Linked Open Data

An introduction

Dr P Sreenivasa Kumar

Professor, CSE Dept, IIT Madras

Chennai, India

What is Linked Open Data?

- Linked Open Data (LOD)

  Data published on the Web such that
  - It is machine-readable – RDF
  - Its meaning is explicitly defined – OWL-DL
  - It is linked to other external data sets
  - And it can in turn be linked to from external data sets

Linked Data - The story so far,
Christian Bizer, Tom Heath
and Tim Berners-Lee, 2009

Linked Open Data

- A recommended best practice for
  - Publishing and
  - Connecting structured data on the Web
  - Adopted by an increasing number of data providers over the past several years
- Leading to
  - Global data space containing billions of assertions
  - The Web of Data
  - Smart data integration

Linked Data Principles

1. Use URIs as names for things.
2. Use HTTP so that people can look up those names.
3. When someone looks up a URI, provide useful RDF information.
4. Include RDF statements that link to other URIs so that they can discover related things.

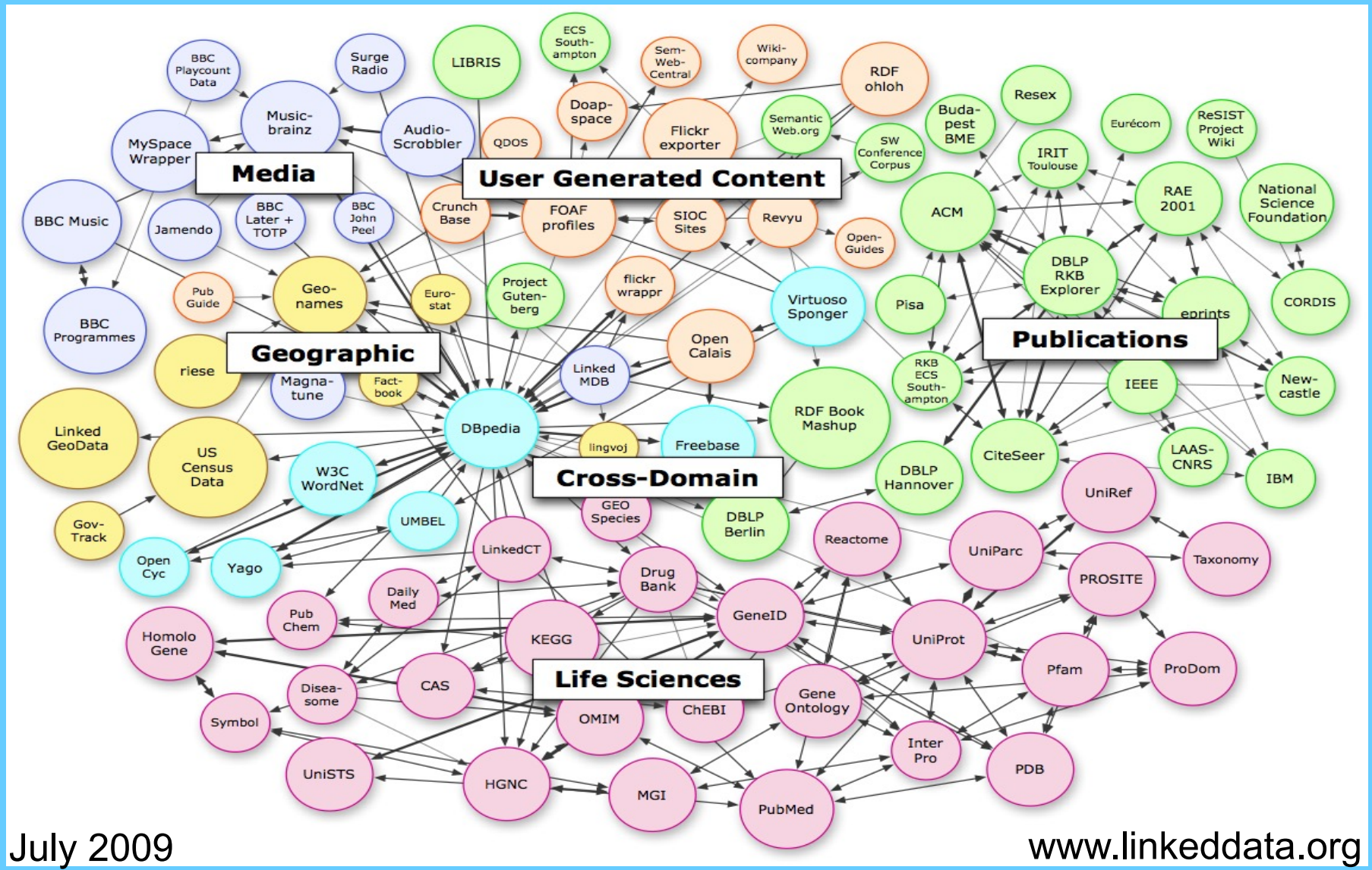
Tim Berners-Lee 2007

http://www.w3.org/DesignIssues/LinkedData.html

Properties of Linked Data

- Anyone can publish data to the Web of Linked Data
  - may include disagreement; contradictory data also
- Entities are connected by links
  - creating a global data graph that spans data sources and enables the discovery of new data sources.
- Data is self-describing
  - unfamiliar vocabulary terms found by an application
    - resolved using URIs of terms in order to find their RDFS or OWL definition.
- The Web of Data is Open
  - meaning that applications can discover new data sources at run-time by following links

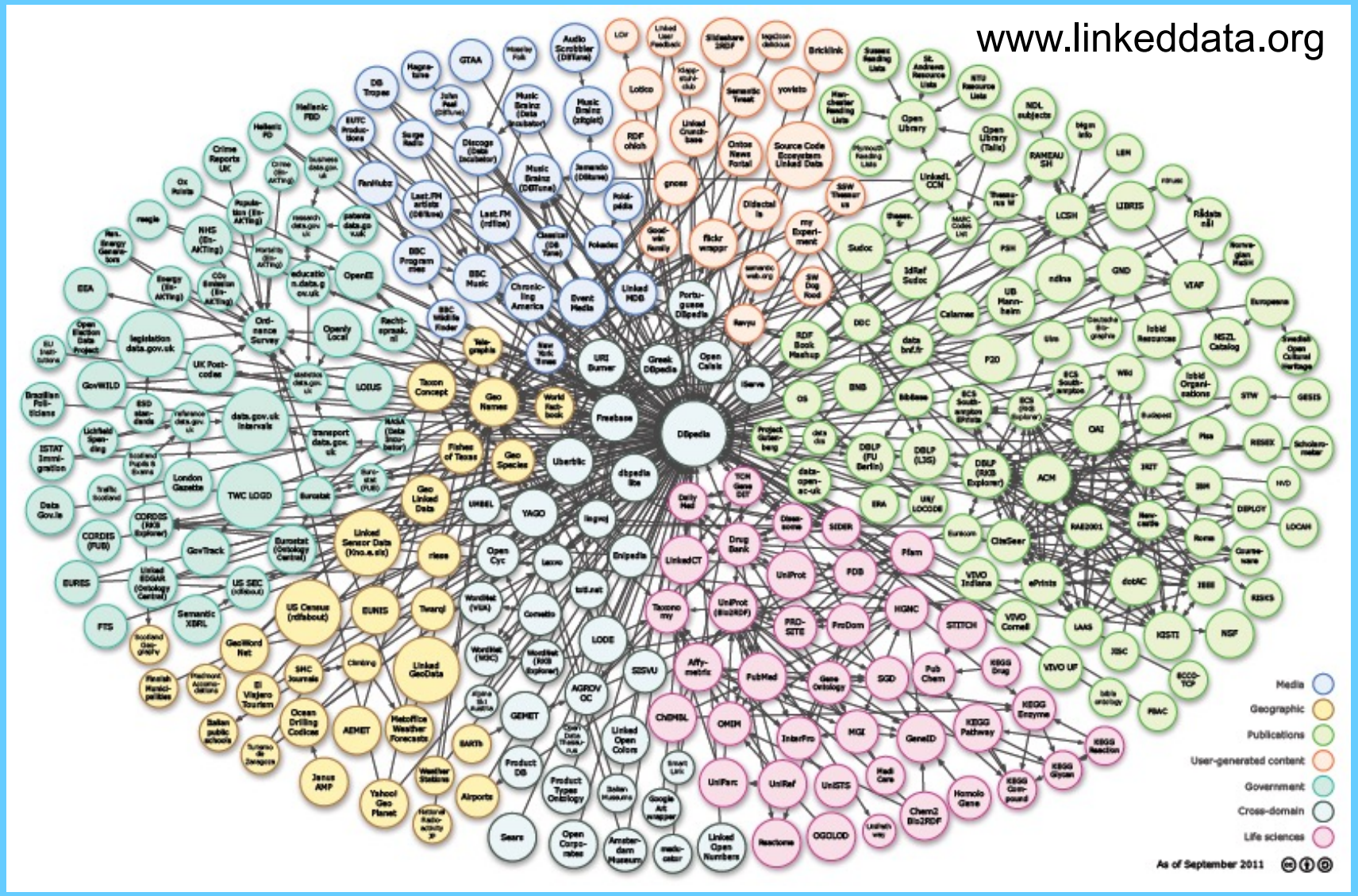
Linked Open Data(LOD) Cloud



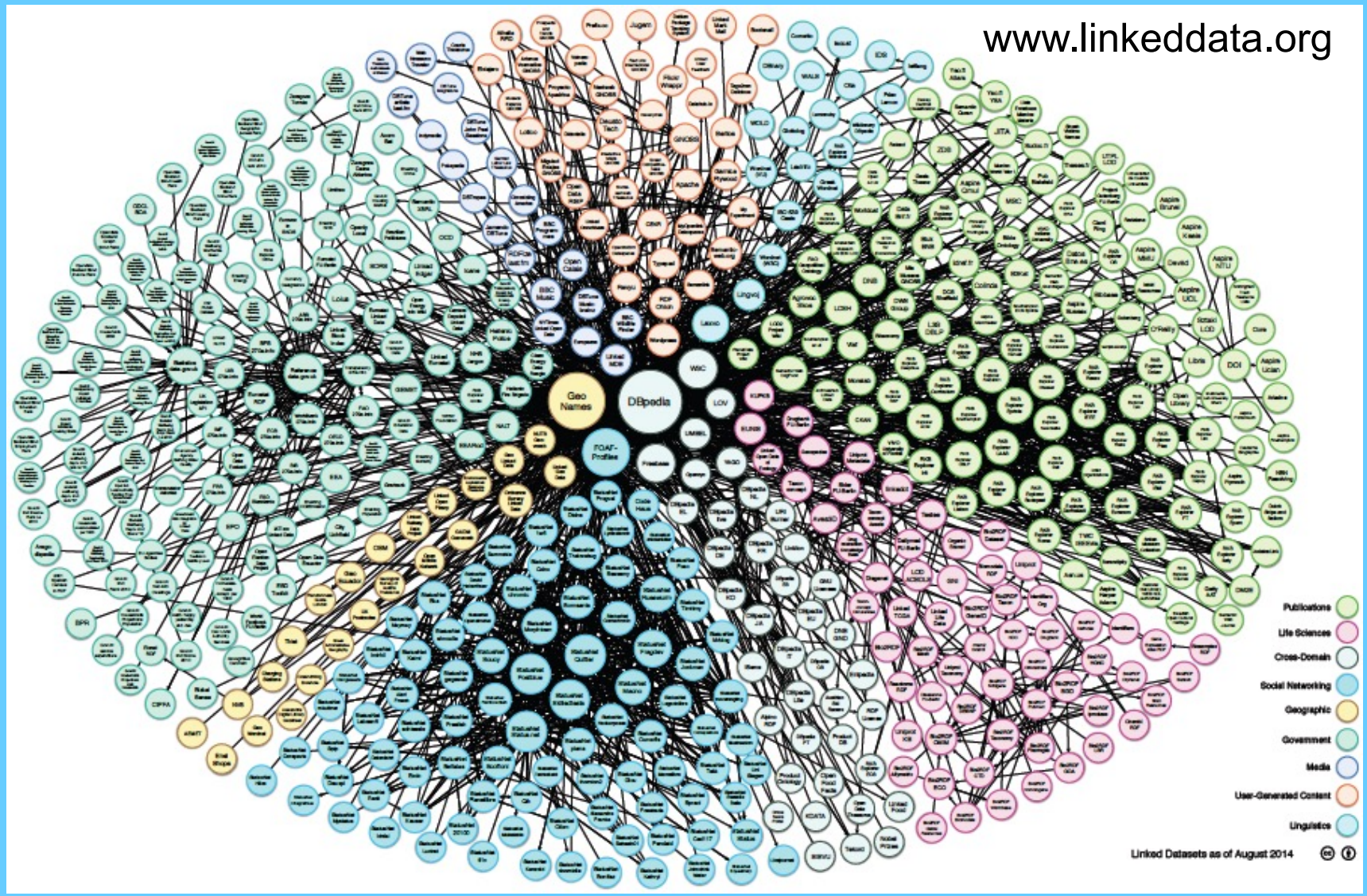
July 2009

www.linkeddata.org

LOD Cloud – Sep 2010

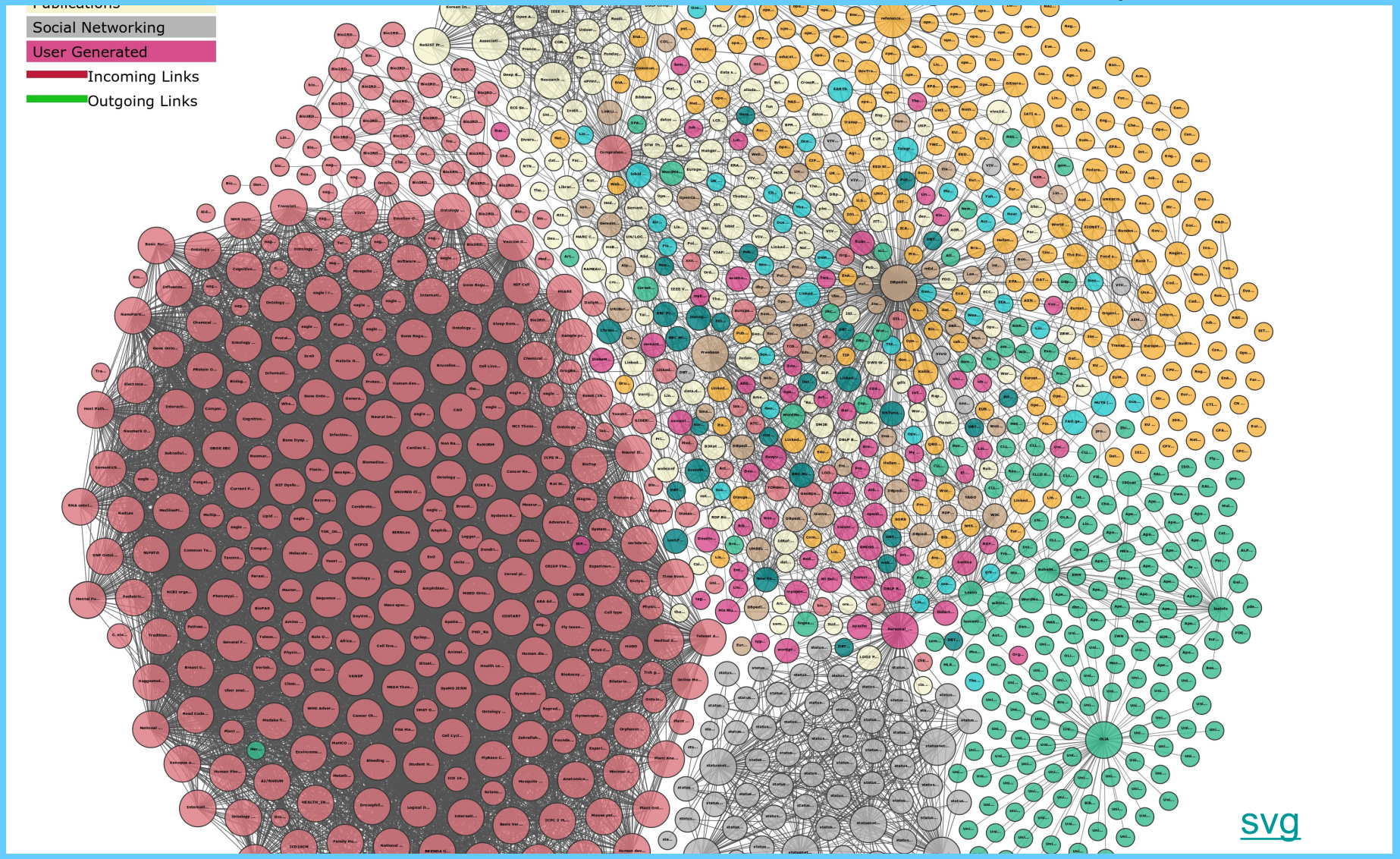


LOD Cloud – Aug 2014



LOD August 2017

http://lod-cloud.net/



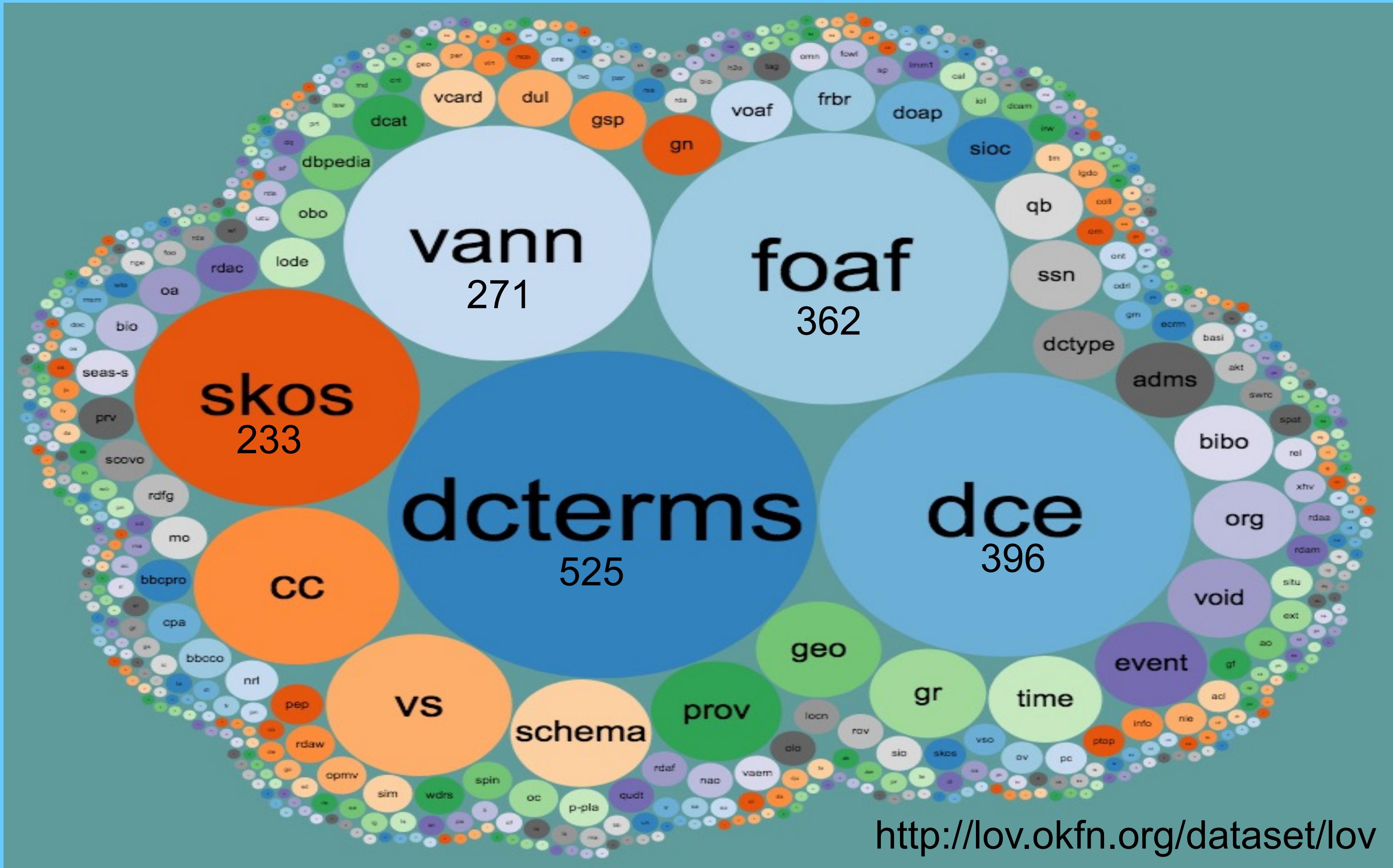
LOD Datasets Statistics - 2017

1139 datasets

http://stats.lod2.eu/stats

Criterion	Average	Min	Max	Median	Total
Triples	67,544.15	0	4,70,54,407	337	19,22,30,648
Entities	18,105.28	0	93,19,918	80	5,42,25,309
Literals	30,137.45	0	3,14,76,008	166	9,02,61,655
Blanks	3,554.83	0	35,65,513	0	1,06,46,711
Blanks as subject	1,742.85	0	19,10,532	0	52,19,831
Blanks as object	1,812.01	0	35,64,789	0	54,26,969
Subclasses	1.6	0	2,000	0	4,779
Typed subjects	7,387.12	0	69,90,722	39	2,21,24,421
Labeled subjects	1,219.97	0	14,40,595	0	36,53,811
Average properties per entity	4.98	0	91.16	3.71	
Average class hierarchy depth	3.24	1	9	None	
Links	15,379.59	0	1,32,52,430	57	4,60,61,873
Average property hierarchy depth	1.5	1	3	None	
Vocabularies	4.27	1	18	3	12,110
Classes	4.36	1	330	3	10,384
Properties	17.58	1	254	16	49,916

LOD Vocabularies



http://lov.okfn.org/dataset/lov

DBpedia

- DBpedia (English) http://wiki.dbpedia.org/about
  - RDF Data – auto-extracted from Wikipedia
  - Describes 4.58 million things
    - 1,445,000 persons, 735,000 places,
    - 411,000 creative works (123,000 music albums, 87,000 films and 19,000 video games),
    - 241,000 organizations (58,000 companies and 49,000 educational institutions),
    - 251,000 species and 6,000 diseases
  - ~685 classes, ~2,795 properties
- And it is growing … ~3 billion triples (2020)

DBpedia Mobile

- Based on the GPS position of the mobile
  - DBpedia Mobile provides
    - A location-centric mashup of nearby locations from DBpedia,
    - Associated reviews from Revyu, and
    - Related photos from Flickr via a Linked Data wrapper
  - Enables users to publish
    - their current location, pictures and reviews to the Web as Linked Data

Becker & Bizer, 2008

DBpedia Mobile

Becker & Bizer, 2008

BBC World

- Uses Linked Data
  - Collects, filters and re-purposes Linked Data
  - From DBpedia, MusicBrainz, WWF etc
- Auto generates web presences for
  - Programs(http://www.bbc.co.uk/programmes)
    - more than 1,500 daily TV and radio broadcasts
  - Music (http://www.bbc.co.uk/music)
  - Wildlife Finder (http://www.bbc.co.uk/nature/wildlife)

[Linked Data – Wood et al, 2014]

References

- Semantic Web for the Working Ontologist
  - Dean Allemang and Jim Hendler
  - Second Edition, Morgan Kaufmann, 2011
- Linked Data: Evolving the Web into a Global Data Space, Tom Heath and Christian Bizer
  - http://linkeddatabook.com/editions/1.0/
- Linked Data
  - David Wood, Marsha Zaidman, Luke Ruth and Michael Hausenblas, Manning Publications, 2014
- http://www.w3.org/