Re-use under licence - what use is quality if your product is illegal?

Anthony Beck

Abstract

Geospatial web services provide general and niche products some of which are based on multiple data sources each data source used under licence. End-users may further chain these services to produce a myriad of other derived data products and services. The background licencing issues can become very complex. Incompatibilities between different licences means that some data products should not be integrated. Hence, licence heterogeneity is a barrier to data integration and interoperability in the way that semantic, syntactic and schematic heterogeneities are also barriers. This presentation will provide an overview of these issues, to describe a potential semantic licence framework, and through discussion better understand where the OGC might be able to help with licence, or legal, interoperability.

This paper was presented at the Open Geospatial Conference hosted at the British Geological Survey in Keyworth, Nottingham on the 16th September 2015.

Figure 1:

This document has been written in CommonMark: an unambiguous implementation of Markdown for scholarly writing.

Contents

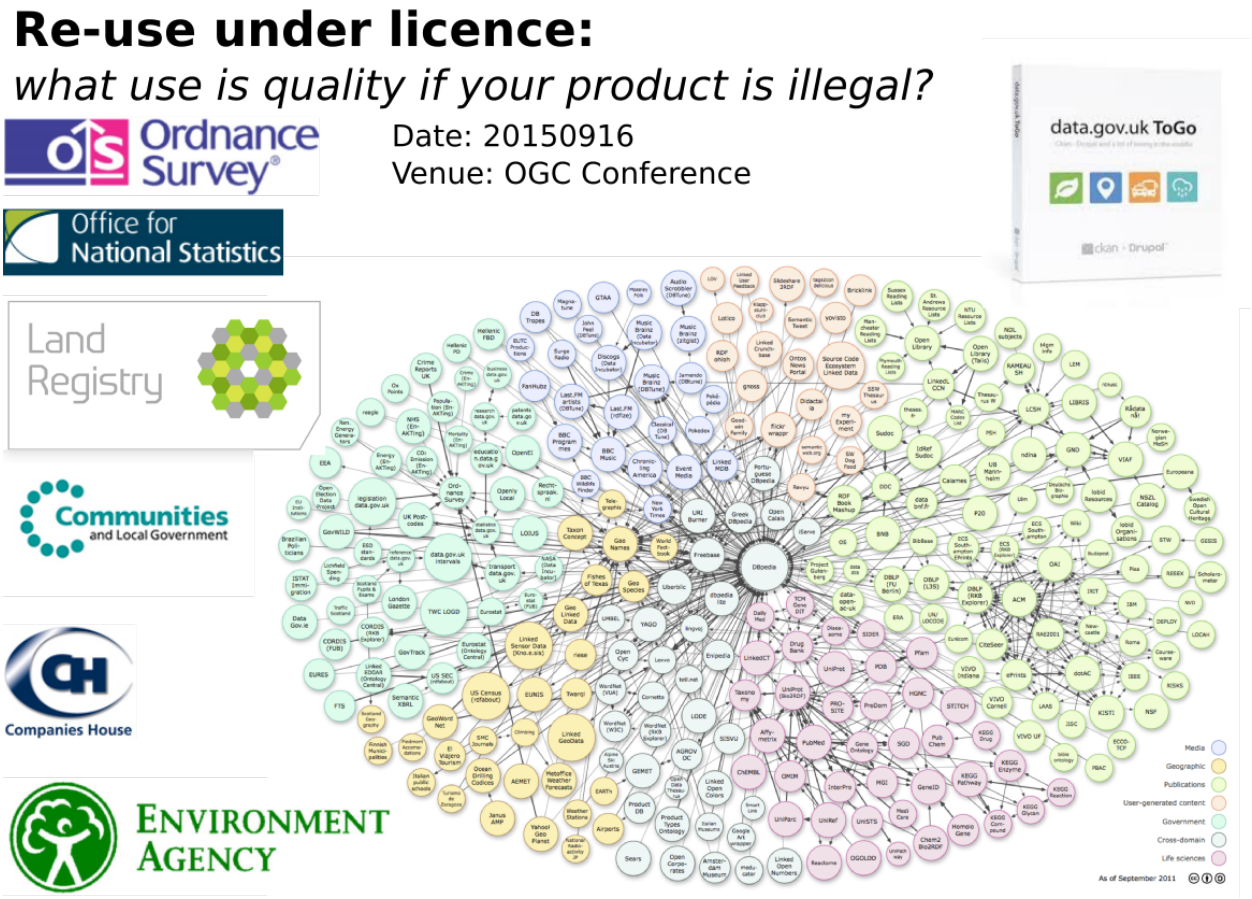


Figure 2:

Go down for licence and other metadata about this presentation

1.6 About me

Figure 4: It's all about me - details about Anthony Beck

- Research Fellow, University of Nottingham: orcid
- Director, Geolytics Limited - A spatial data analytics consultancy

1.7 About this presentation

- Available on GitHub - https://github.com/AntArch/Presentations_Github/
- Fully referenced PDF

2 A potted history of mapping

2.1 In the beginning was the geoword

and the word was ***cartography***

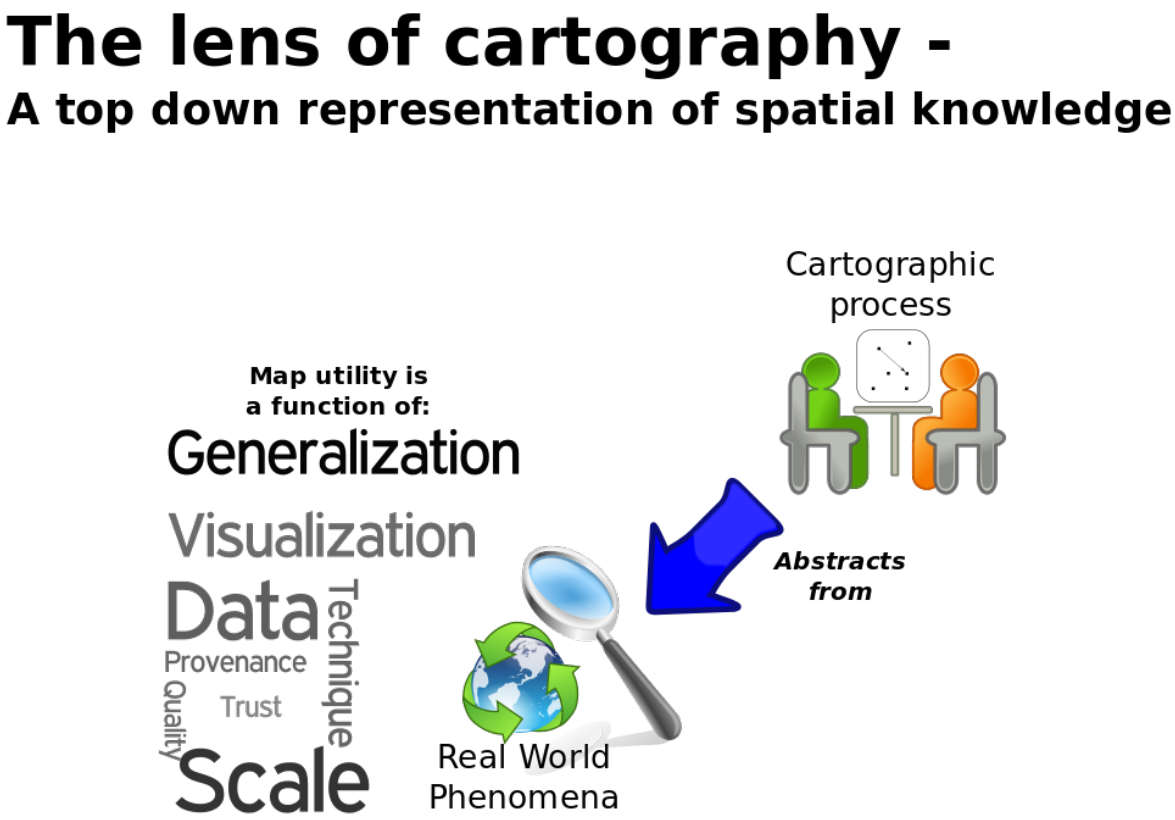


Figure 5: The lens of cartography Beck (2015g)

Figure 6: A static map (public domain)

Cartography was king. Static representations of spatial knowledge with the cartographer deciding what to represent.

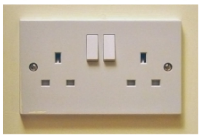
3 And then there was data .....

Figure 7: Data Beck (2015h)

At the end of the 20th Century National Mapping and Cadastral agencies characterised by:

Heterogeneous and incompatible data

Syntactic - data flavours
Schematic - data designs
Semantic - data description



Restrictive licences

Figure 8: But the data was siloed (restricted use)

Restrictive data

Figure 9: Concerted efforts to de-silo data and make data interoperable (restricted use)

Making data interoperable and open

4 Technical interoperability - levelling the field

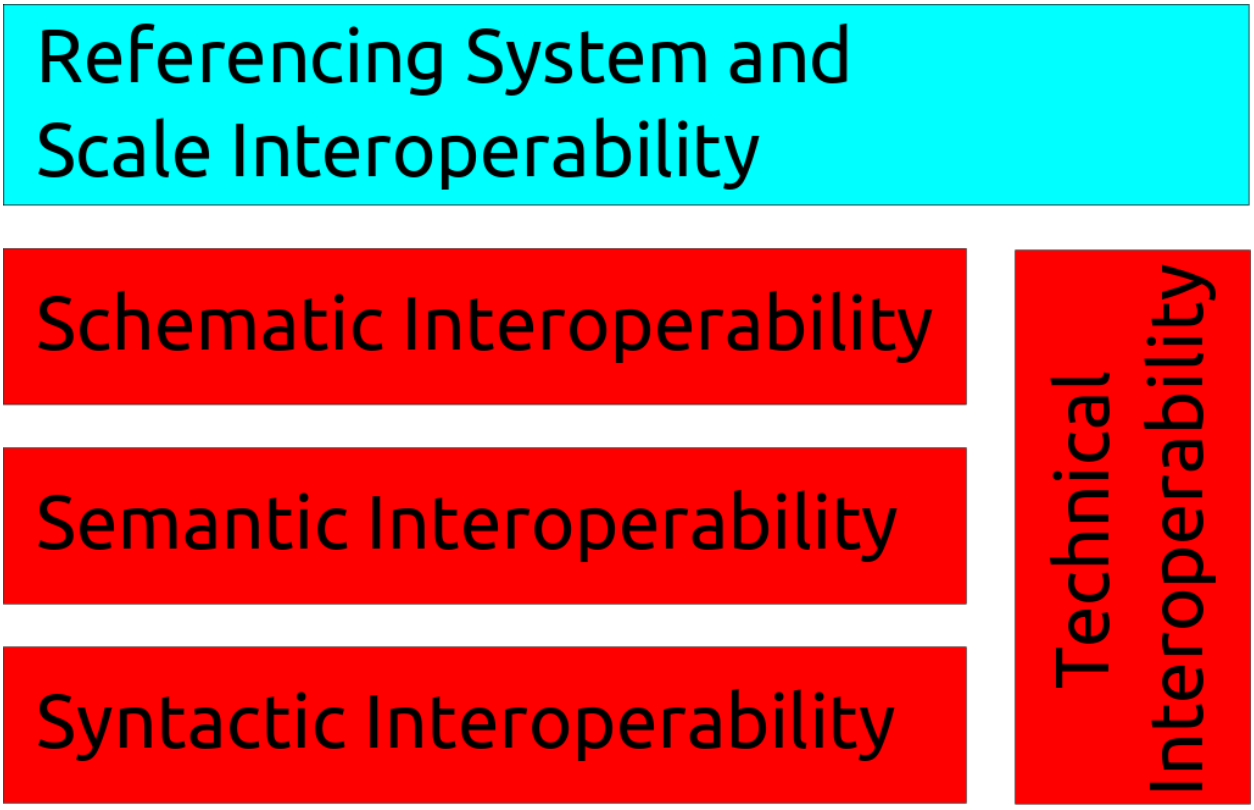


Figure 10: Interoperable integration of spatial data - the technological issues Beck (2015d)

4.1 Facilitating data driven visualization

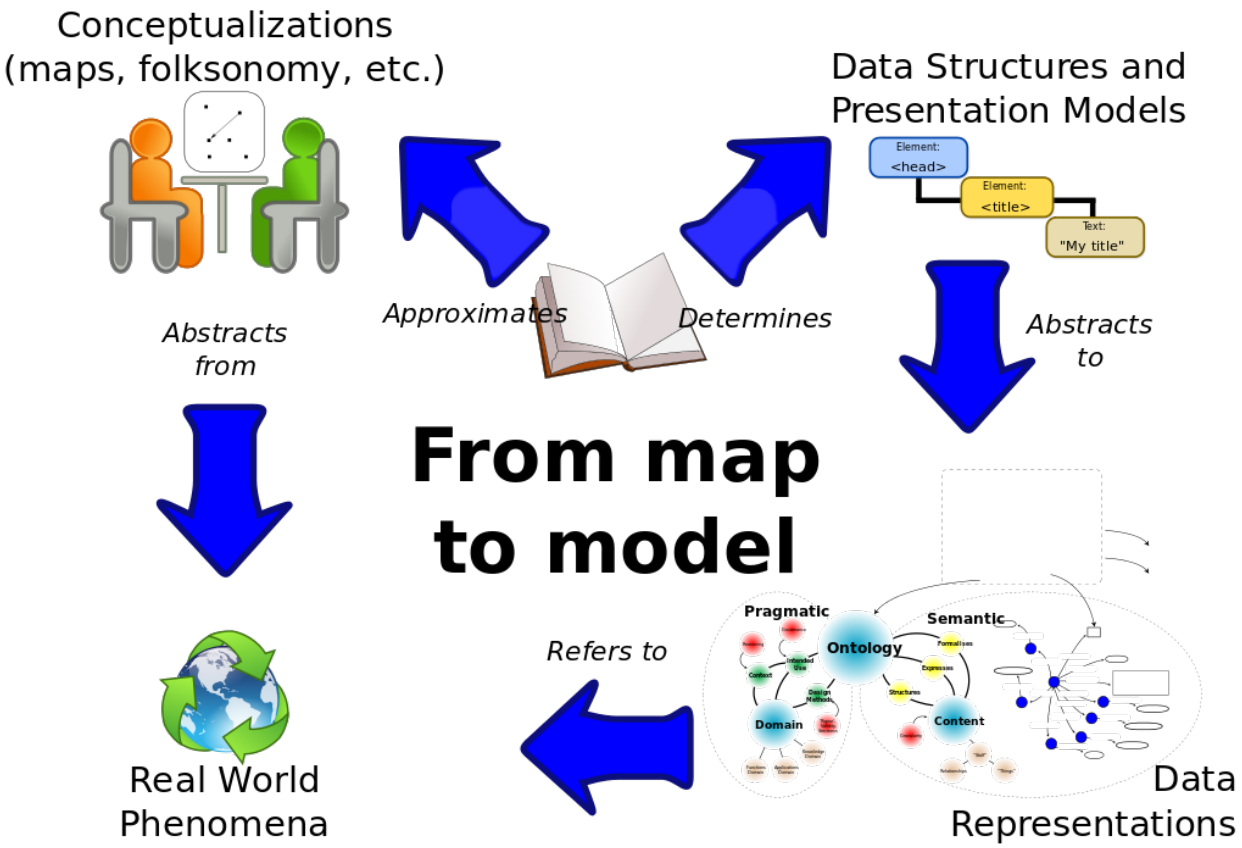


Figure 11: From Map to Model The changing paradigm of map creation from cartography to data driven visualization Beck (2015c)

From Map to Model The changing paradigm of map creation from cartography to data driven visualization

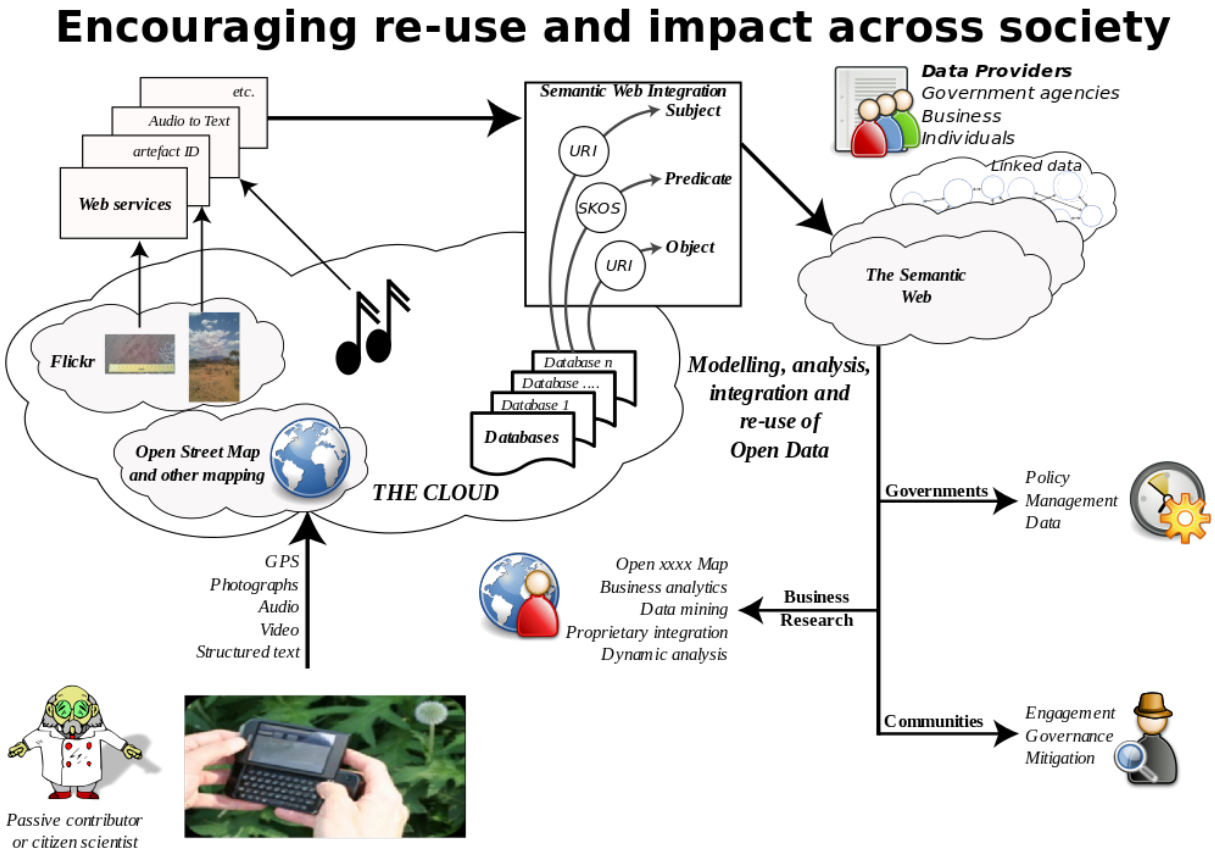


Figure 12: Local To Global integration of data to create multiple generic products Beck (2015e)

Figure 13: A new working paradigm (public domain)

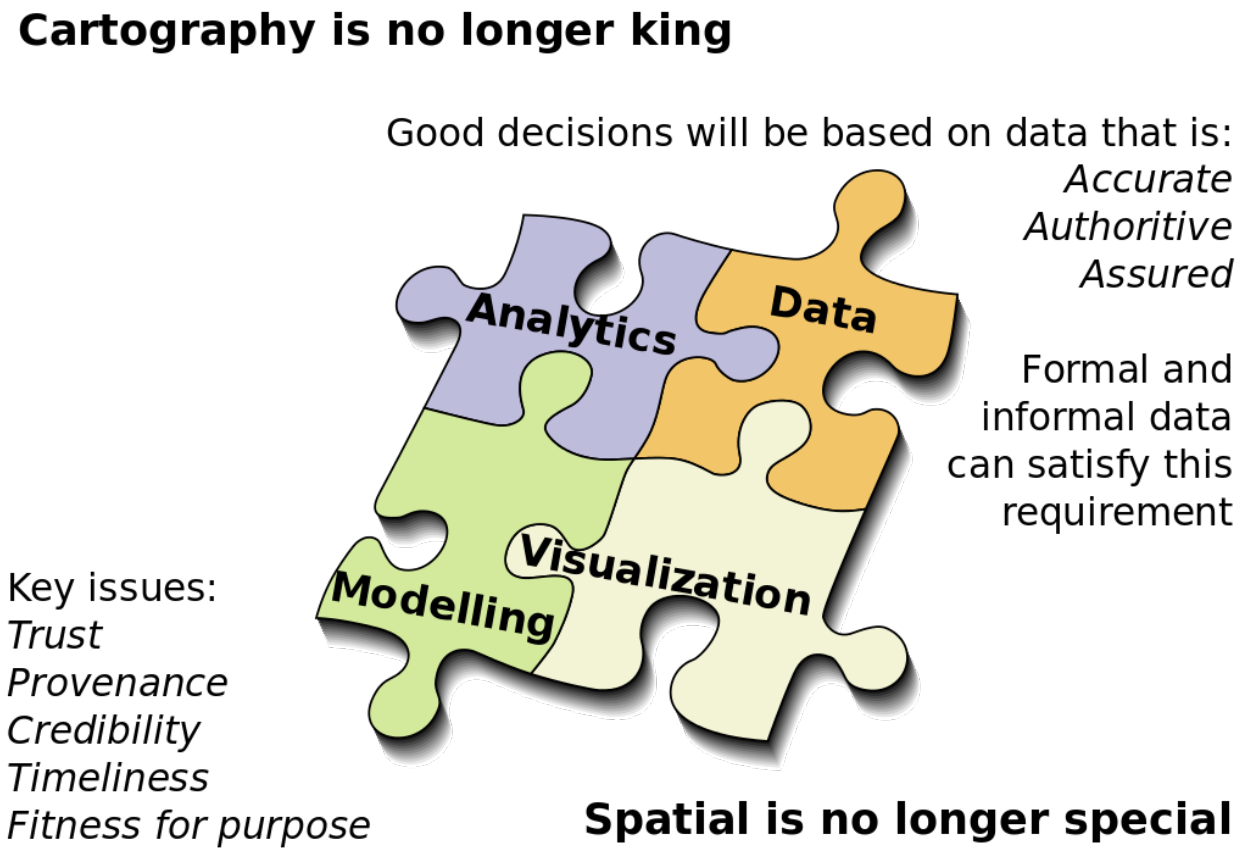


Figure 14: Cartography is no longer key. Spatial mapping is now about the the formal and informal data stack. Elements such as provenance, credibility are much more important for use and re-use of this data. Beck (2015b)

5 What about non-technical interoperability issues?

Issues surrounding non-technical interoperability include:

- Policy interoperabilty
- Licence interoperability
- Legal interoperability
- Social interoperability

We will focus on licence interoperability

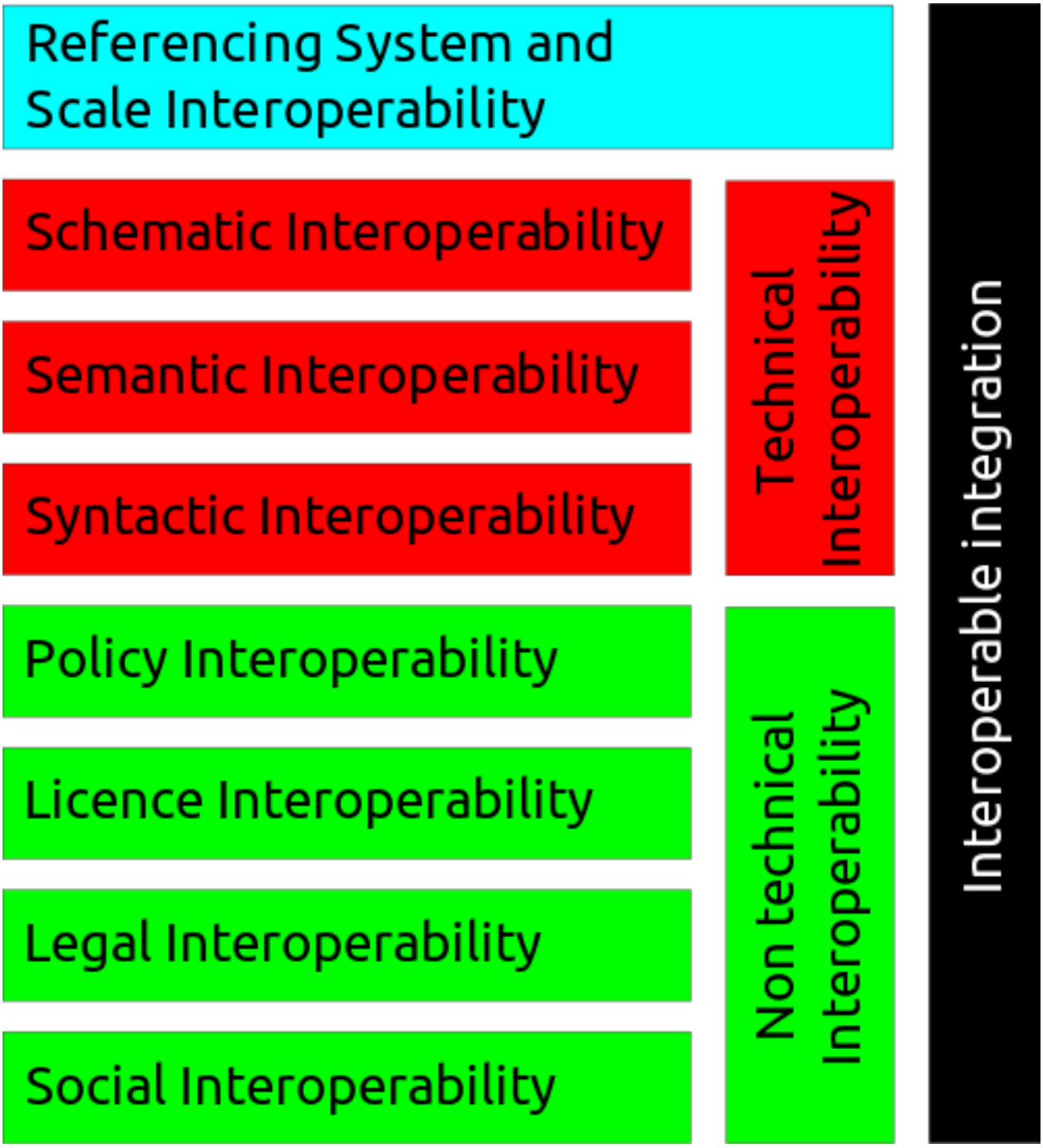


Figure 15: The full stack that enables interoperable integration of spatial data Beck (2015f)

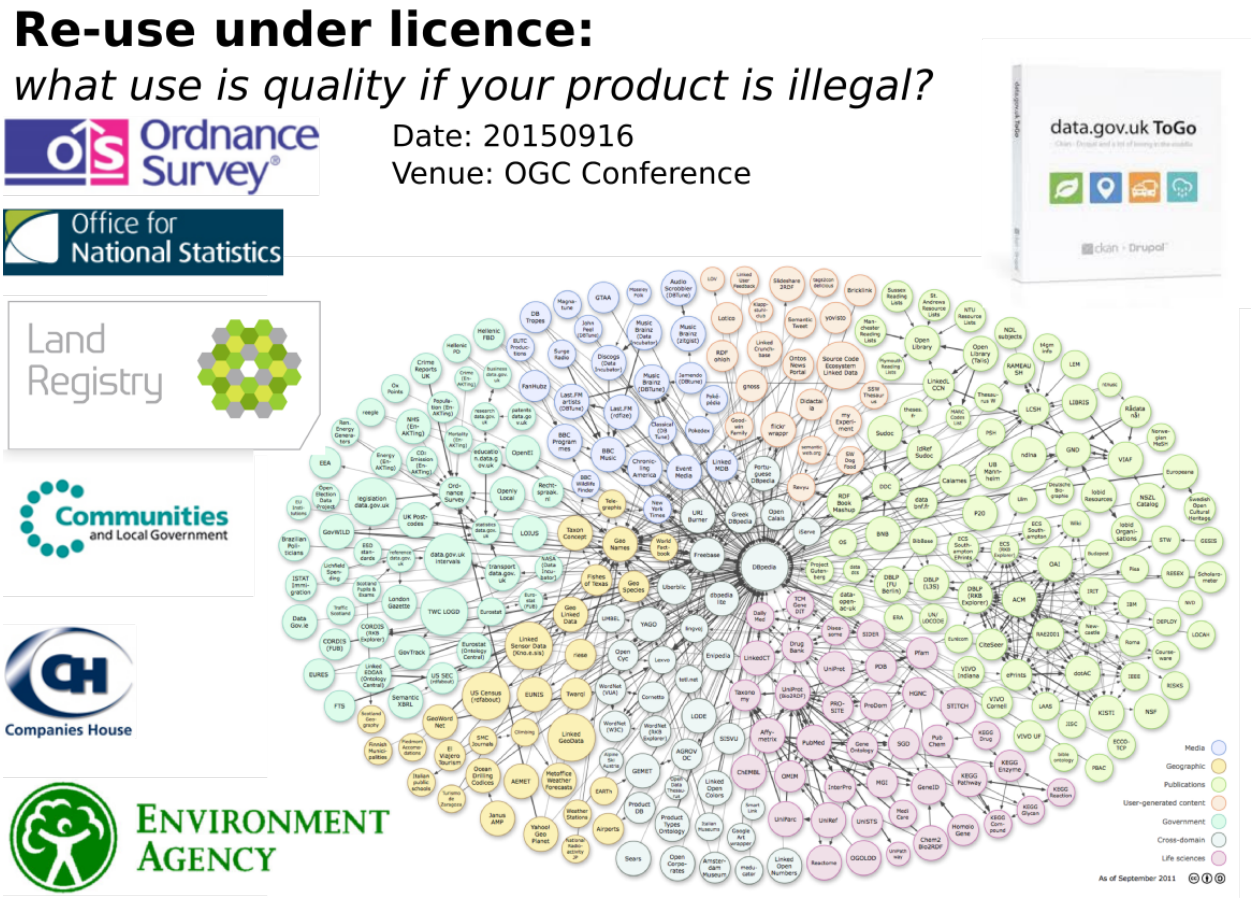


Figure 16: The modern data landscape (restricted)

There is a multitude of formal and informal data.

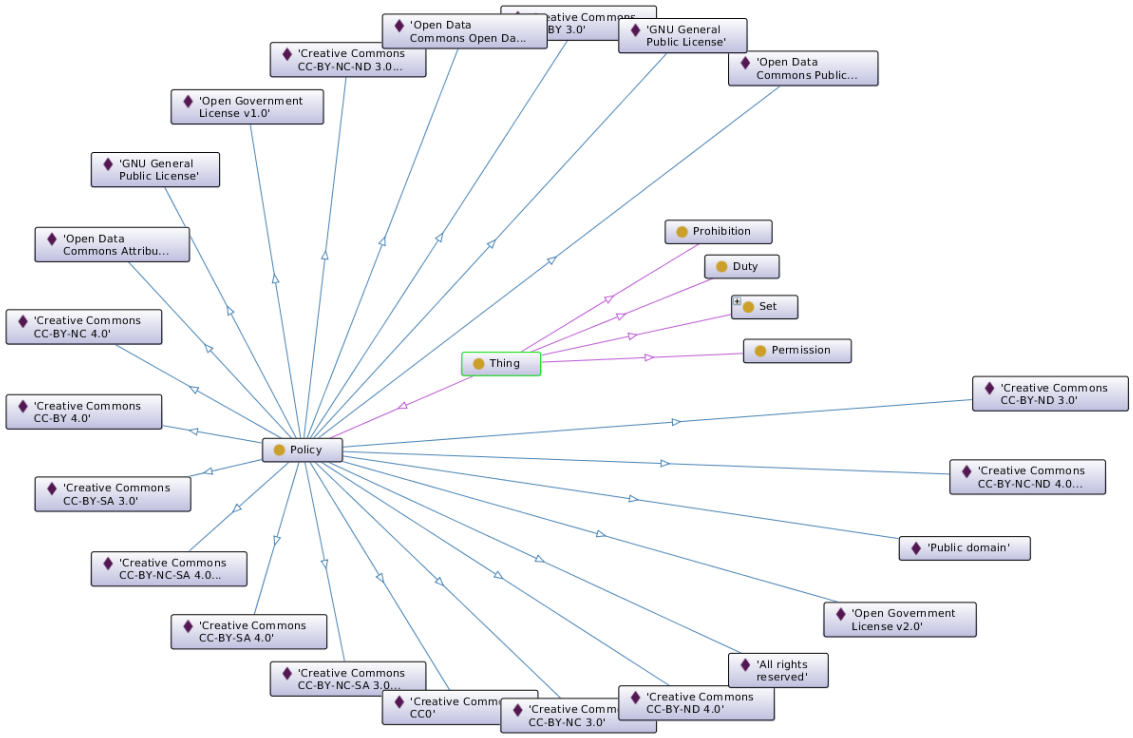


Figure 17: Some licences (Anon. n.d.)

Each of these data objects can be licenced in a different way. This shows some of the licences described by the RDFLicence ontology

5.1 What is a licence?

Wikipedia state:

> A license may be granted by a party ("licensor") to another party ("licensee") as an element of an agreement between those parties.
>
> A shorthand definition of a license is "an authorization (by the licensor) to use the licensed material (by the licensee)."

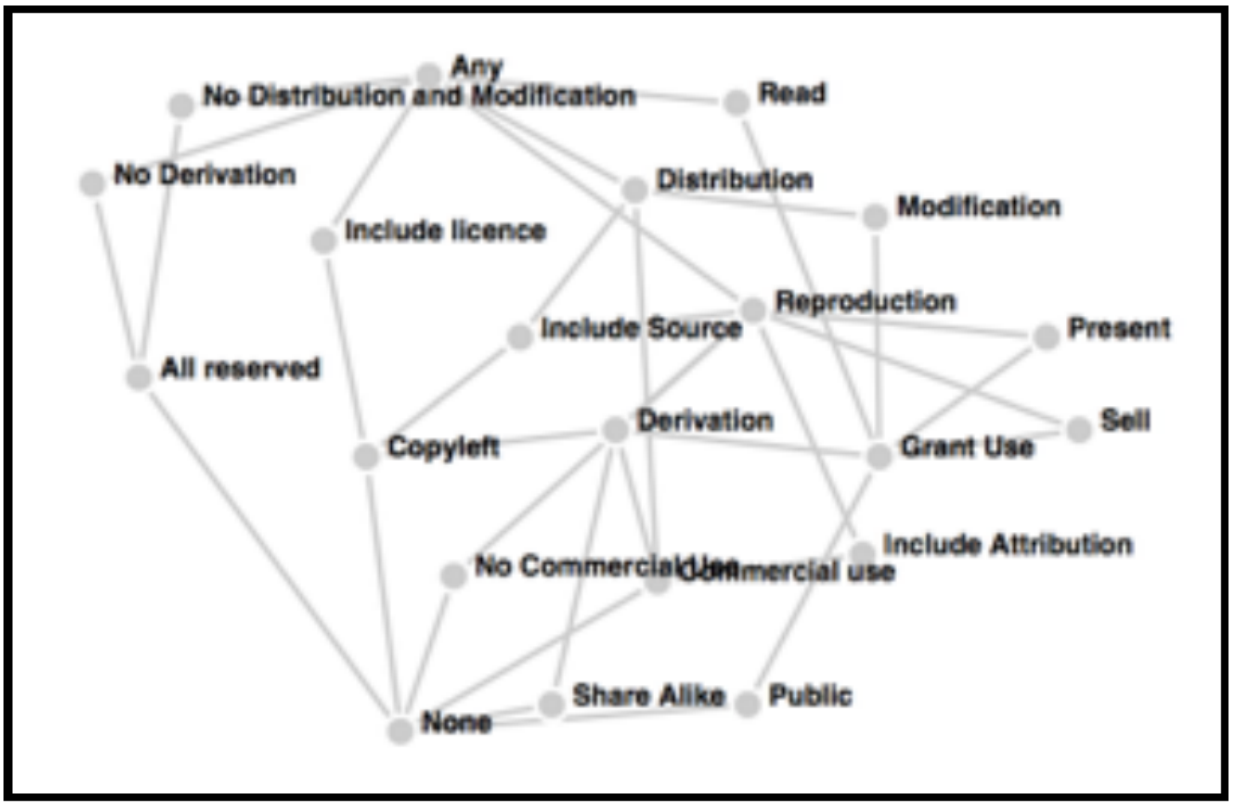


Figure 18: A licence describes what you can and cannot do to a data object (Anon. n.d.)

Concepts (derived from Formal Concept Analysis) surrounding licences

Two lead organisations have developed legal frameworks for content licensing:

- Creative Commons (CC) and
- Open Data Commons (ODC).

Until the release of CC version 4, published in November 2013, the CC licence did not cover data. Between them, CC and ODC licences can cover all forms of digital work.

- **There are many other licence types**
- Many are bespoke
  - Bespoke licences are difficult to manage
  - Many legacy datasets have bespoke licences

Figure 19: Creative Commons Gianni (2008)

I'll describe CC in more detail

5.2 Creative Commons Zero

Creative Commons Zero (CC0) is essentially public domain which allows:

- Reproduction
- Distribution
- Derivations

5.2.1 Constraints on CC0

The following clauses constrain CC0:

- Permissions
  - ND – No derivatives: the licensee can not derive new content from the resource.
- Requirements
  - BY – By attribution: the licensee must attribute the source.
  - SA – Share-alike: if the licensee adapts the resource, it must be released under the same licence.
- Prohibitions
  - NC – Non commercial: the licensee must not use the work commercially without prior approval.

5.2.2 CC license combinations

Table 1: Creative Commons license combinations

License	Reproduction	Distribution	Derivation	ND	BY	SA	NC
CC0	X	X	X				
CC-BY-ND	X	X		X	X		
CC-BY-NC-ND	X	X		X	X		X
CC-BY	X	X	X		X		
CC-BY-SA	X	X	X		X	X	
CC-BY-NC	X	X	X		X		X
CC-BY-NC-SA	X	X	X		X	X	X

6 Why are licenses important?

- They tell you what you can and can't do with 'stuff'
- Very significant when multiple datasets are combined
  - It then becomes an issue of license compatibility

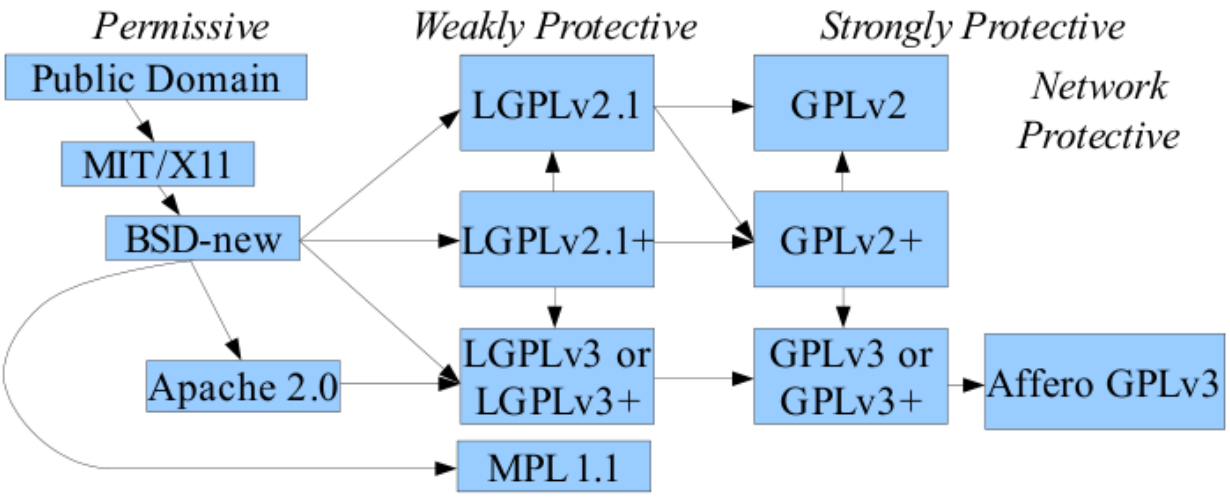


Figure 20: Compatibility of common open-source software licenses Wheeler (2007)

6.1 Which is important when we mash up data

Certain licences when combined:

- Are incompatible
  - Creating data islands
- Inhibit commercial exploitation (NC)
- Force the adoption of certain licences
  - If you want people to commercially exploit your stuff don't incorporate CC-BY-NC-SA data!
- Stops the derivation of *new works*

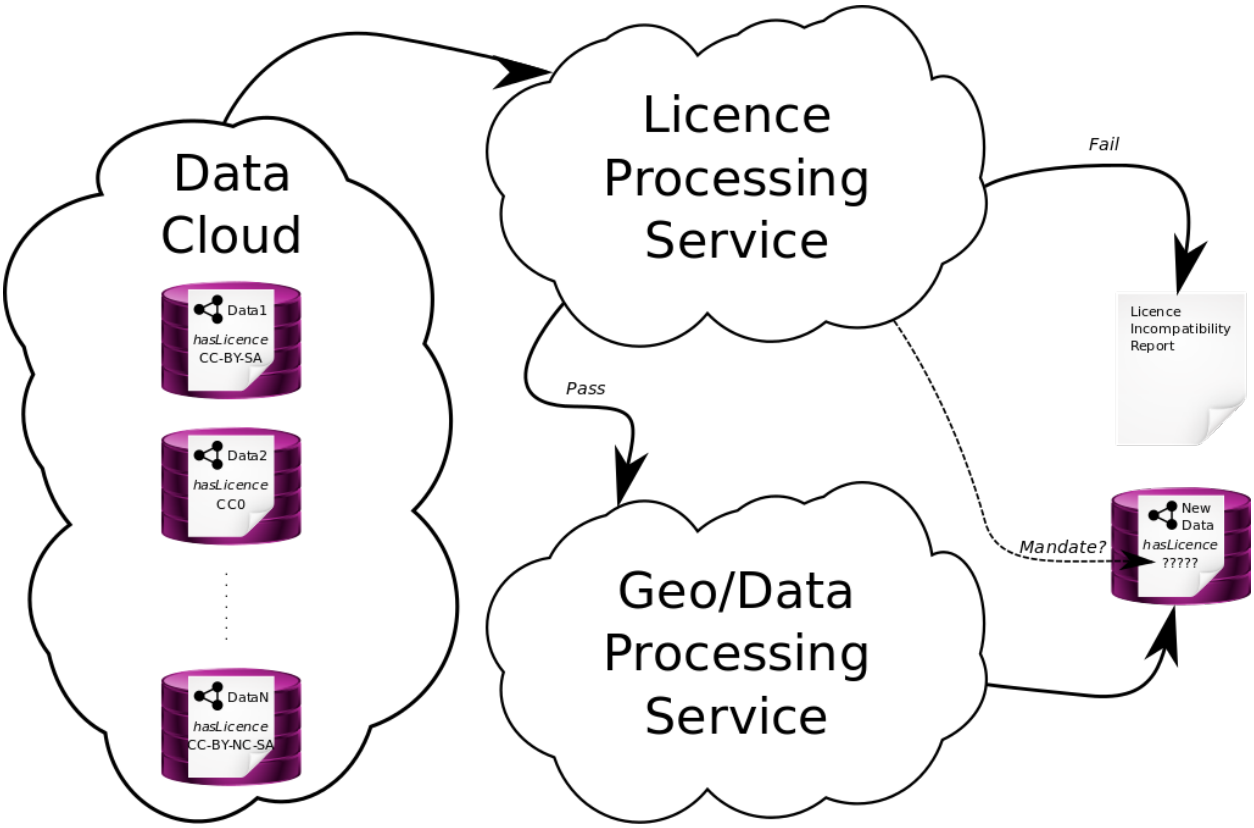


Figure 21: A conceptual licence processing workflow Beck (2015a)

A conceptual licence processing workflow. The licence processing service analyses the incoming licence metadata and determines if the data can be legally integrated and any resulting licence implications for the derived product.

7 A rudimentry logic example

```
Data1 hasDerivedContentIn NewThing.

Data1 hasLicence a cc-by-sa.

What hasLicence a cc-by-sa? #reason here

If X hasDerivedContentIn Y and hasLicence Z then Y hasLicence Z. #reason here

Data2 hasDerivedContentIn NewThing.

Data2 hasLicence a cc-by-nc-sa.

What hasLicence a cc-by-nc-sa? #reason here

Nothing hasLicence a cc-by-nc-sa and hasLicence a cc-by-sa. #reason here
```

And processing this within the Protege reasoning environment

```
from IPython.display import YouTubeVideo
YouTubeVideo('jUzGF401vLc')
```

```
    <iframe
        width="400"
        height="300"
        src="https://www.youtube.com/embed/jUzGF401vLc"
        frameborder="0"
        allowfullscreen
    ></iframe>
```

7.1 Here's something I prepared earlier

A live presentation (for those who weren't at the event).....

```
from IPython.display import YouTubeVideo
YouTubeVideo('tkRB5Rp1_W4')
```

```
    <iframe
        width="400"
        height="300"
        src="https://www.youtube.com/embed/tkRB5Rp1_W4"
        frameborder="0"
        allowfullscreen
    ></iframe>
```

8 A more robust logic

- Would need to decouple licence incompatibility from licence name into licence clause (see table below)
- Deal with all licence type
- Provide recommendations based on desired derivative licence type
- Link this through to the type of process in a workflow:
  - data derivation is, from a licence position, very different to contextual display
- etc..... for discussion?

Table 2: Creative Commons license combinations

License	Reproduction	Distribution	Derivation	ND	BY	SA	NC
CC0	X	X	X				
CC-BY-ND	X	X		X	X		
CC-BY-NC-ND	X	X		X	X		X
CC-BY	X	X	X		X		
CC-BY-SA	X	X	X		X	X	
CC-BY-NC	X	X	X		X		X
CC-BY-NC-SA	X	X	X		X	X	X

9 OGC and Licence interoperability

- The geo business landscape is increasingly based on integrating heterogeneous data to develop new products
- Licence heterogeneity is a barrier to data integration and interoperability
- A licence calculus can help resolve and identify heterogenties leading to
  - legal compliance
  - confidence
- Use of standards and collaboration with organisations is crucial
  - Open Data Licensing ontology
  - The Open Data Institute
- Failure to do this could lead to breaches in data licenses
  - and we all know where that puts us……..

Figure 22: Breaching a data license can be serious (restricted = randomly!)

10 Questions

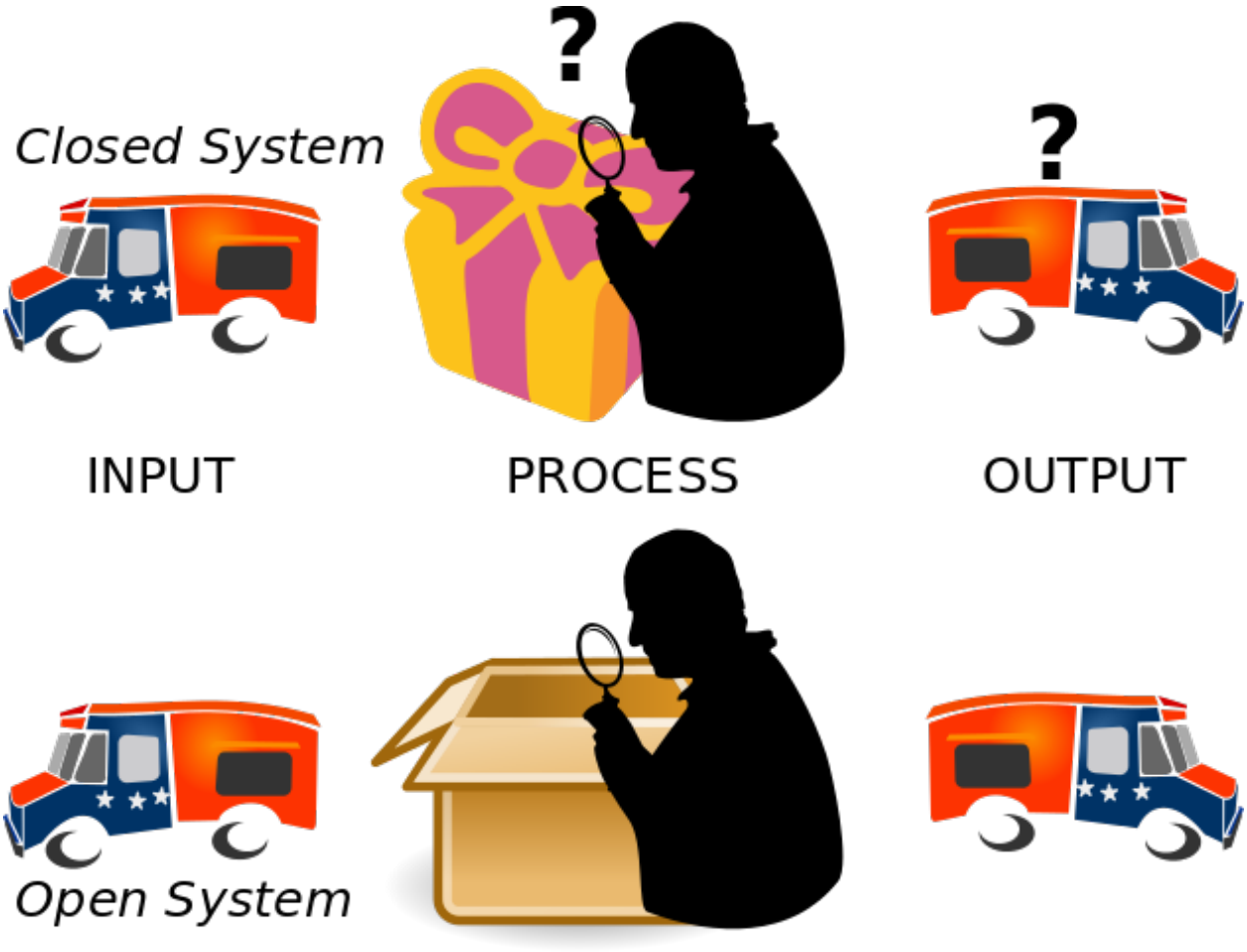


Figure 23: Processing transparency between open and closed systems

References

Anon., A bottom up approach for licences classification and selection. Available at: http://www.slideshare.net/enricodaga/a-bottom-up-approach-for-licences-classification-and-selection [Accessed September 10TZ, 2015a].

Anon., RDFLicense - the datahub. Available at: http://datahub.io/dataset/rdflicense [Accessed September 10TZ, 2015b].

Beck, A., 2015a. A conceptual licence processing workflow. Available at: https://commons.wikimedia.org/wiki/File:ConceptualLicenceProcessingWorkflow.svg [Accessed September 10, 2015].

Beck, A., 2015b. Cartography is no longer key. Available at: https://commons.wikimedia.org/wiki/File:CartographyNoLongerKing.svg [Accessed September 7, 2015].

Beck, A., 2015c. From map to model the changing paradigm of map creation from cartography to data driven visualization. Available at: https://commons.wikimedia.org/wiki/File:FromMapToModel.svg [Accessed September 7, 2015].

Beck, A., 2015d. Interoperable integration of spatial data - the technological issues. Available at: https://commons.wikimedia.org/wiki/File:TechnicalInteroperableIntegrationOfSpatialData.svg [Accessed September 7, 2015].

Beck, A., 2015e. Local to global integration of data to create multiple generic products. Available at: https://commons.wikimedia.org/wiki/File:Local_To_Global_integration_of_data_to_create_multiple_generic_products.svg [Accessed September 7, 2015].

Beck, A., 2015f. The full stack that enables interoperable integration of spatial data. Available at: https://commons.wikimedia.org/wiki/File:InteroperableIntegrationOfSpatialData.svg [Accessed September 7, 2015].

Beck, A., 2015g. The lens of cartography. Available at: https://commons.wikimedia.org/wiki/File:TheLensOfCartography.svg [Accessed September 7, 2015].

Beck, A., 2015h. Types of data. Available at: https://commons.wikimedia.org/wiki/File:Data_types_-_en.svg [Accessed September 7, 2015].

Gianni, S., 2008. We love creative commons. Available at: https://www.flickr.com/photos/steren/2732488224/ [Accessed July 27, 2015].

Wheeler, D.A., 2007. Compatibility of common open-source software licenses. Available at: https://en.wikipedia.org/wiki/File:Floss-license-slide-image.png [Accessed July 27, 2015].