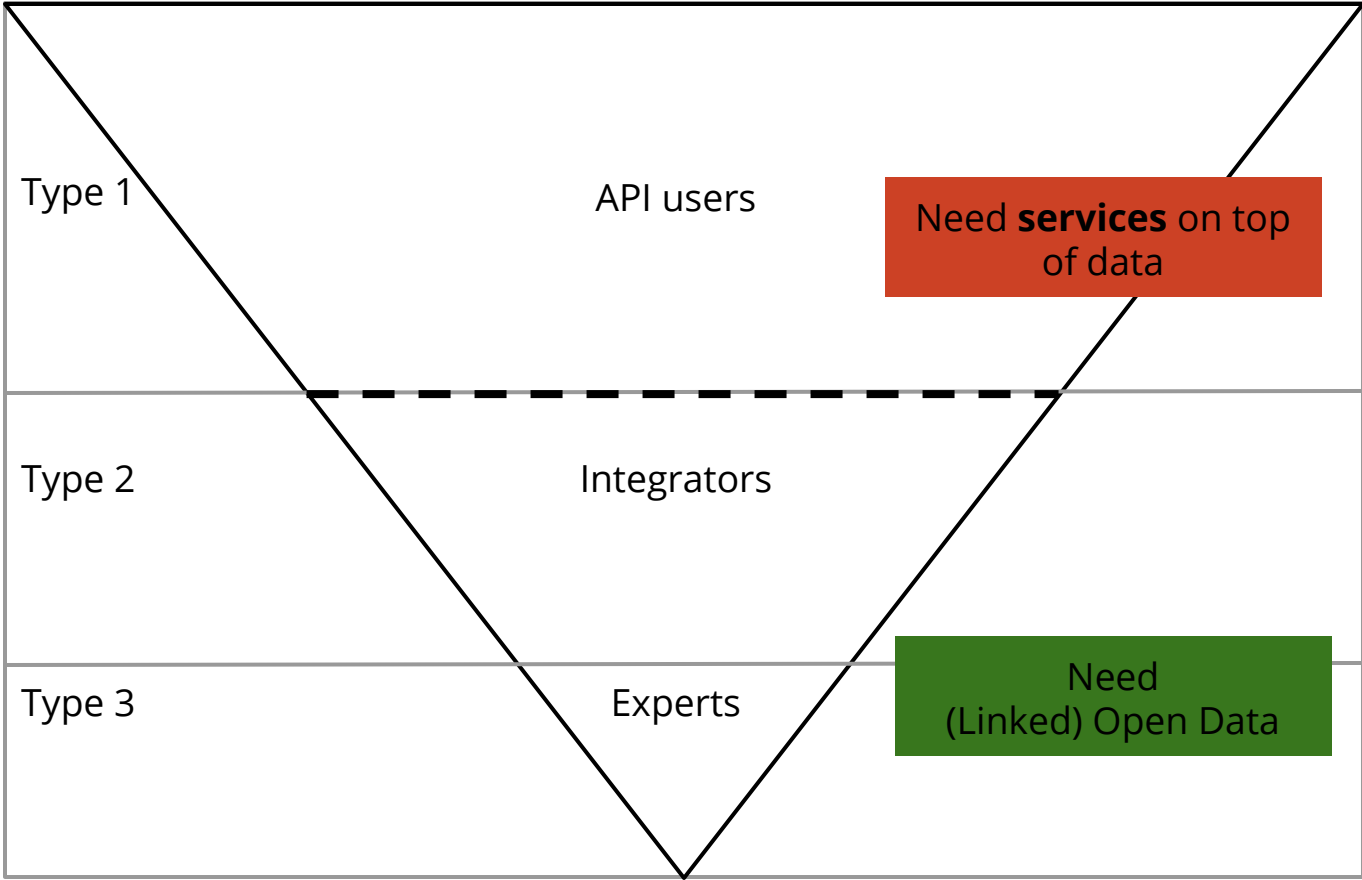
Everything with phylum Chordata?

Query the Web yourself at  
<http://client.linkeddatafragments.org>

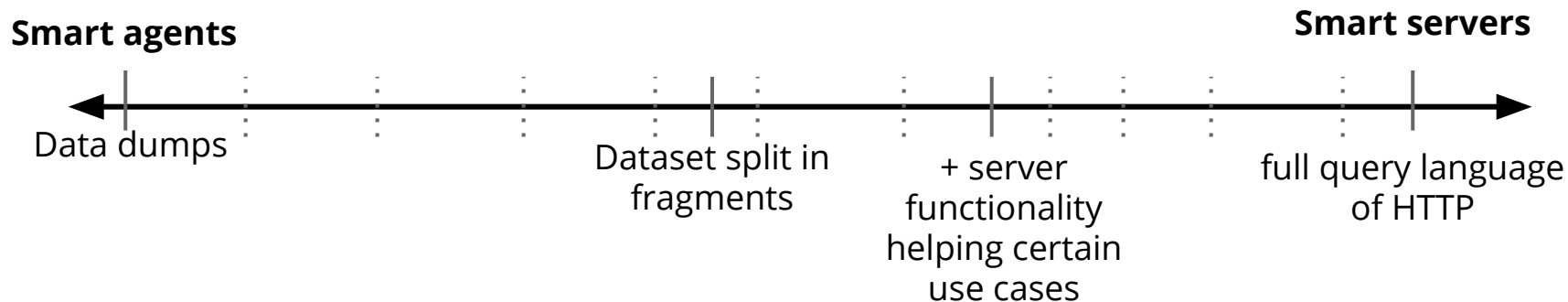


## **2. Maximising reuse**

Or maximising functionalities?

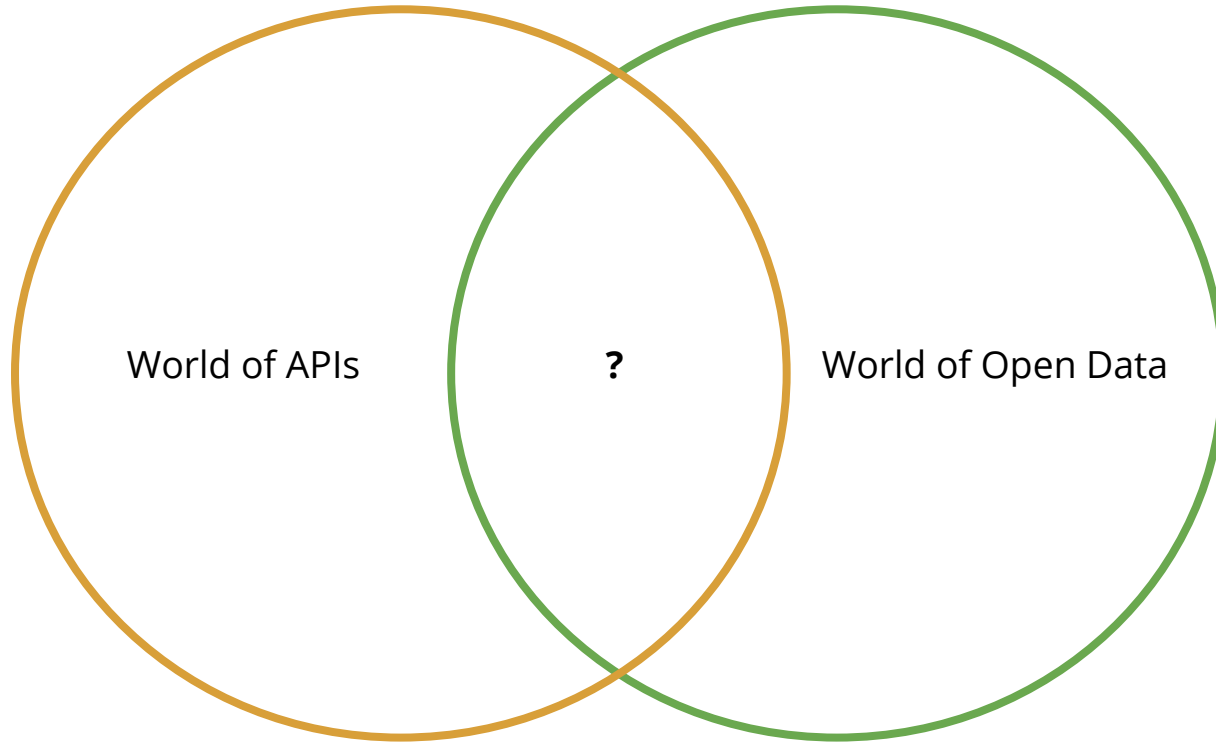


# What do we want?



Read more at <http://linkeddatafragments.org>

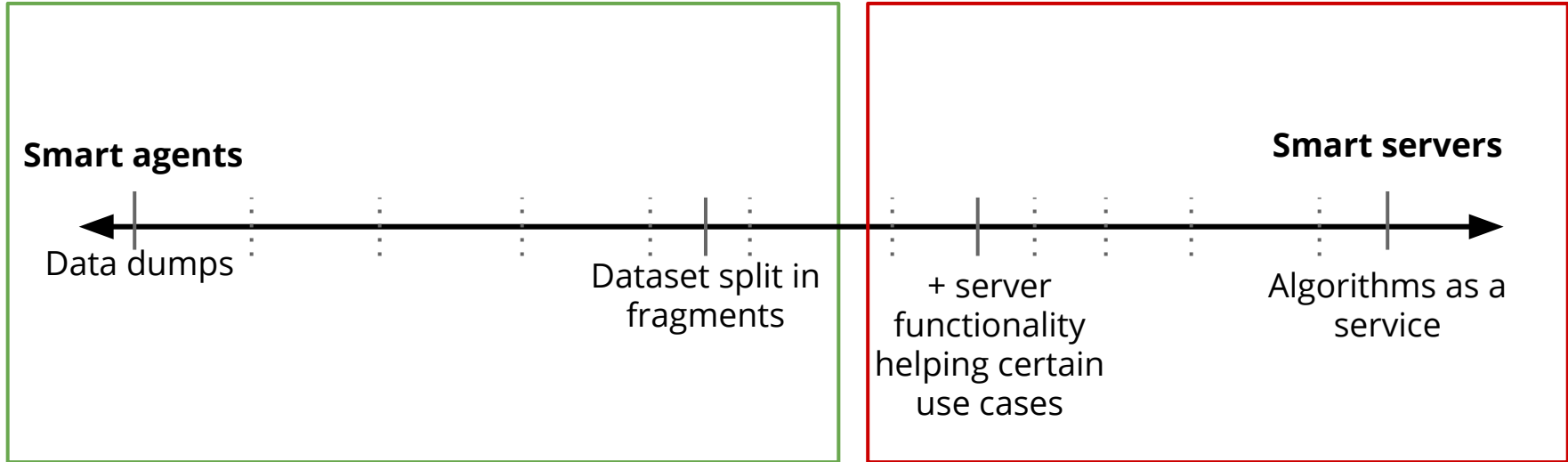
# Confusion arose



# Publishing data vs. data services

Data publishing  
(cheap)

Data services  
(rather expensive)



# 3. Why?

Where's the money?

# Servers publishing Open Data e.g.,

- all the planned and actual arrivals and departures
- the network of roads in a certain region



## web-services e.g.,

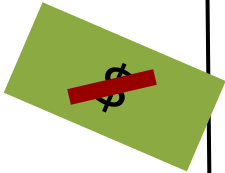
- a route planner: from → to
- the closest station to your current location?



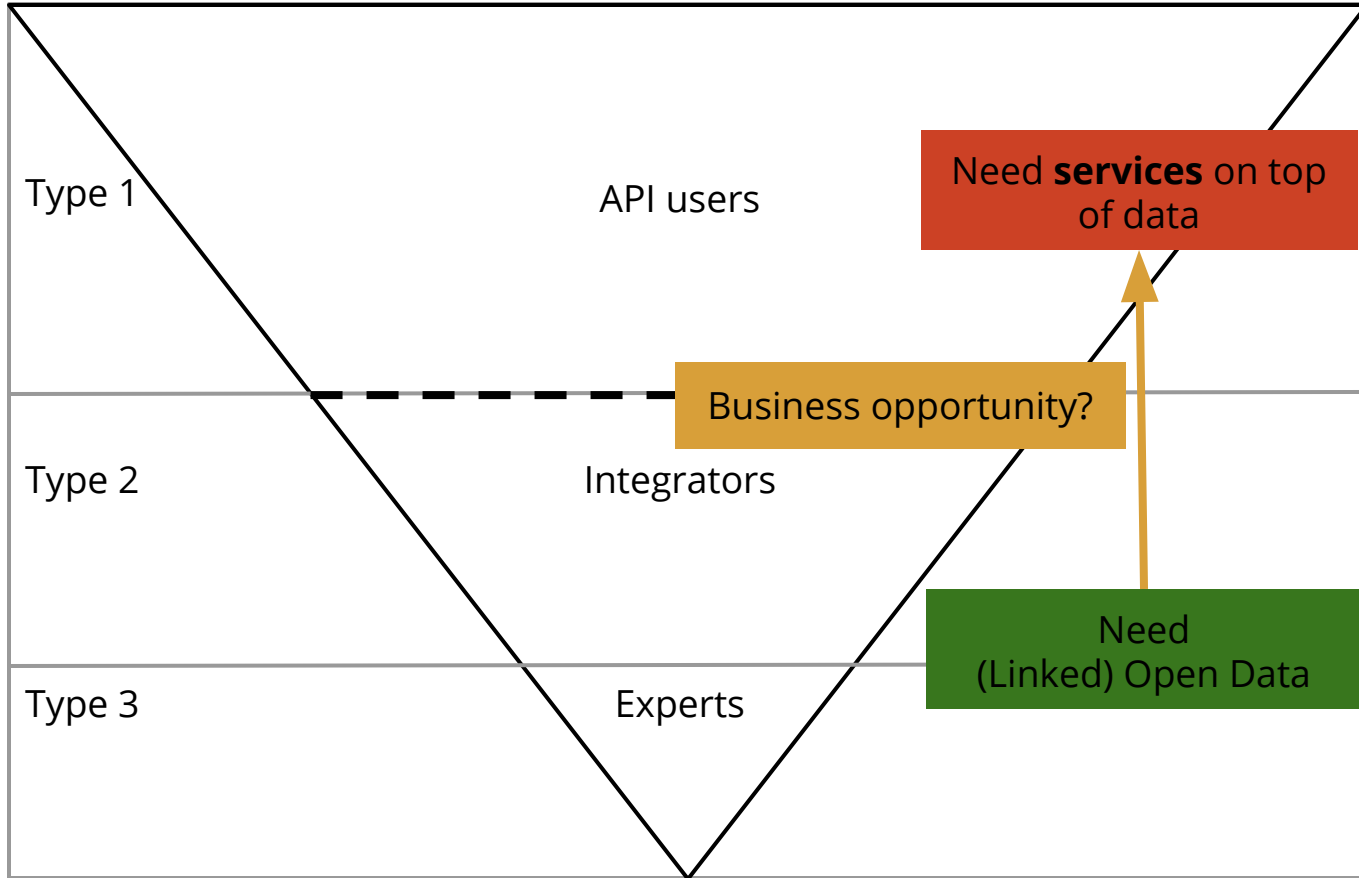
end-user apps



end-users

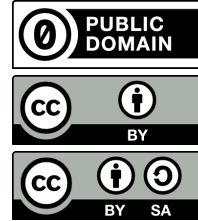






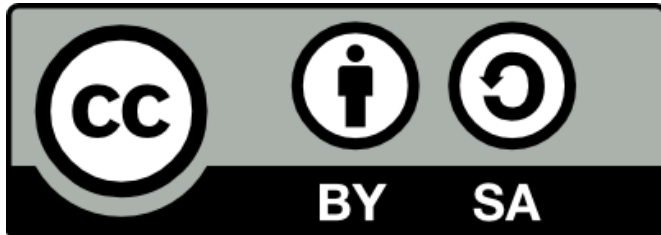
# Wrapping up

1.



2. Web of documents → web of data

We want a world where  
knowledge creates power for the many,  
not the few.



@pietercolpaert