

# Rights Statements on the Web of Data

Leigh Dodds  
Programme Manager, Talis Platform  
Talis  
25<sup>th</sup> October 2009



<http://creativecommons.org/licenses/by/2.0/uk/>



# Why Are Rights Statements Important?

A photograph showing the interior of a large, traditional wooden barn. The ceiling is made of a complex network of light-colored wooden beams and rafters. The floor is made of light-colored wood planks. In the background, there are several wooden barrels and a small wooden cart. On the left side, there is a wooden desk with a small model house on it. On the right side, there is a person sitting at a desk. The overall atmosphere is rustic and historical.

Clear framework for all forms of reuse

A close-up view of the astronomical clock face at the Prague Astronomical Clock. The face is circular with multiple concentric rings. The outermost ring features Roman numerals from I to XII. Inside this are several other rings, each containing different symbols, likely representing the zodiac signs. A small sunburst or starburst symbol is positioned near the top center. The background is a light blue color.

Building for the Future

# Some Terminology

# License

**“...permission to do or not to do something...”**

**“To permit or authorize by license...”**

**“The document granting such permission...”**

# Waiver

- “1. an intentional relinquishment of some right, interest, or the like.
- 2. an express or written statement of such relinquishment.”



There are many different legal tools



Different communities have different priorities



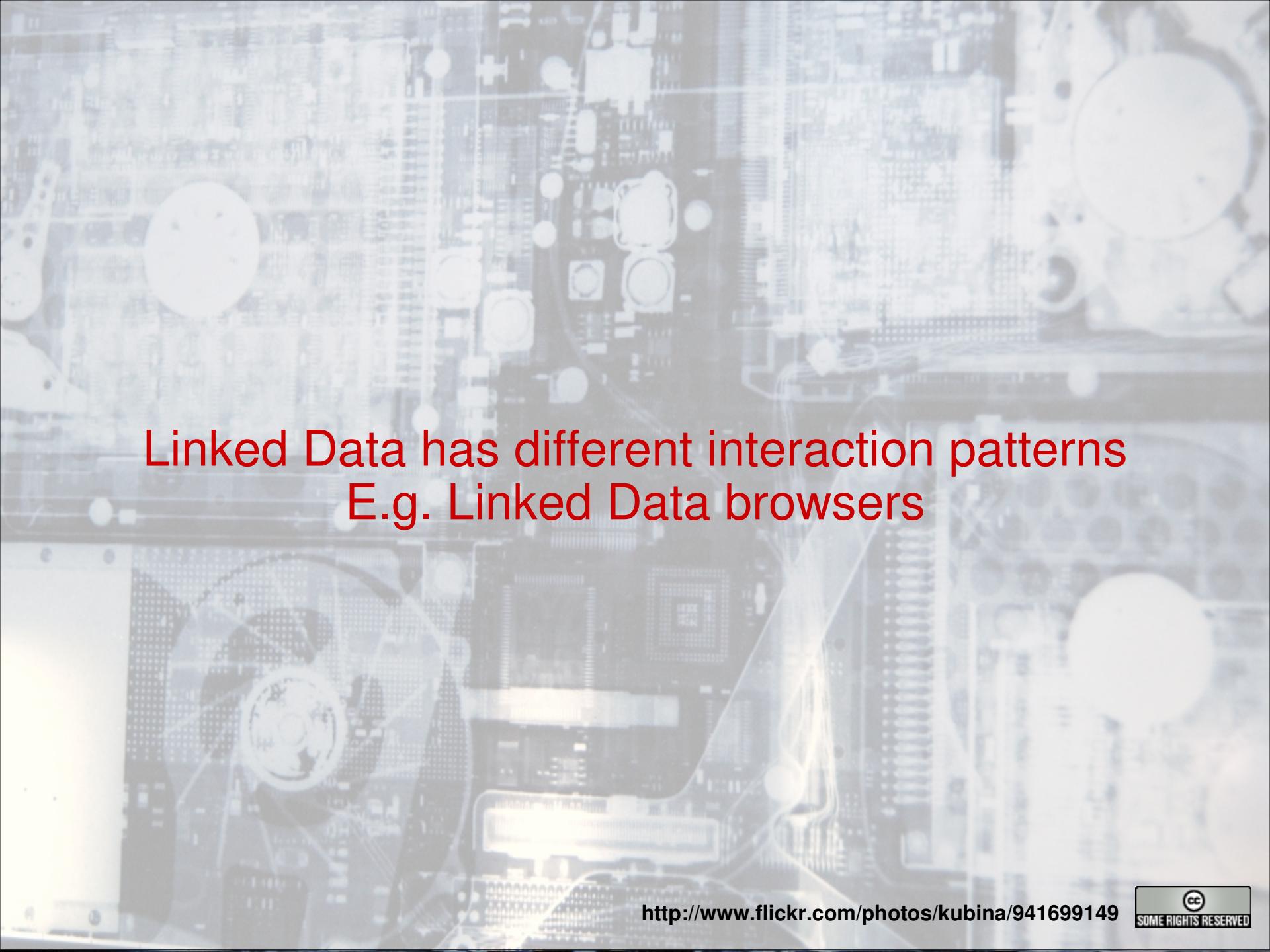
Legal tools need both human and machine  
readable expressions



As of July 2009

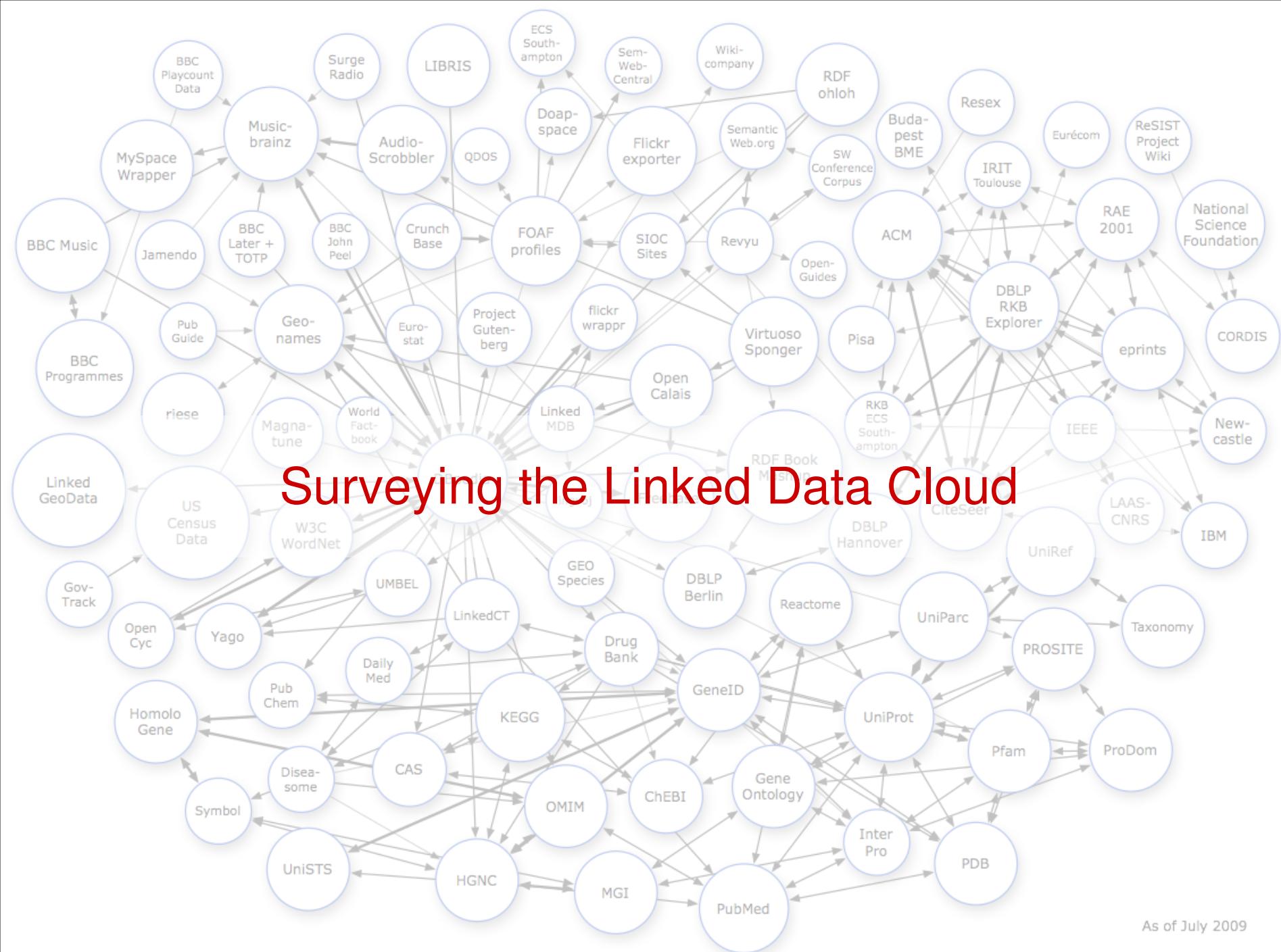


## Explicit agreement involved in using a Web 2.0 API

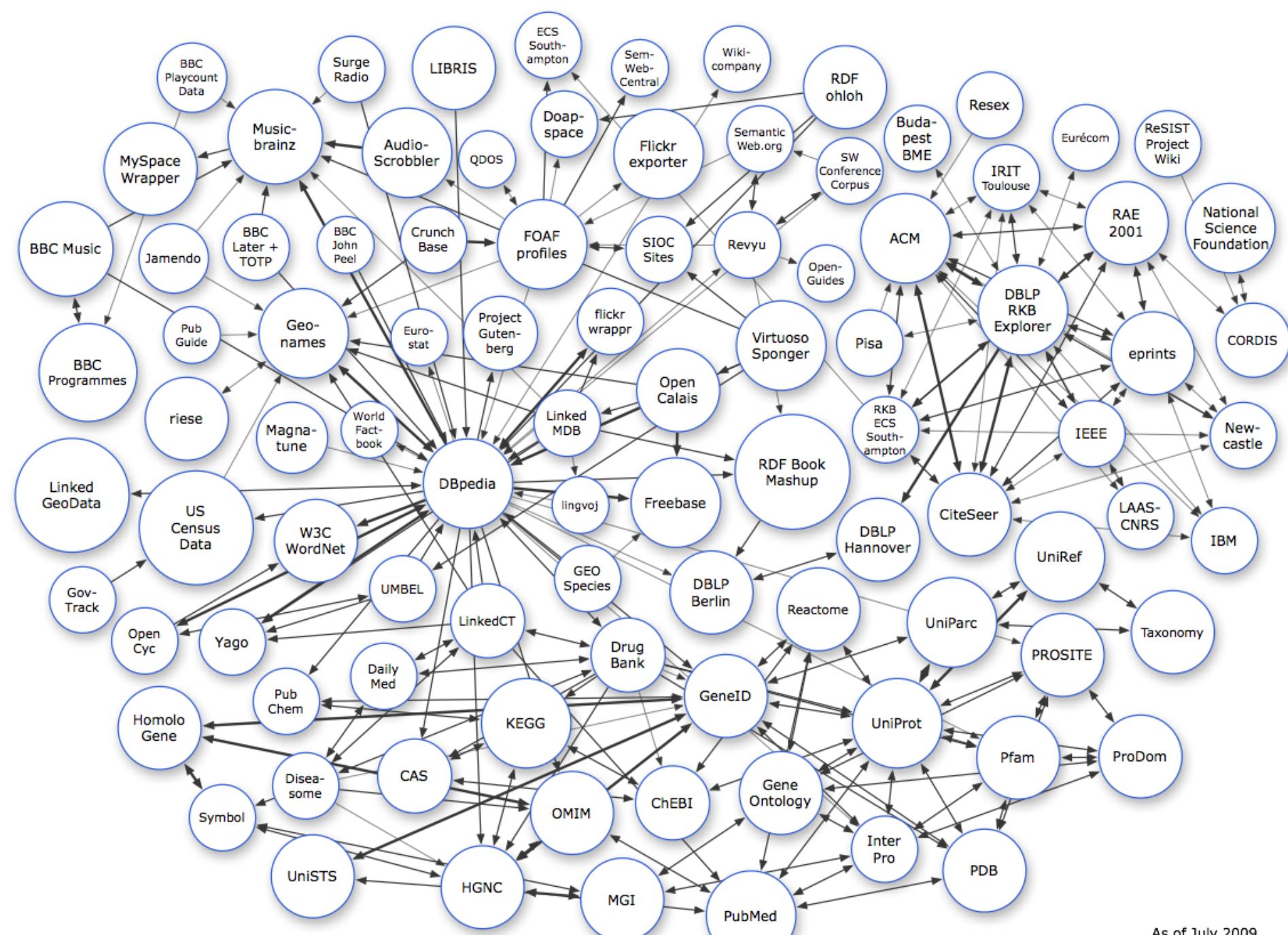


Linked Data has different interaction patterns  
E.g. Linked Data browsers

# Surveying the Linked Data Cloud



As of July 2009



As of July 2009



# Can we find explicit `an` rights statement associated with these datasets?

And which licensing option does each use?

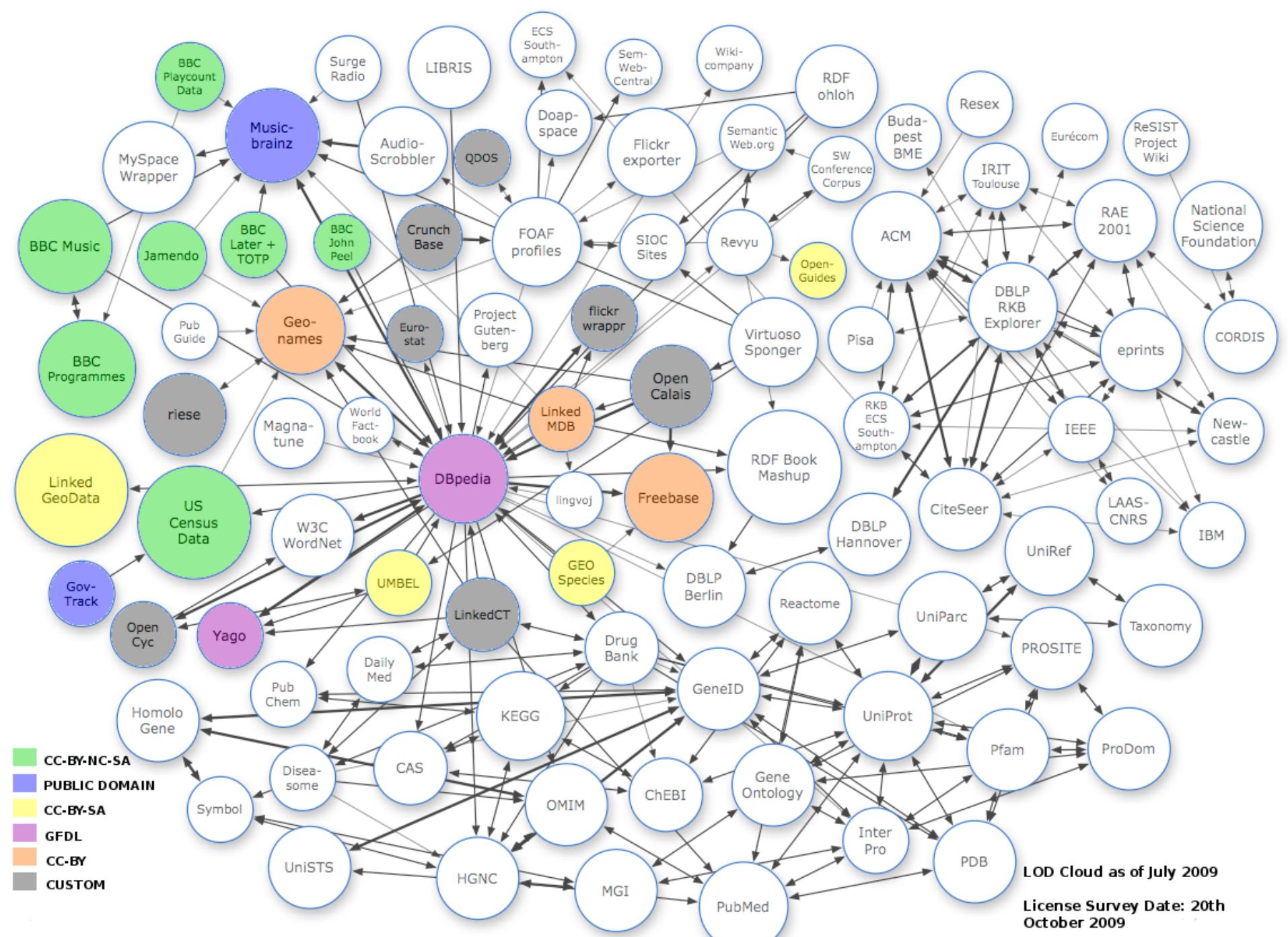
This diagram illustrates a complex network of interconnected semantic web projects and datasets. Nodes are represented by blue circles, and connections between them are shown as grey arrows. The network is highly interconnected, with many nodes having multiple links to others. Key clusters include:

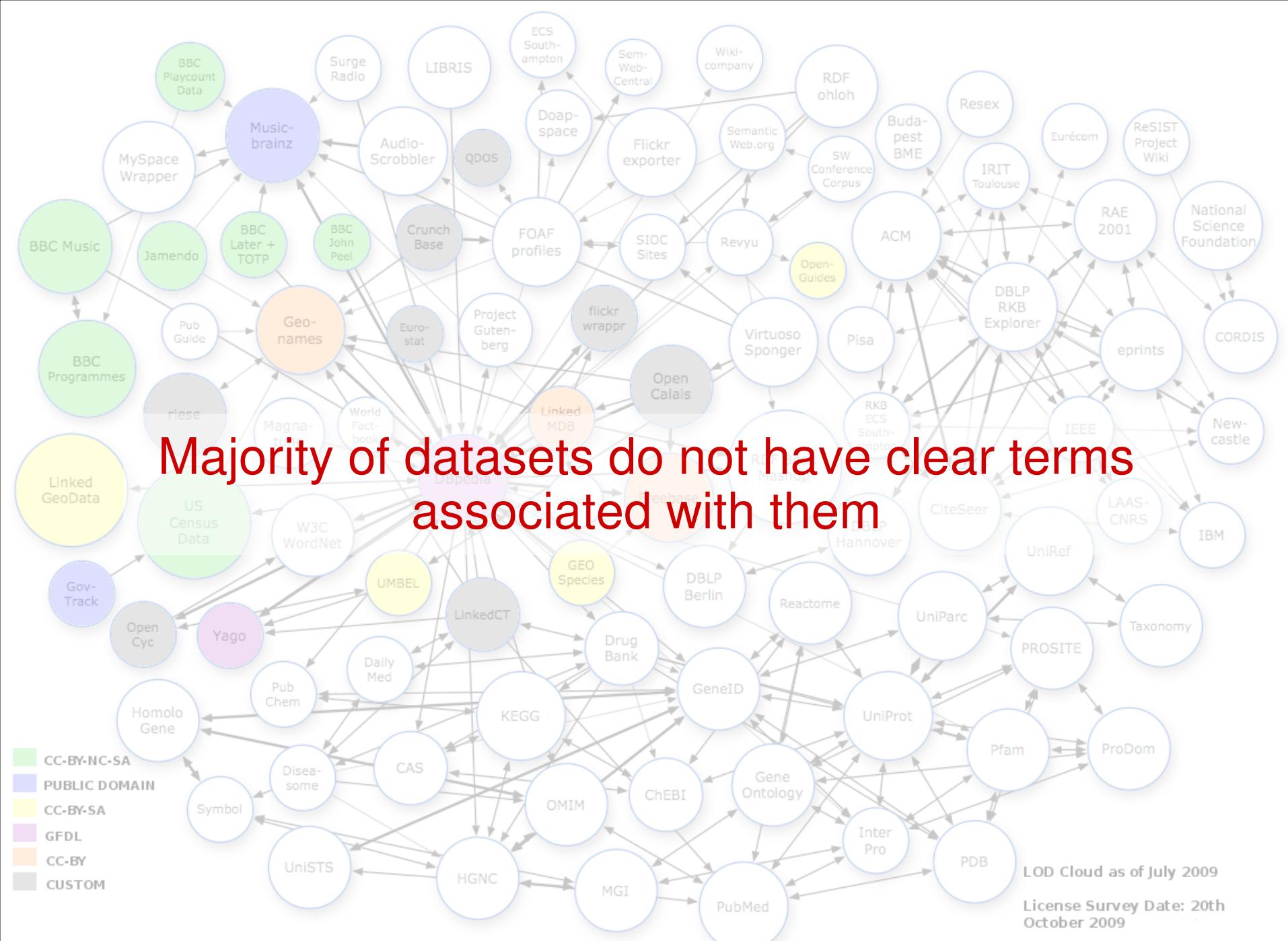
- A central cluster around "FOAF profiles" connected to "Geonames", "Pub Guide", "BBC Music", "BBC Programmes", "BBC Playcount Data", "MySpace Wrapper", "Surge Radio", "LIBRIS", "ECS Southampton", "Sem-Web-Central", "Wiki-company", "RDF ohloh", "Resex", "Eurécom", "ReSIST Project Wiki", "National Science Foundation", and "CORDIS".
- A cluster around "DBLP RKB Explorer" connected to "IEEE", "eprints", "RAE 2001", "ACM", "Budapest BME", "IRIT Toulouse", "Open-Guides", "Revyu", "Semantic Web.org", "SW Conference Corpus", "Pisa", "RKB ECS Southampton", "RDF Book Machine", "DBLP Hannover", "CiteSeer", "DBLP Berlin", "Reactome", "UniProt", "UniParc", "PROSITE", "Pfam", "ProDom", "UniRef", "Taxonomy", "IBI", "EAS-CNRS", "Newcastle", and "IBM".
- A cluster around "HGNC" connected to "KEGG", "Drug Bank", "GEO Species", "UMBEL", "LinkedCT", "Yago", "Open Cyc", "Gov-Track", "US Census Data", "W3C WordNet", "World Fact-book", "Eurostat", "Project Gutenberg", "Crunch Base", "QDOS", "Doap-space", "Flickr exporter", "SIOC Sites", "Revyu", "Virtuoso Sponger", "Open Calais", "RDF Book Machine", "DBLP Berlin", "Reactome", "GeneID", "OMIM", "ChEBI", "Gene Ontology", "UniProt", "UniParc", "PROSITE", "Pfam", "ProDom", "UniRef", "Taxonomy", "IBI", "EAS-CNRS", "Newcastle", and "IBM".
- A cluster around "HGNC" connected to "HGNC", "MGI", "PubChem", "Daily Med", "CAS", "Symbol", "UniSTS", "Disease-some", "Homolo Gene", "UniProt", "UniParc", "PROSITE", "Pfam", "ProDom", "UniRef", "Taxonomy", "IBI", "EAS-CNRS", "Newcastle", and "IBM".

# The Results

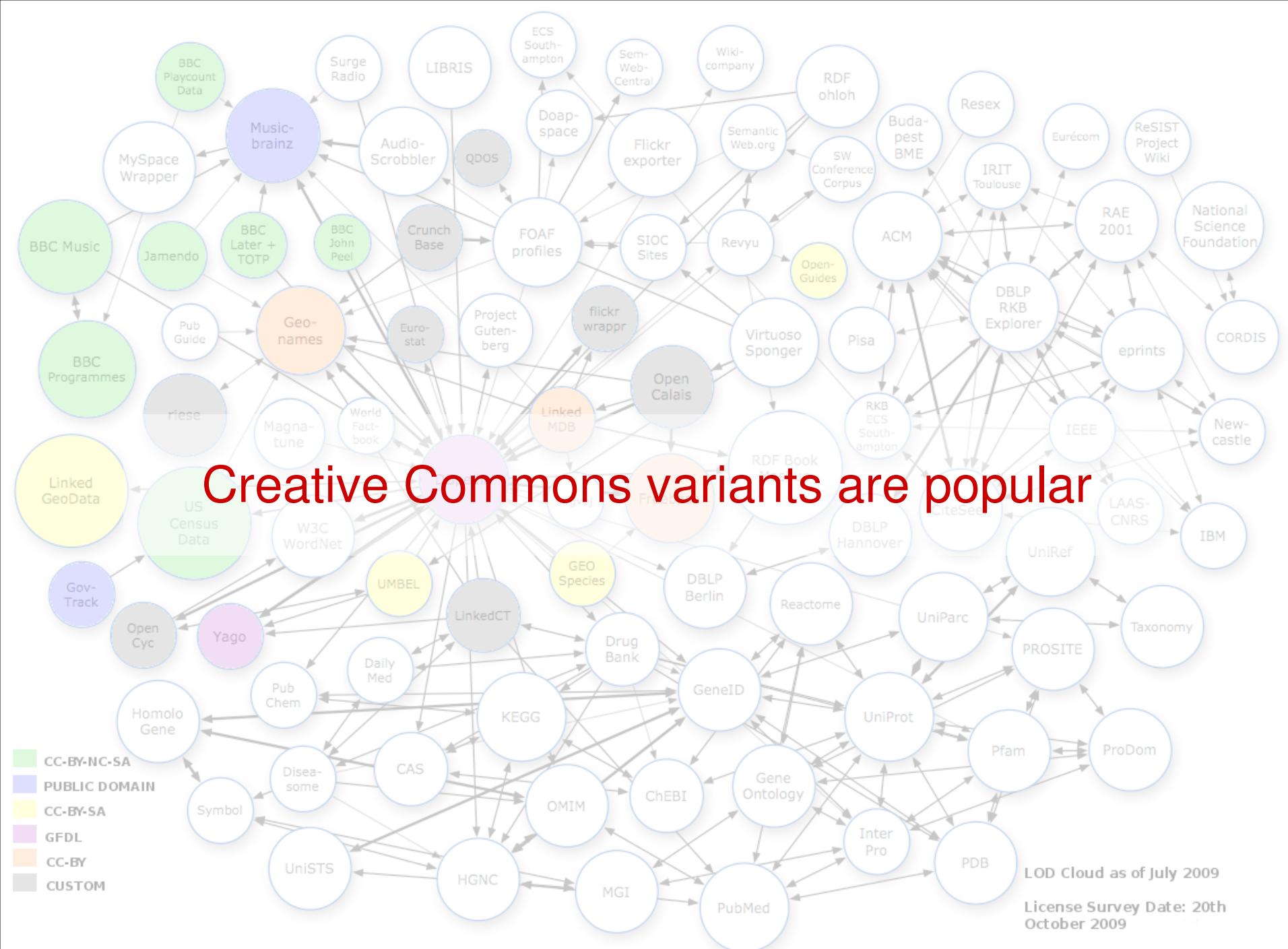


As of July 2009

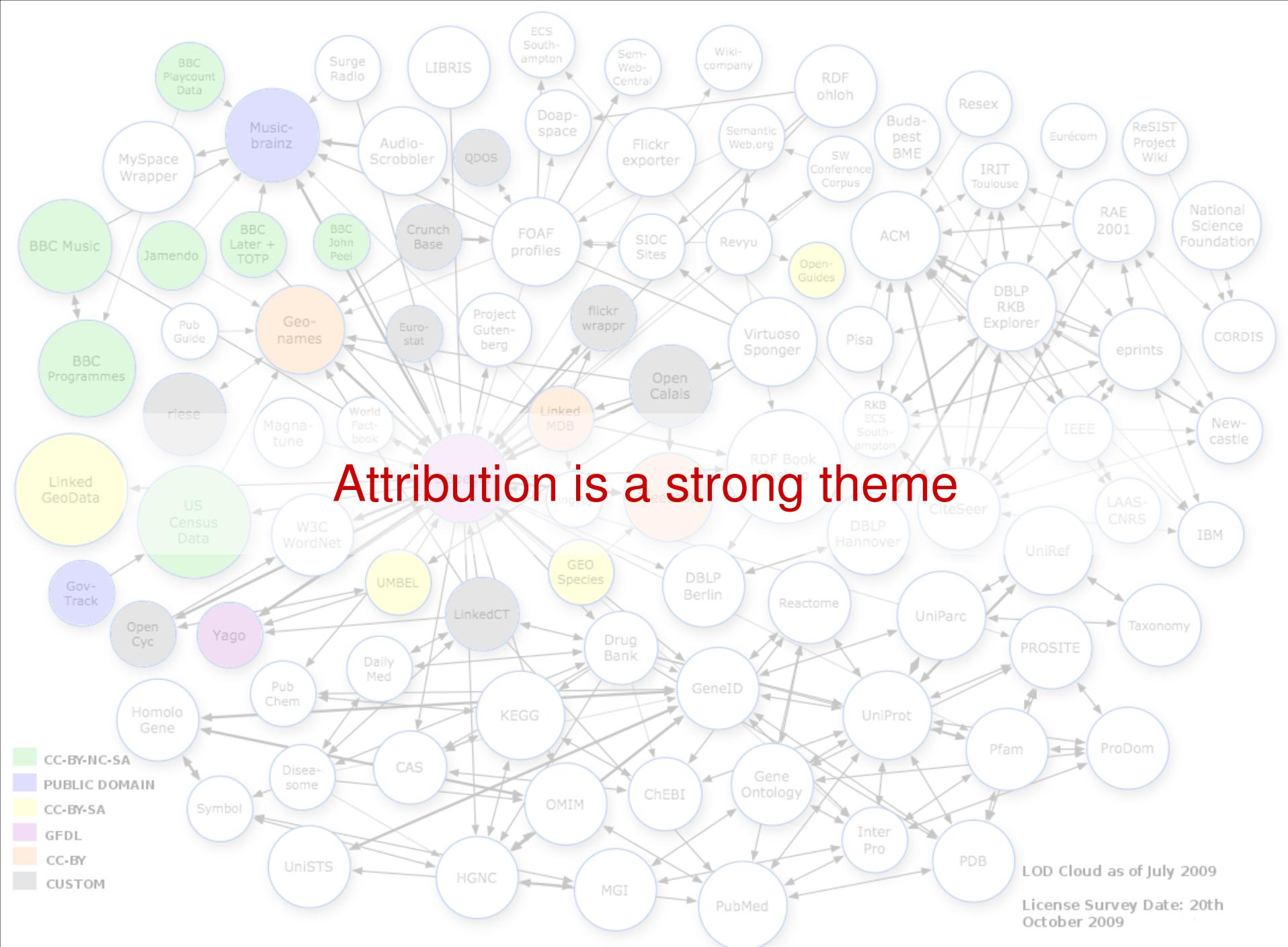




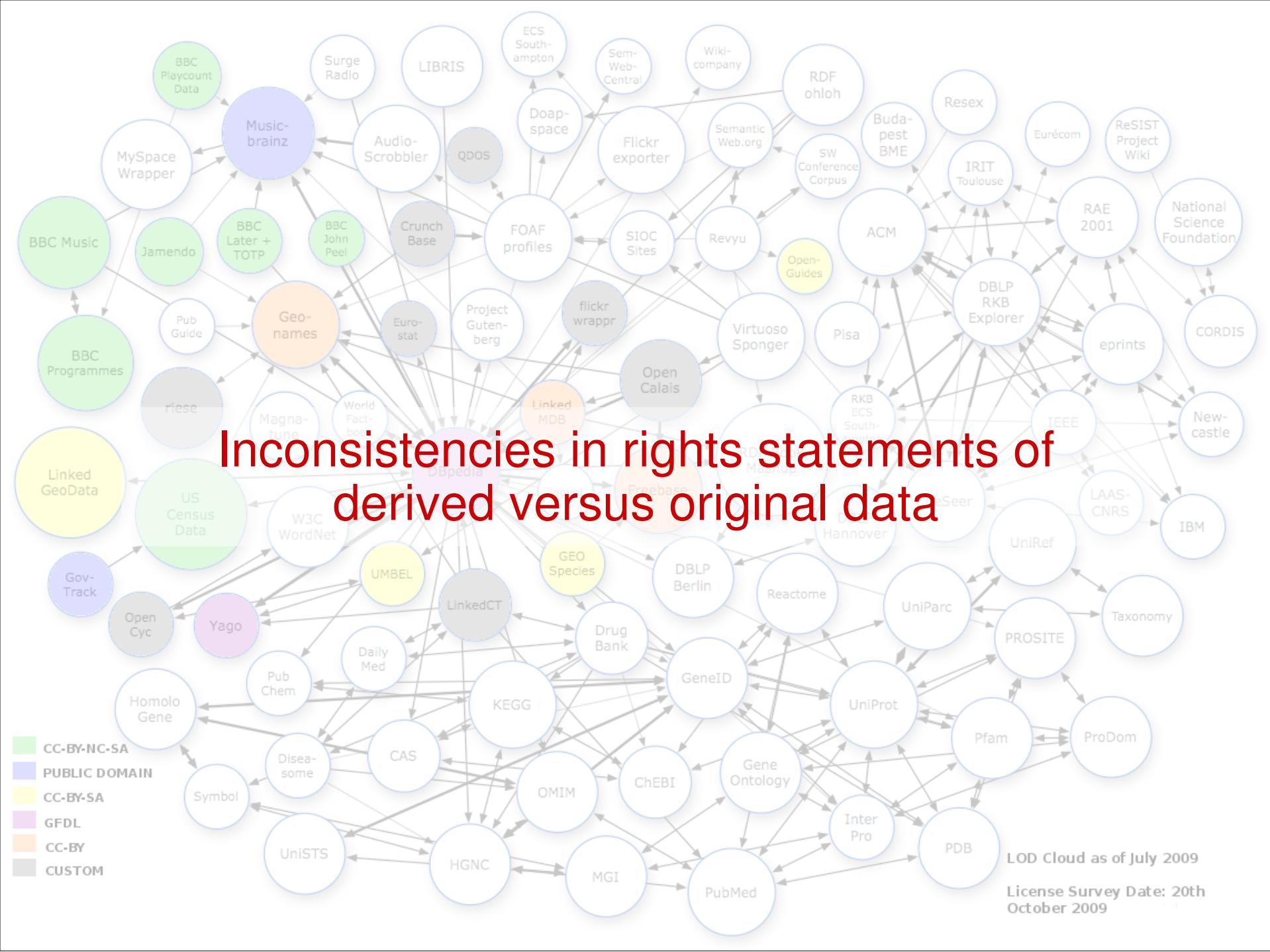
# Creative Commons variants are popular

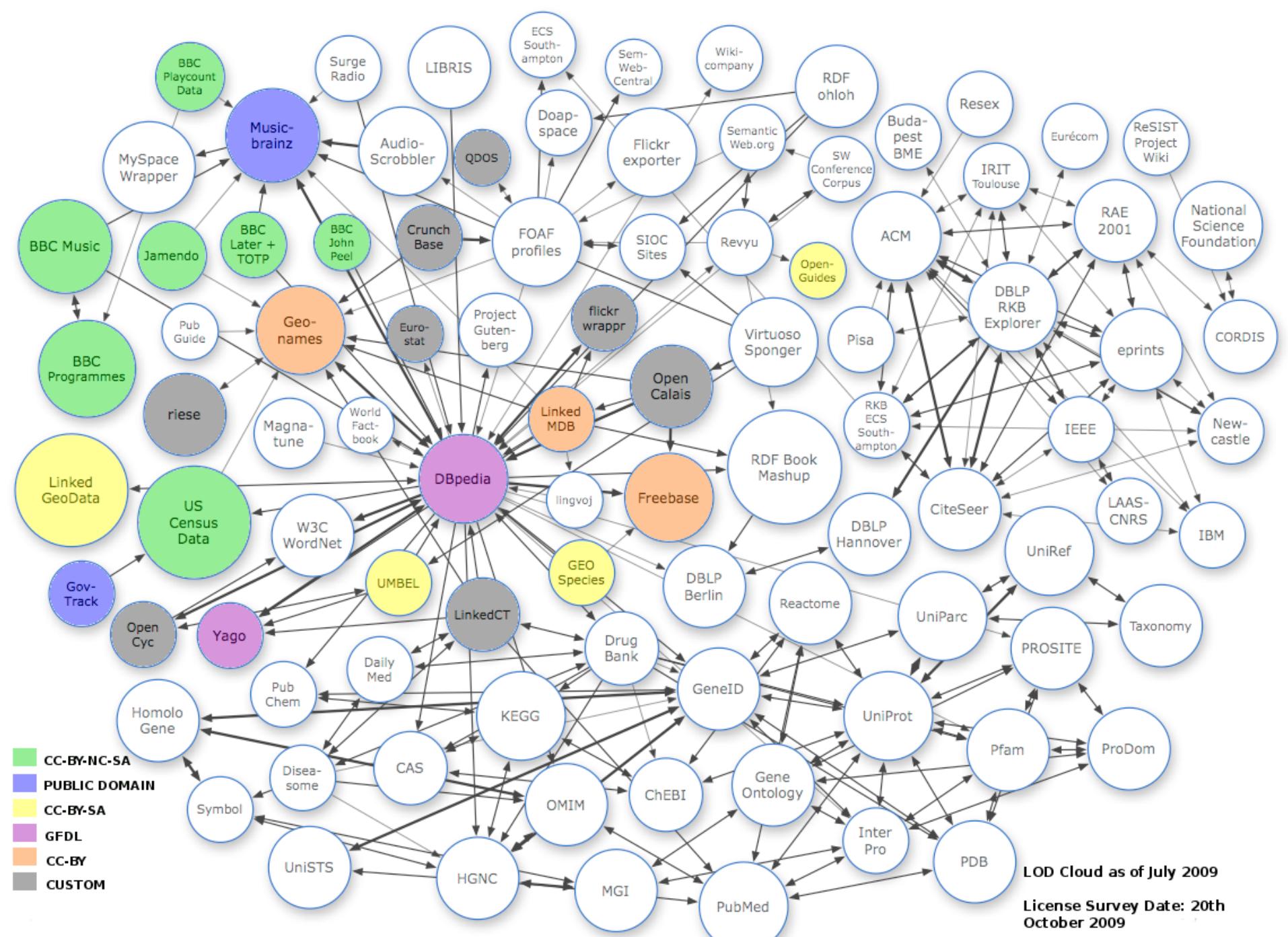


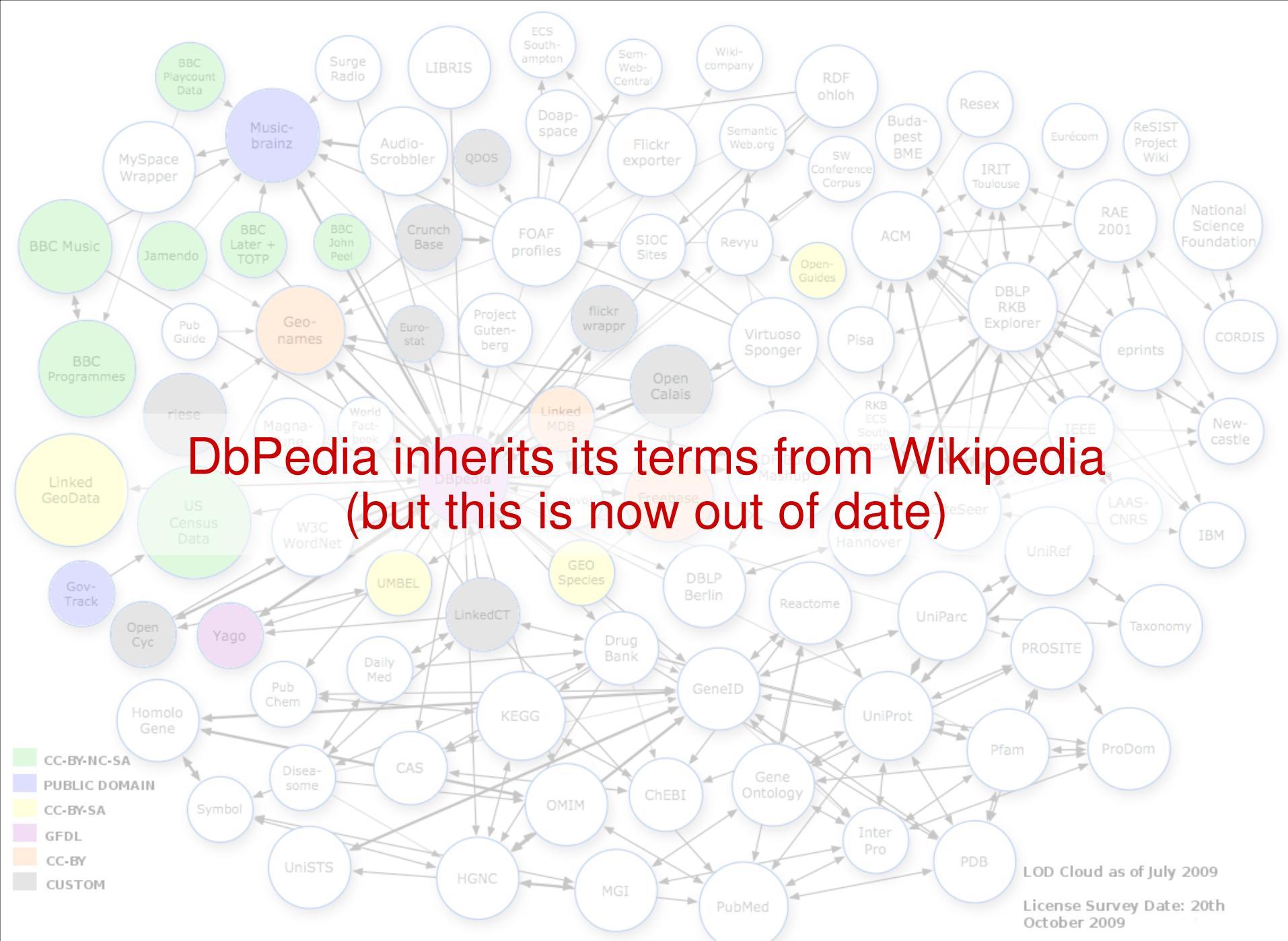
# Attribution is a strong theme

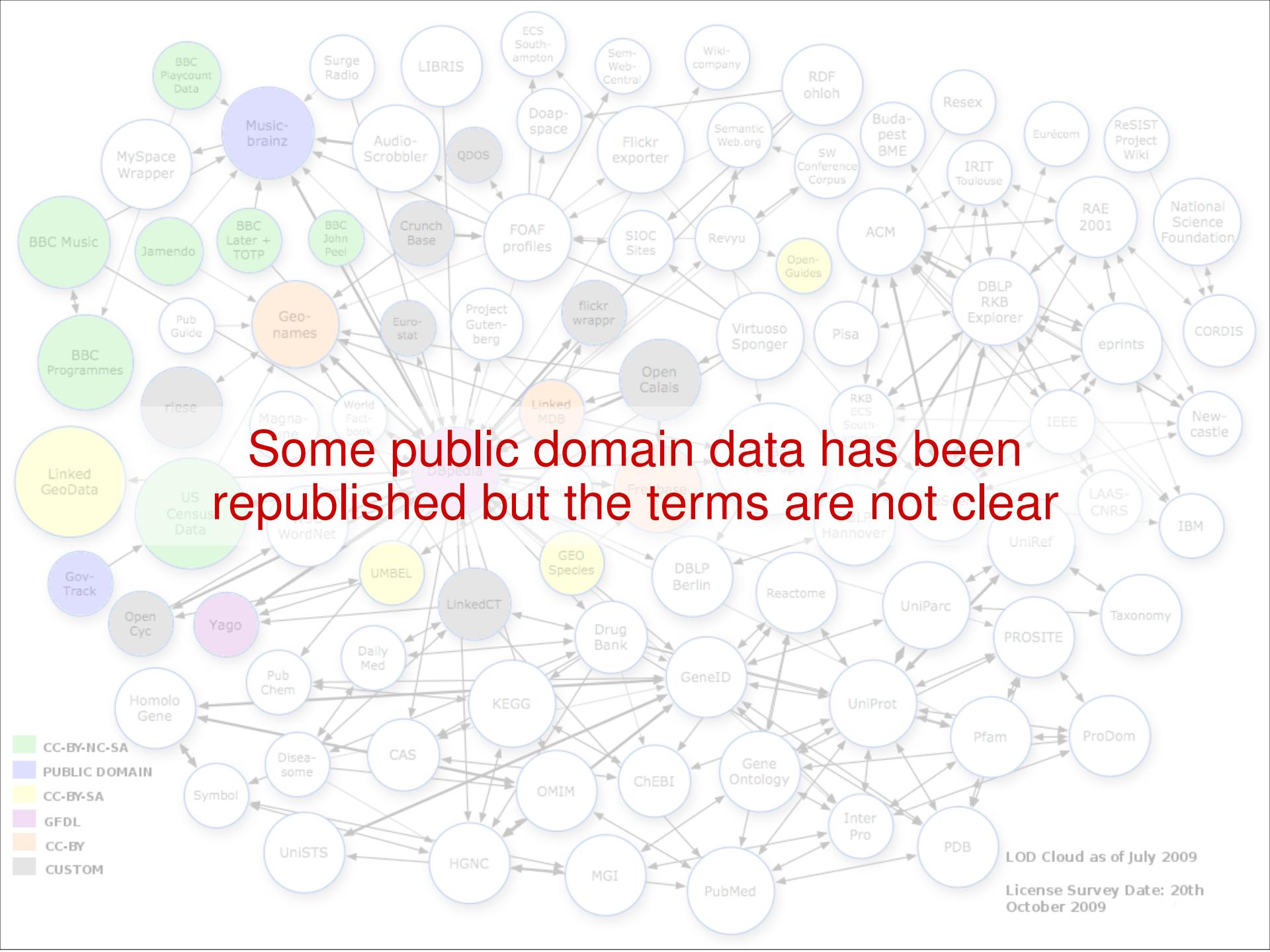


# Inconsistencies in rights statements of derived versus original data









Some public domain data has been republished but the terms are not clear

This diagram illustrates the complex relationships between various LOD Cloud entities. Entities are represented by circles, and connections between them are shown as arrows. The color of the entities indicates their license type:

- CC-BY-NC-SA (light green)
- PUBLIC DOMAIN (light purple)
- CC-BY-SA (light yellow)
- GFDL (light pink)
- CC-BY (light orange)
- CUSTOM (light grey)

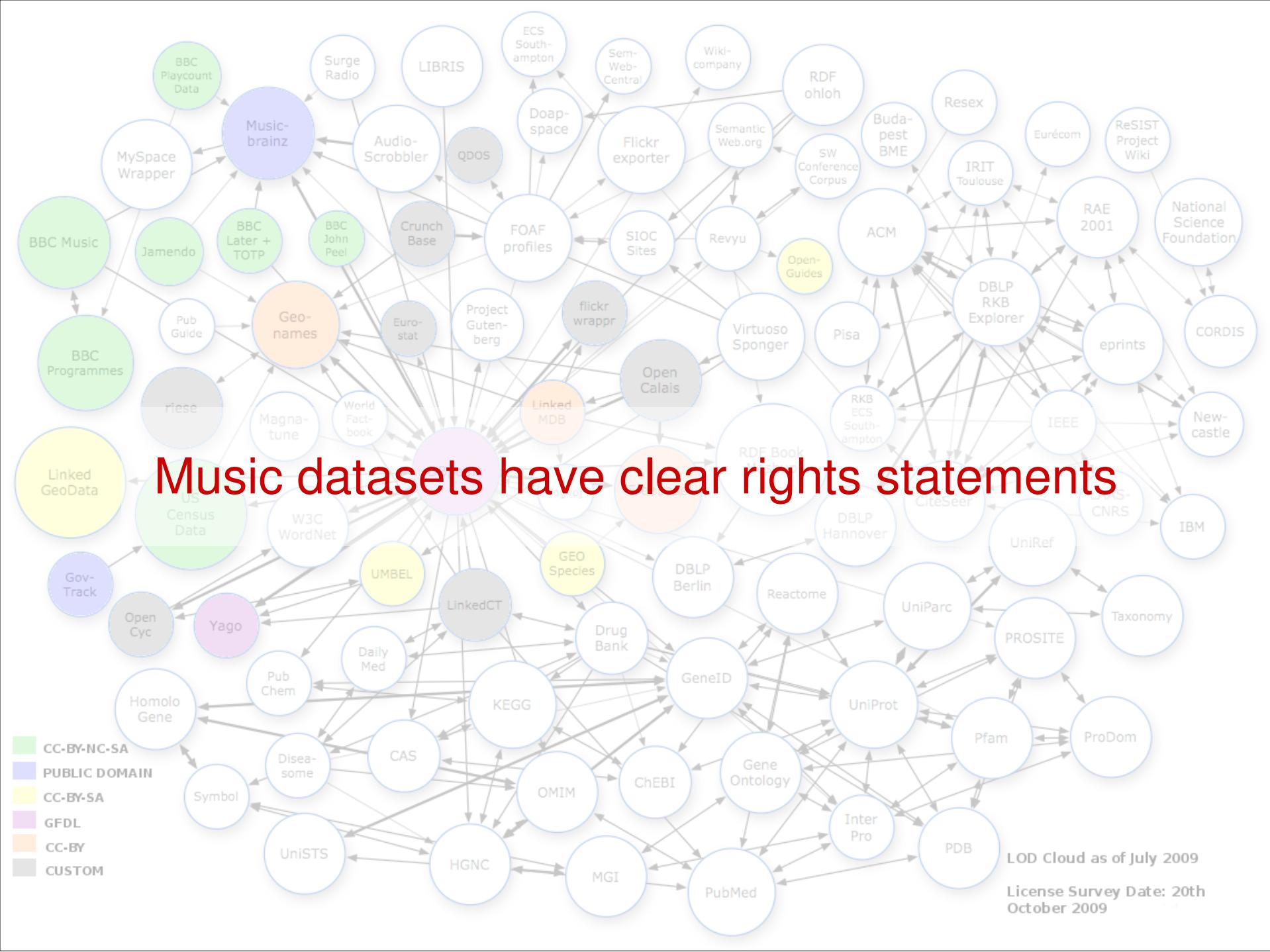
The network includes many well-known datasets and services such as DBpedia, Wikidata, GeoNames, DBLP, UniProt, and various RDF exporters and crawlers.

- CC-BY-NC-SA
- PUBLIC DOMAIN
- CC-BY-SA
- GFDL
- CC-BY
- CUSTOM

LOD Cloud as of July 2009

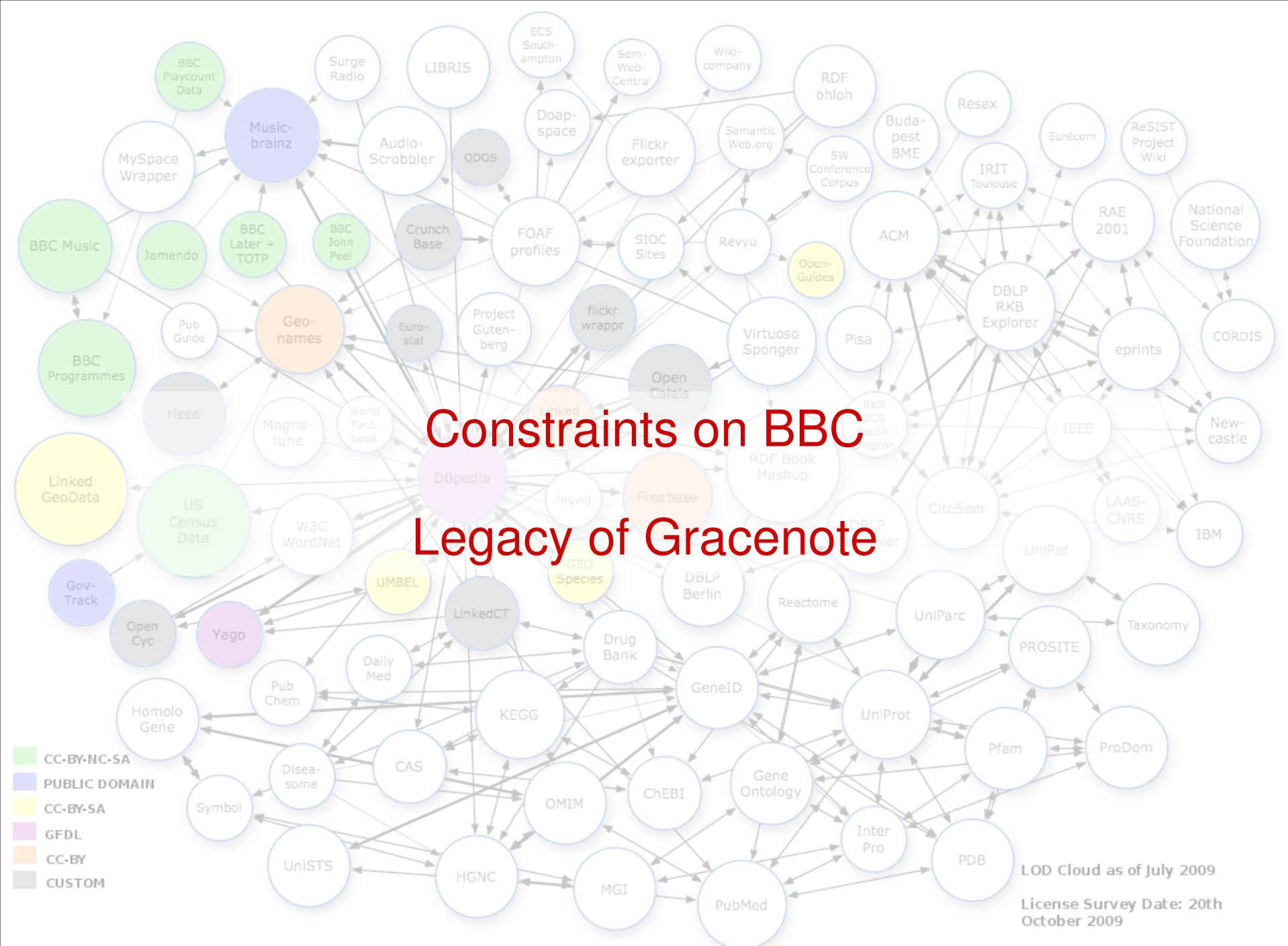
License Survey Date: 20th October 2009

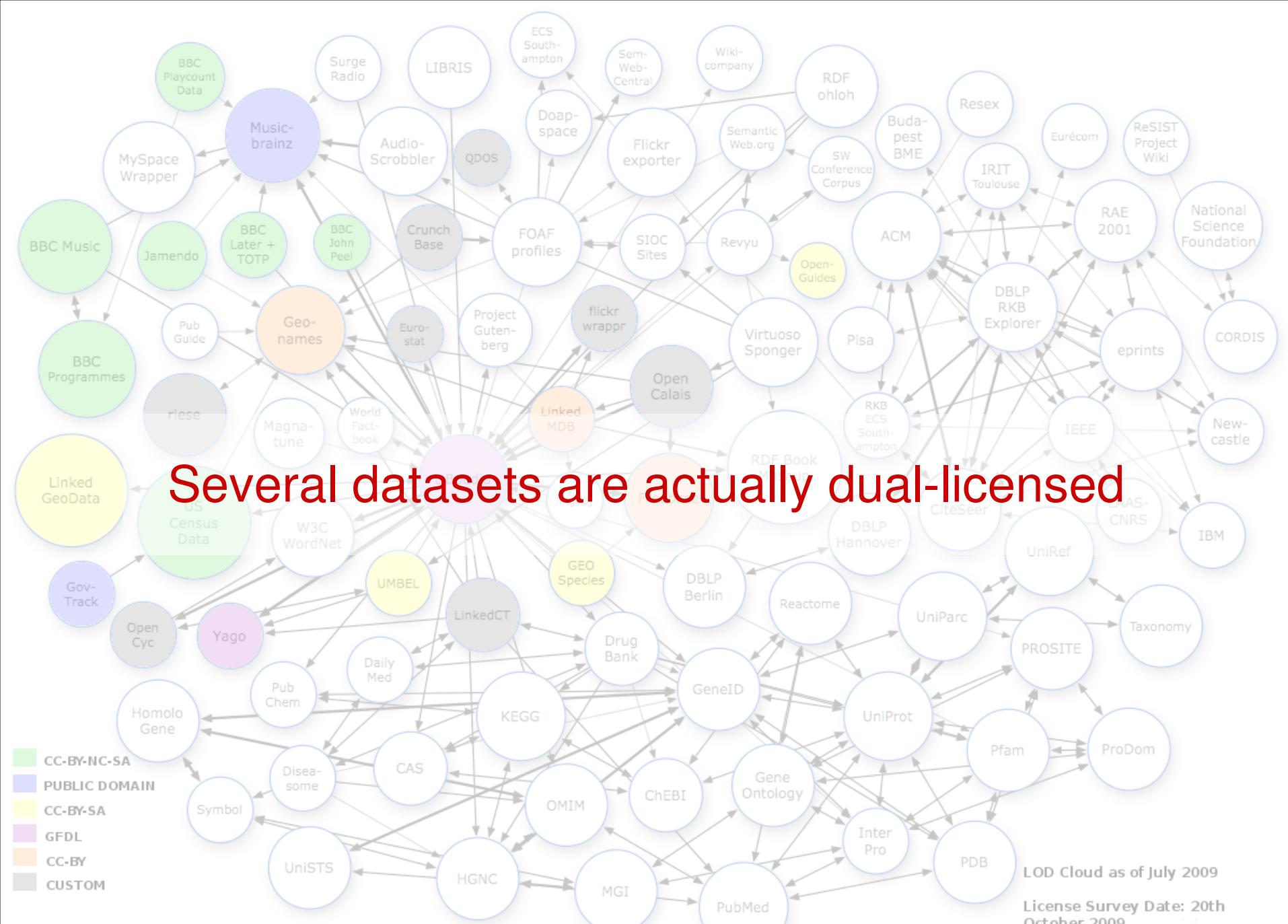
# Music datasets have clear rights statements



# Constraints on BBC

## Legacy of Gracenote





# Several datasets are actually dual-licensed



How do we solve these issues?

CHANGE

shared innovation