Semantic e-Science
in the
Bibliographic Cloud

Connecting the data with the literature

~~Semantic~~ e-Science

in the

Bibliographic Cloud

Connecting the data with the literature

~~Semantic~~ e-Science

in the

~~Bibliographic Cloud~~

Connecting the data with the literature

~~Semantic~~ e-Science

in the

~~Bibliographic Cloud~~

Connecting the data with the literature

Connecting the data with the literature

(using Linked Data)

Image: ariz, http://flickr.com/photos/ariz/

Linked Data

Linked Data is about using the Web to connect related data that wasn't previously linked, or using the Web to lower the barriers to linking data currently linked using other methods. More specifically, Wikipedia defines Linked Data as "a term used to describe a recommended best practice for exposing, sharing, and connecting pieces of data, information, and knowledge on the Semantic Web using URIs and RDF."

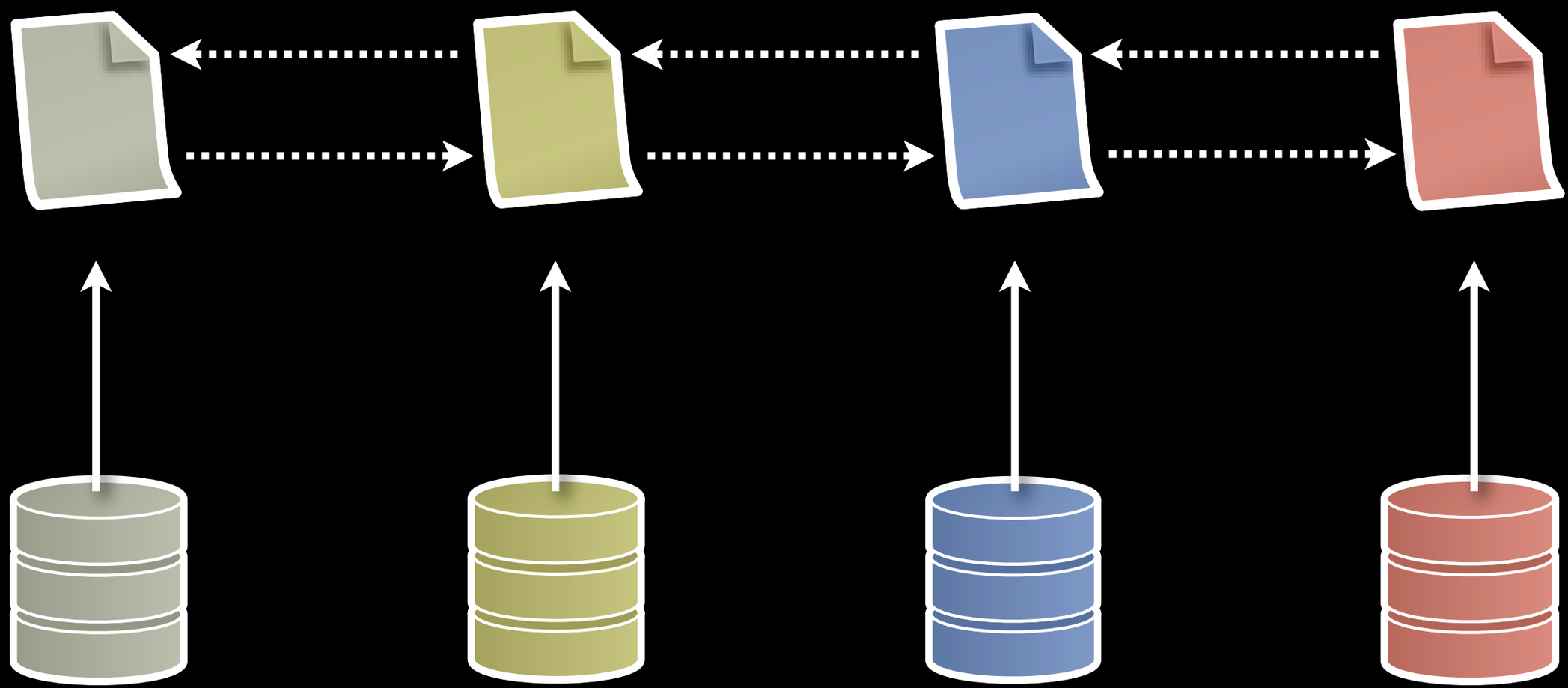
Linked Data

using the Web to connect related data that wasn't previously linked

The Web of Documents

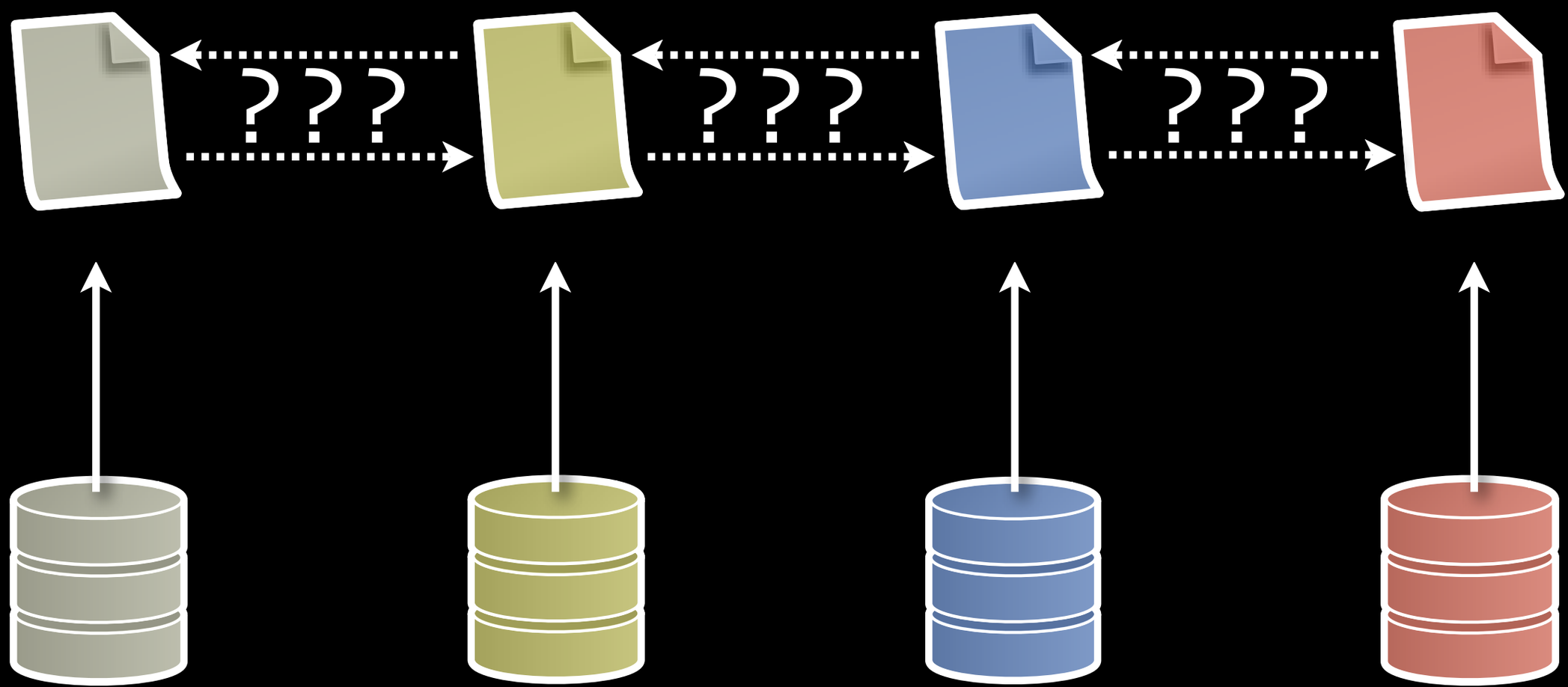
- A global filesystem
- Designed to be human-readable
- Documents are primary objects
- Links are between documents
- Link semantics are implicit

The Web of Documents



The Web of Documents

Problem:
How are these documents related?



The Web of Documents

"What's the favorite recording artist of all famous people born in the city of West Lafayette who are depicted in this photo?"

Image: Michael Stephens, http://flickr.com/photos/mstephens7/

Limitations

- Disconnected data (many silos)
- Lack of structure
- Duplication across documents
- Difficult to integrate documents
- Can't execute complex queries

Image: zoomzoom, http://flickr.com/photos/zoomzoom/

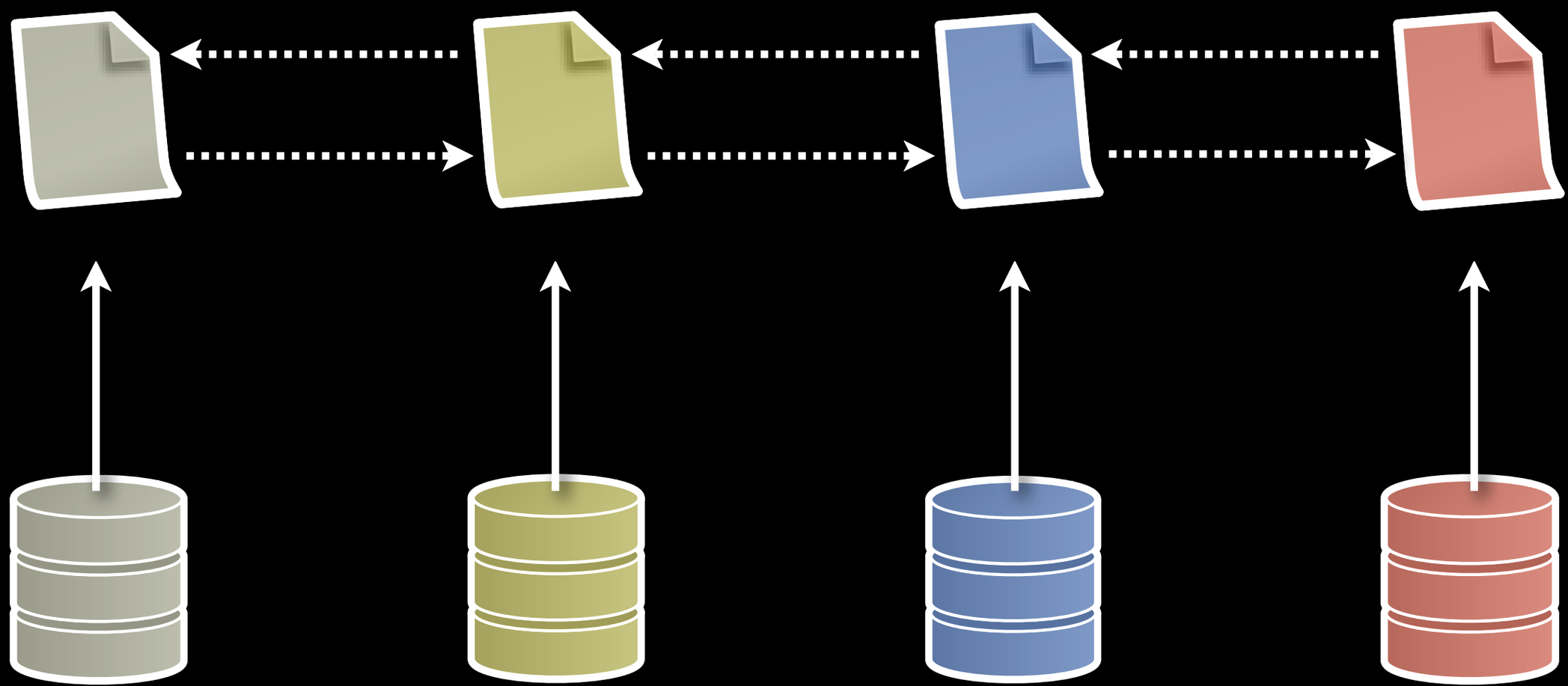
Linked Data Principles

- Use URIs as names for things.
- Use HTTP URIs so people can look up those names.
- When someone looks up a URI, provide useful information.
- Include links to other URIs so they can discover more things.

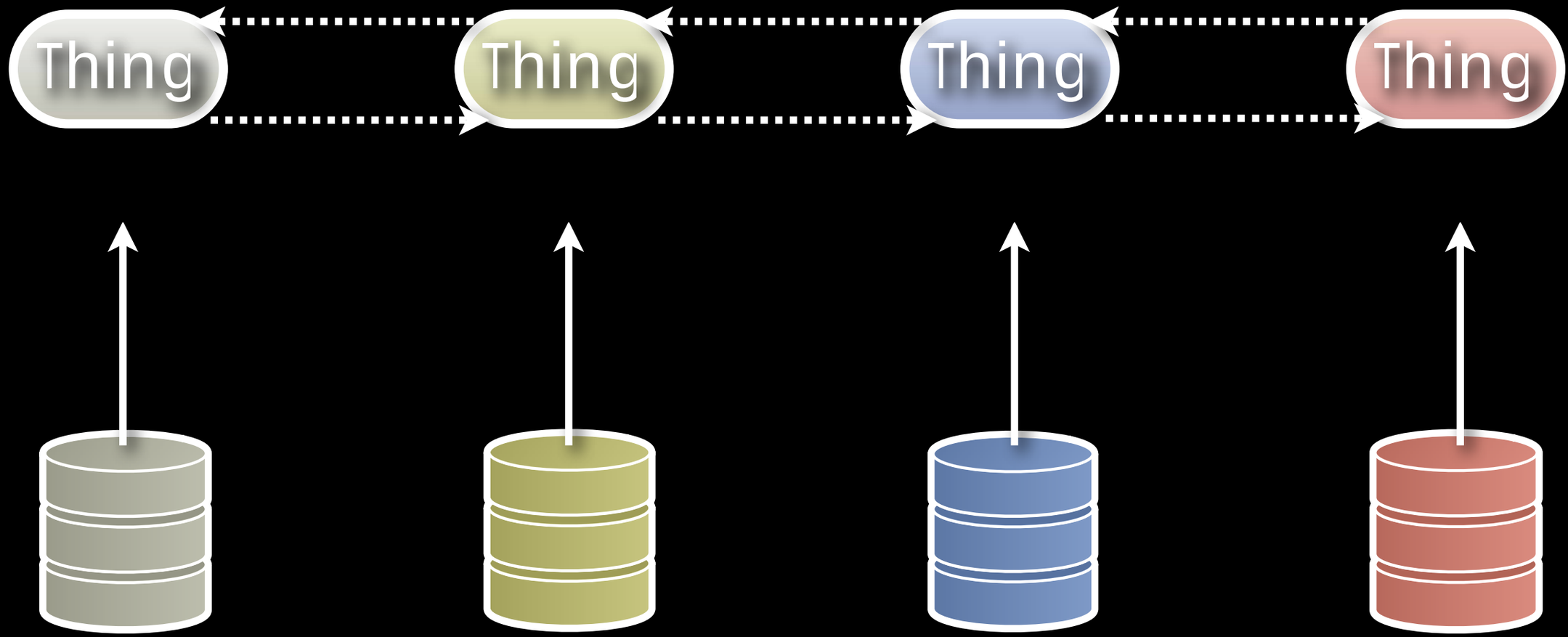
The Web of Data

- A global database
- Designed to be machine-readable
- Primary objects are *things*
- Links are between *things*
- Link semantics are explicit

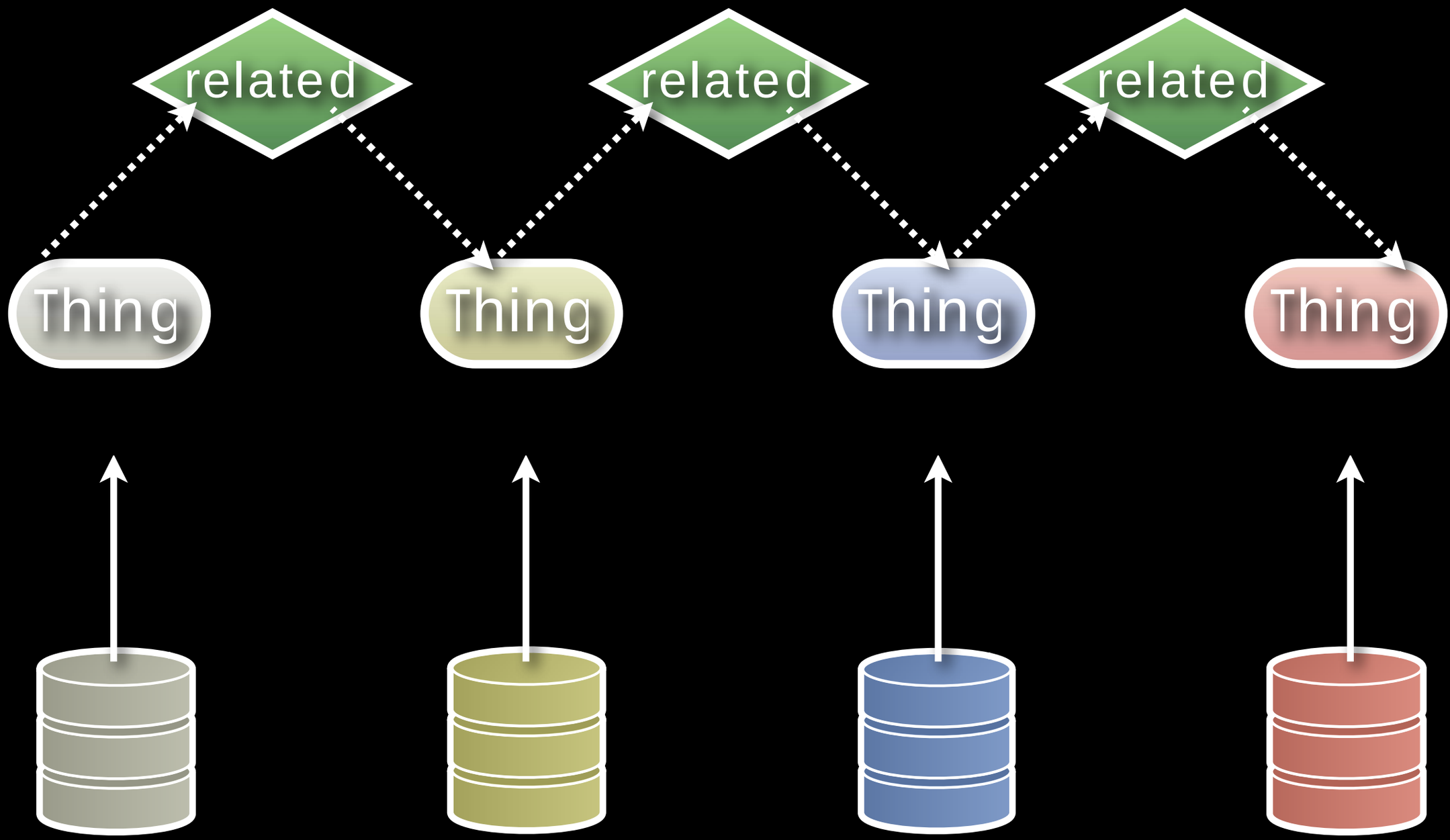
The Web of Data



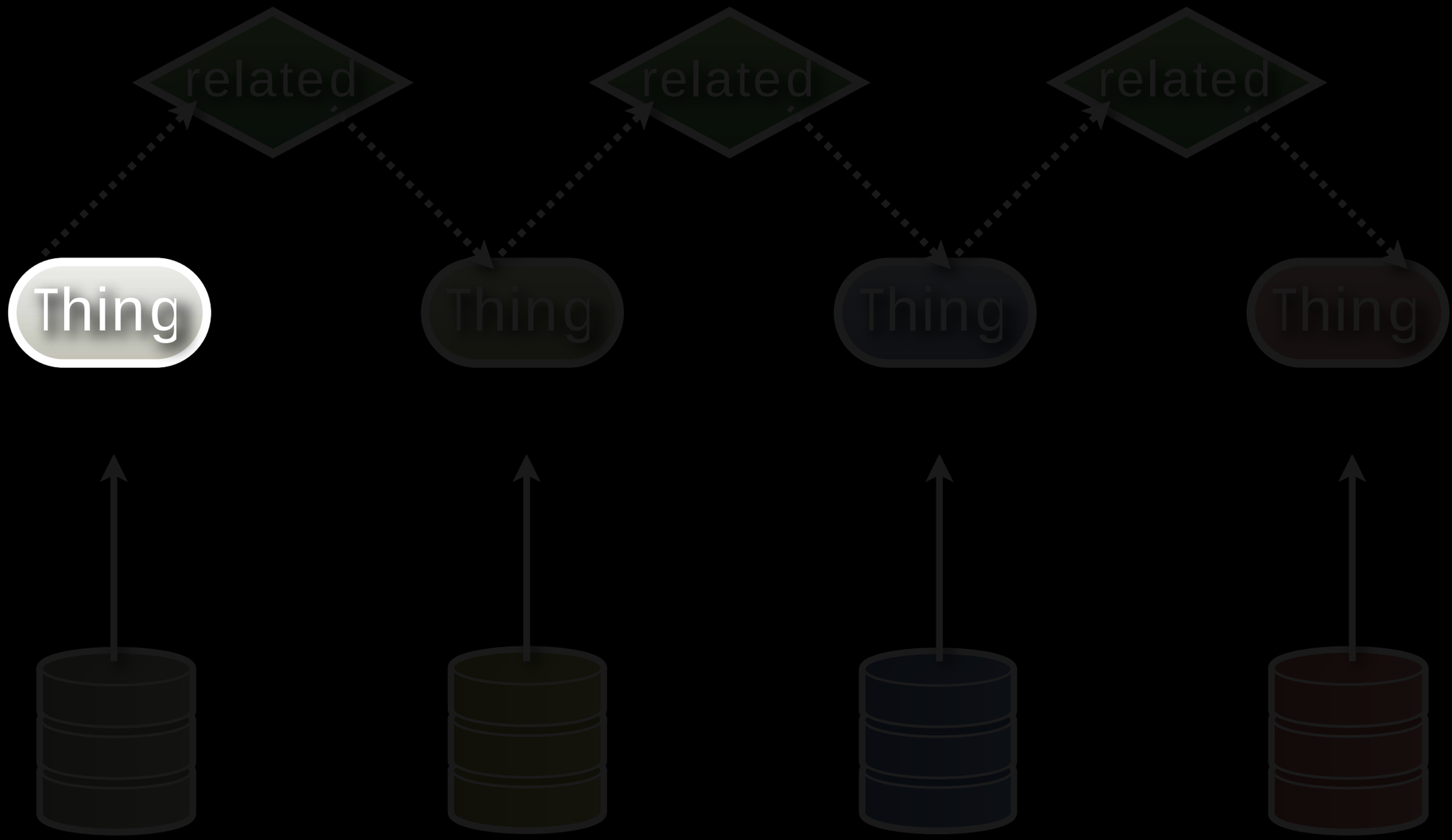
The Web of Data



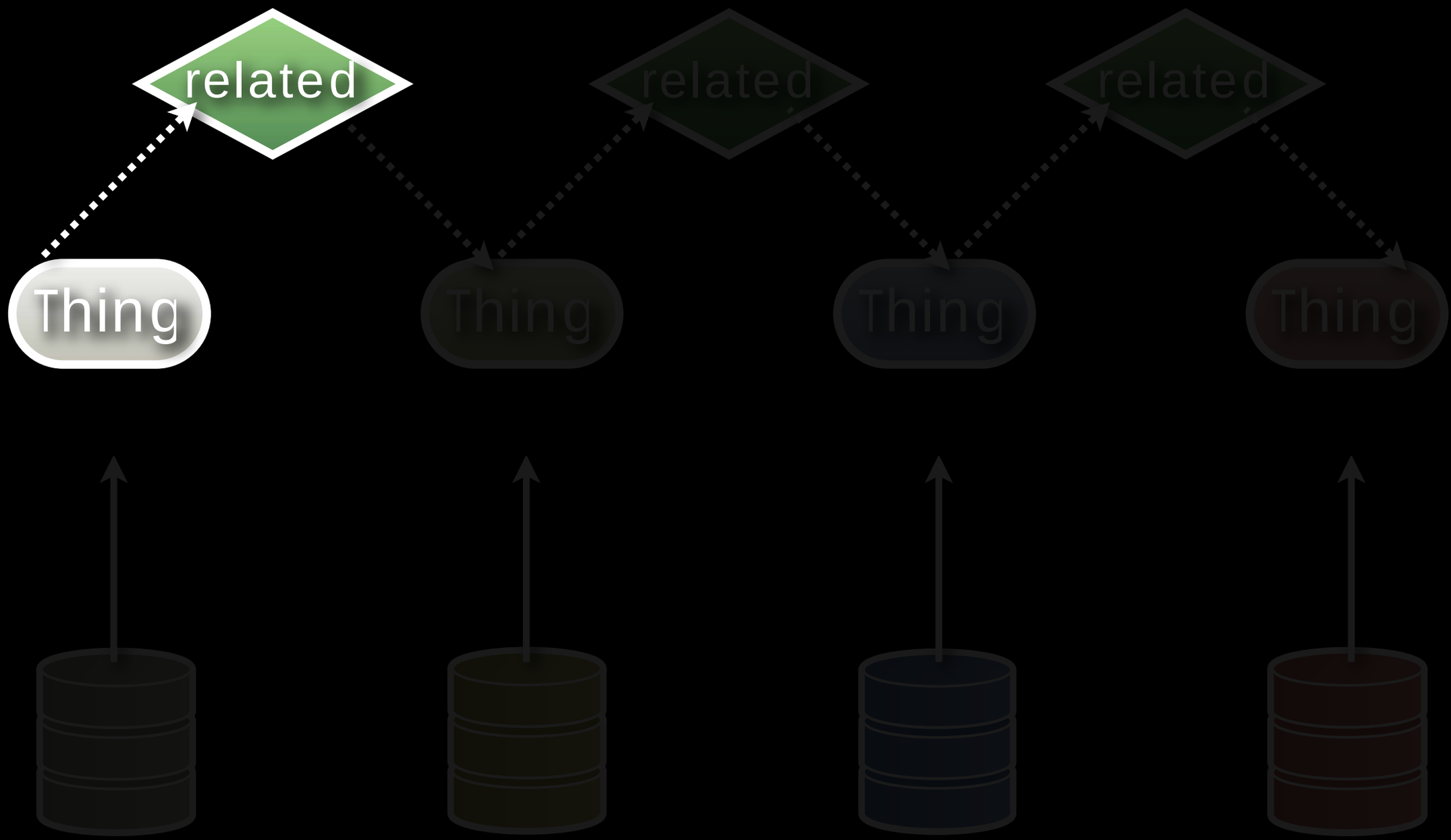
The Web of Data



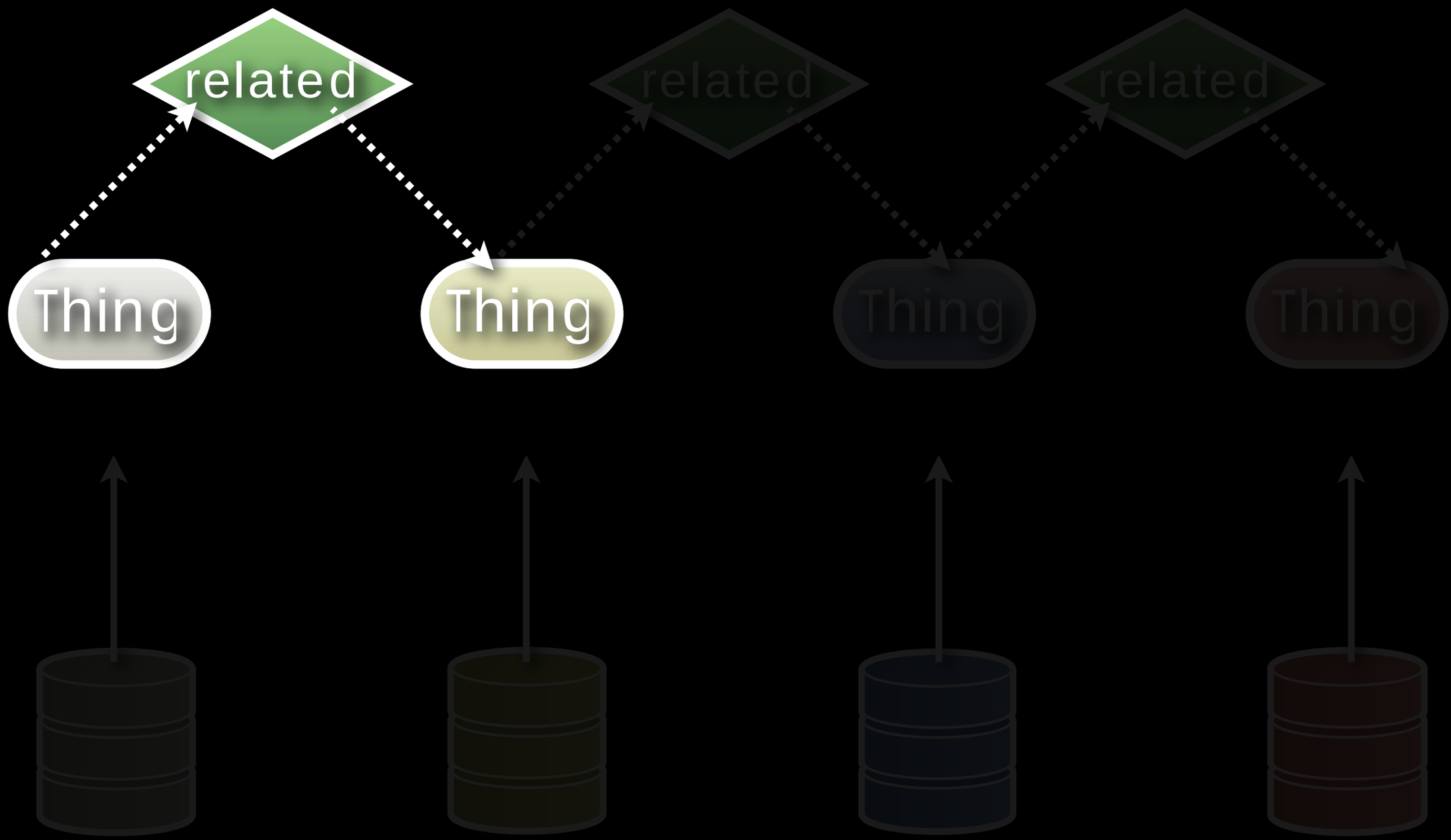
The Web of Data



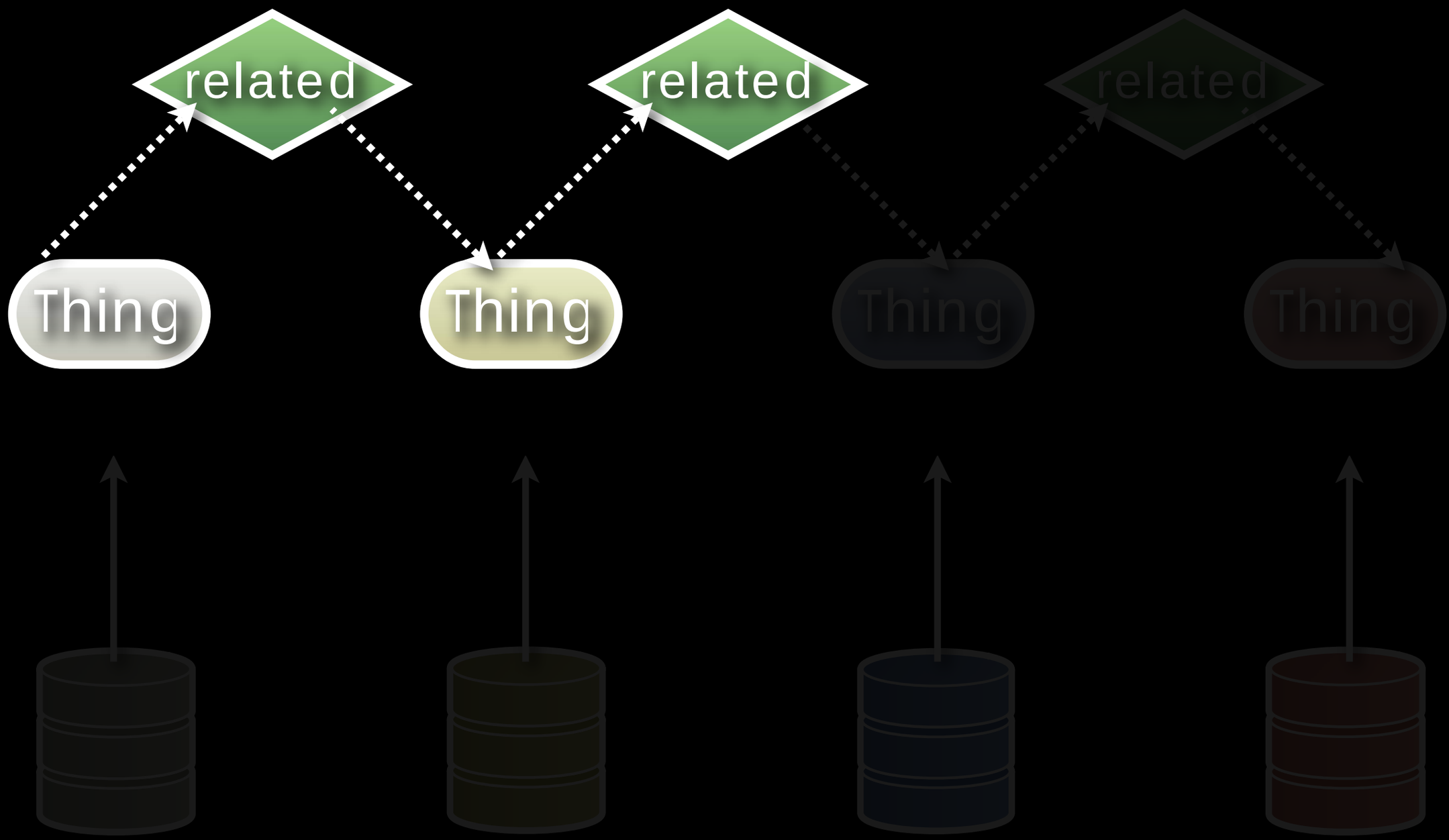
The Web of Data



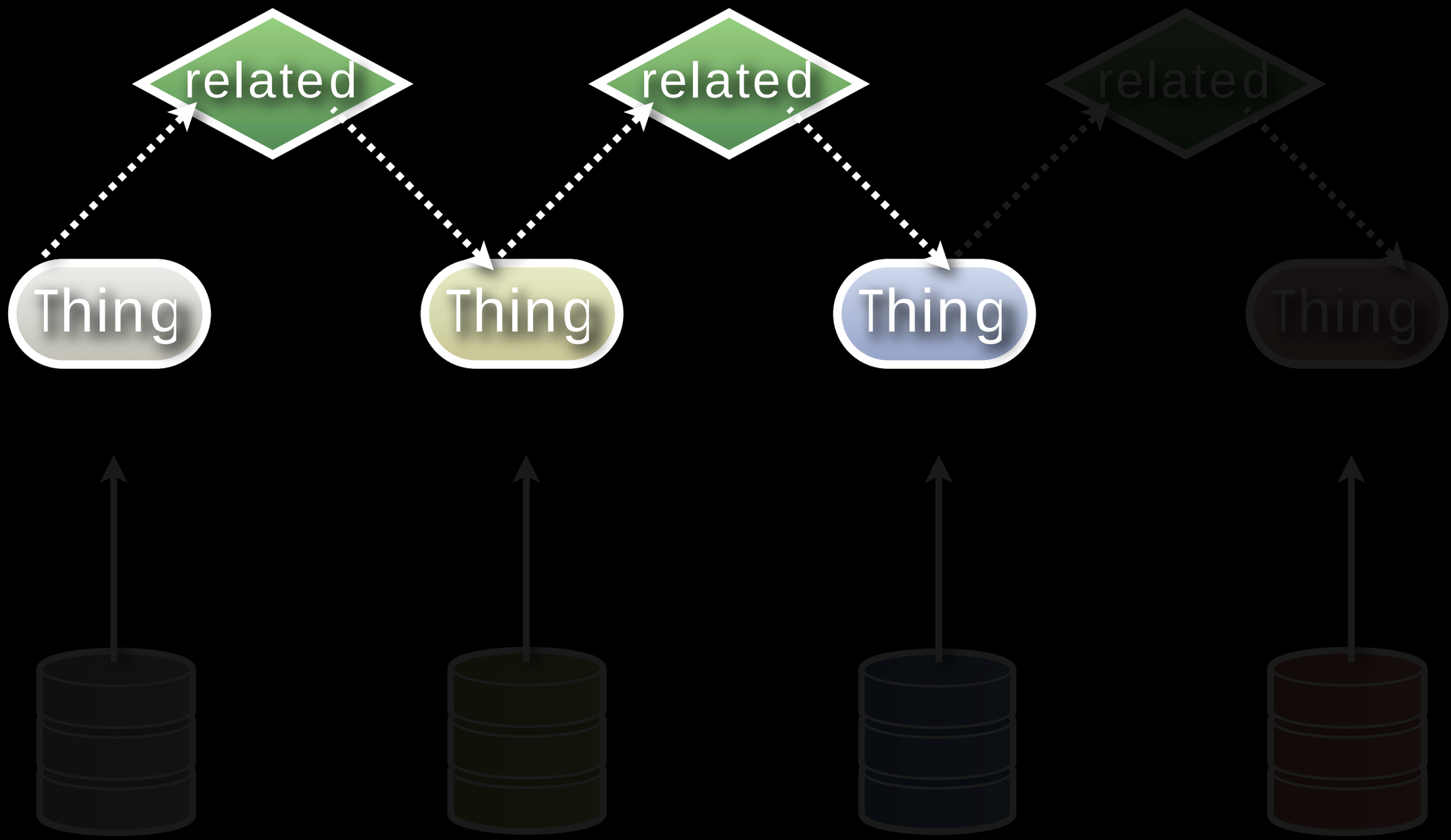
The Web of Data



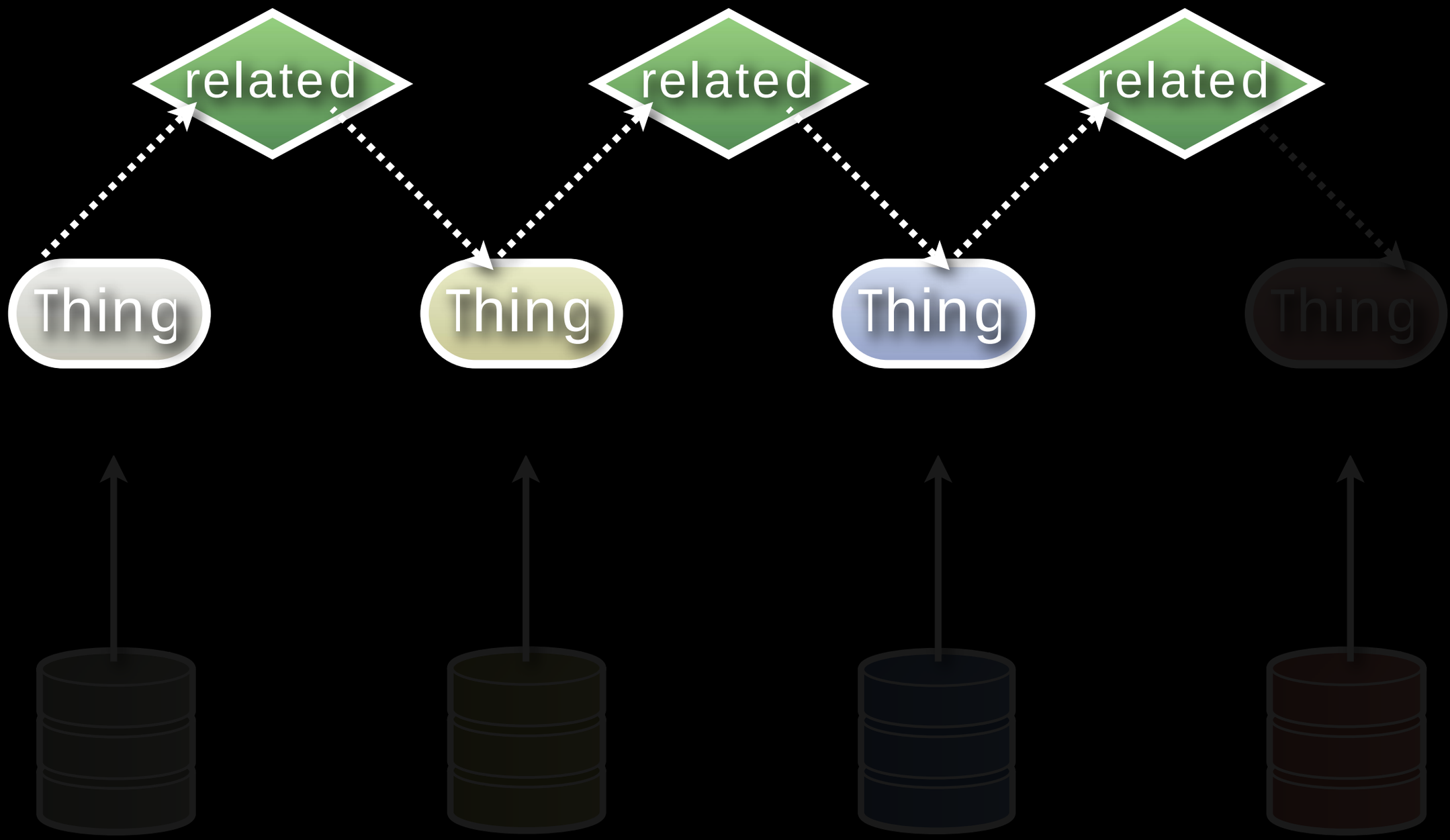
The Web of Data



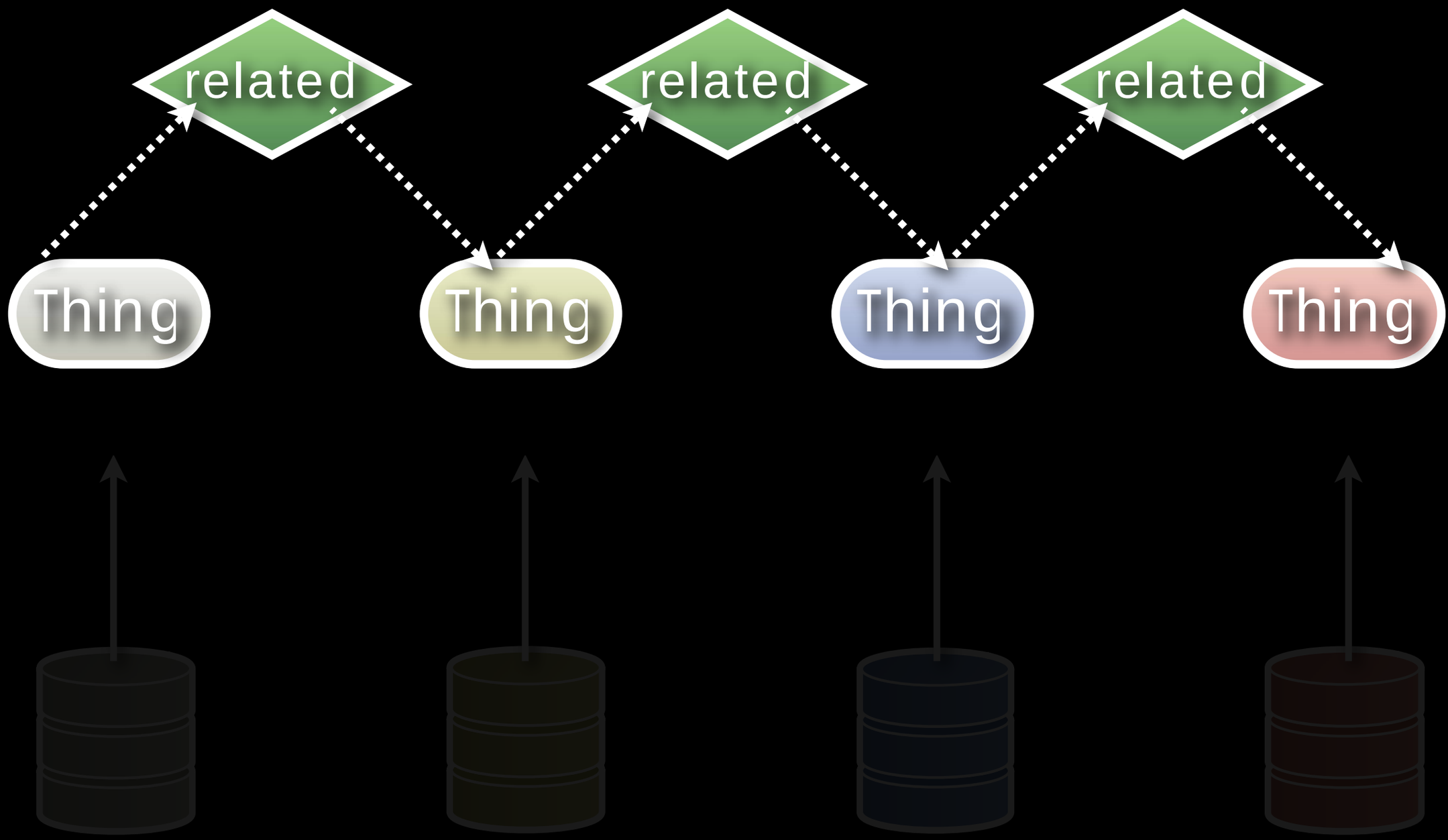
The Web of Data



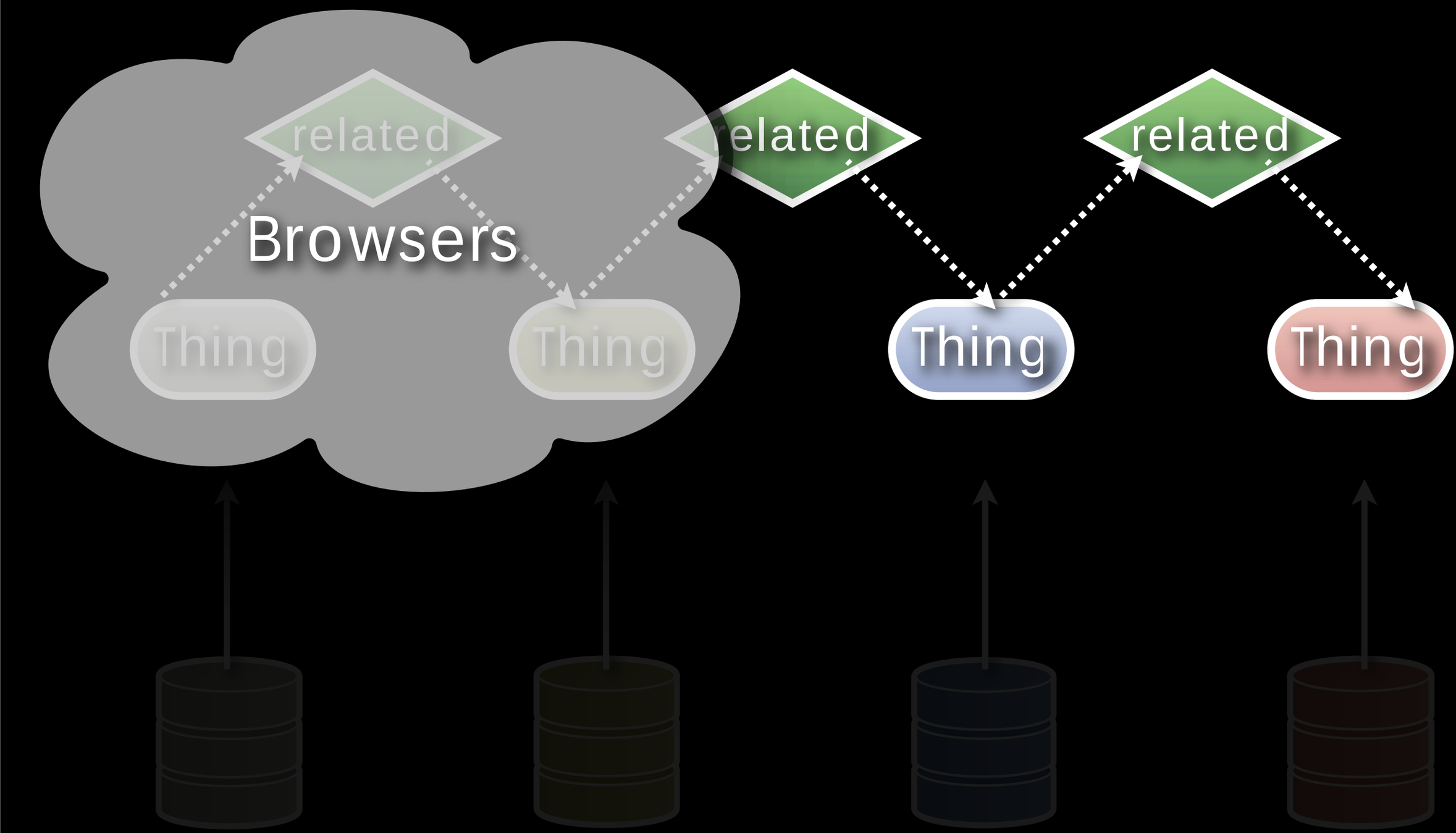
The Web of Data



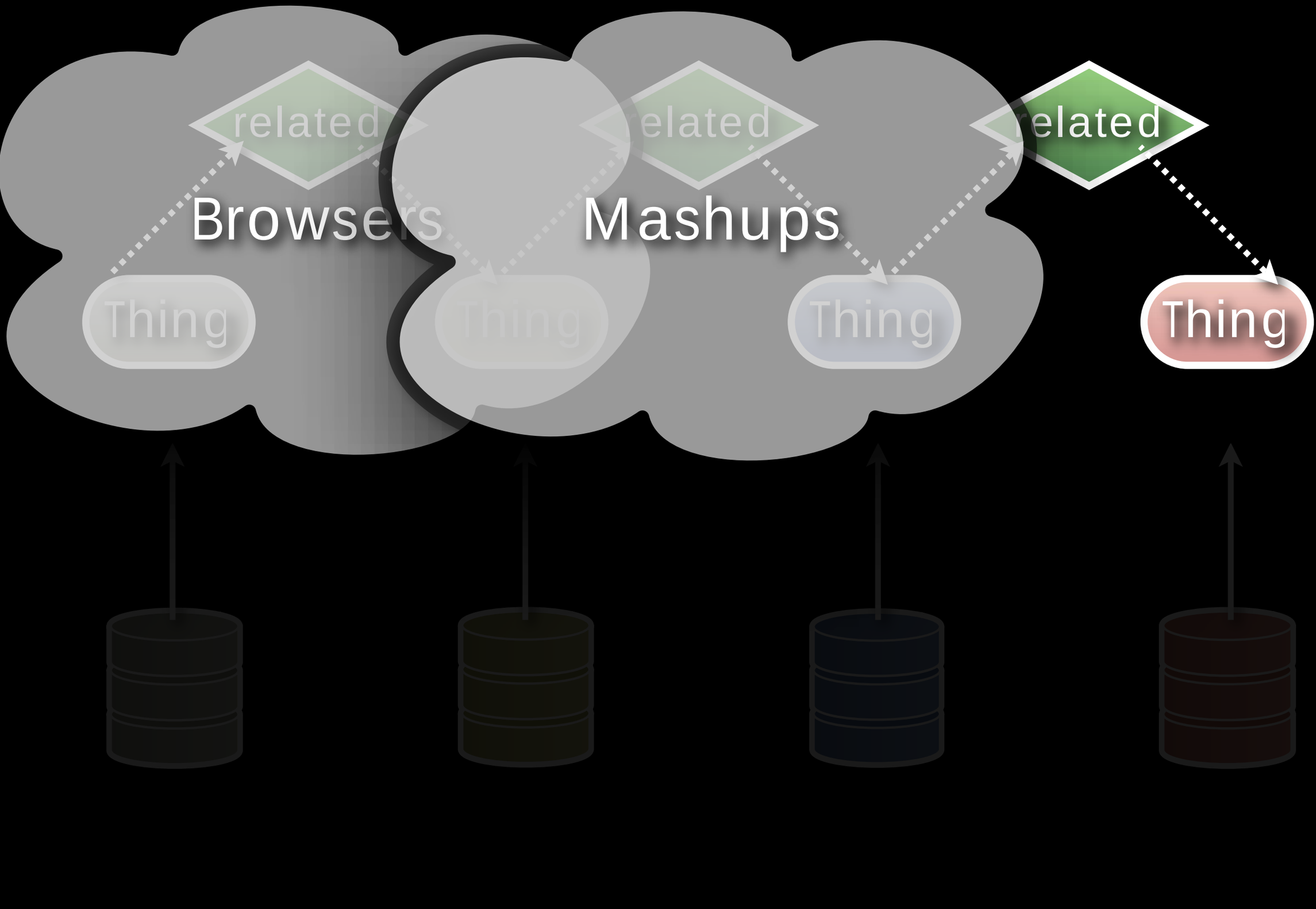
The Web of Data



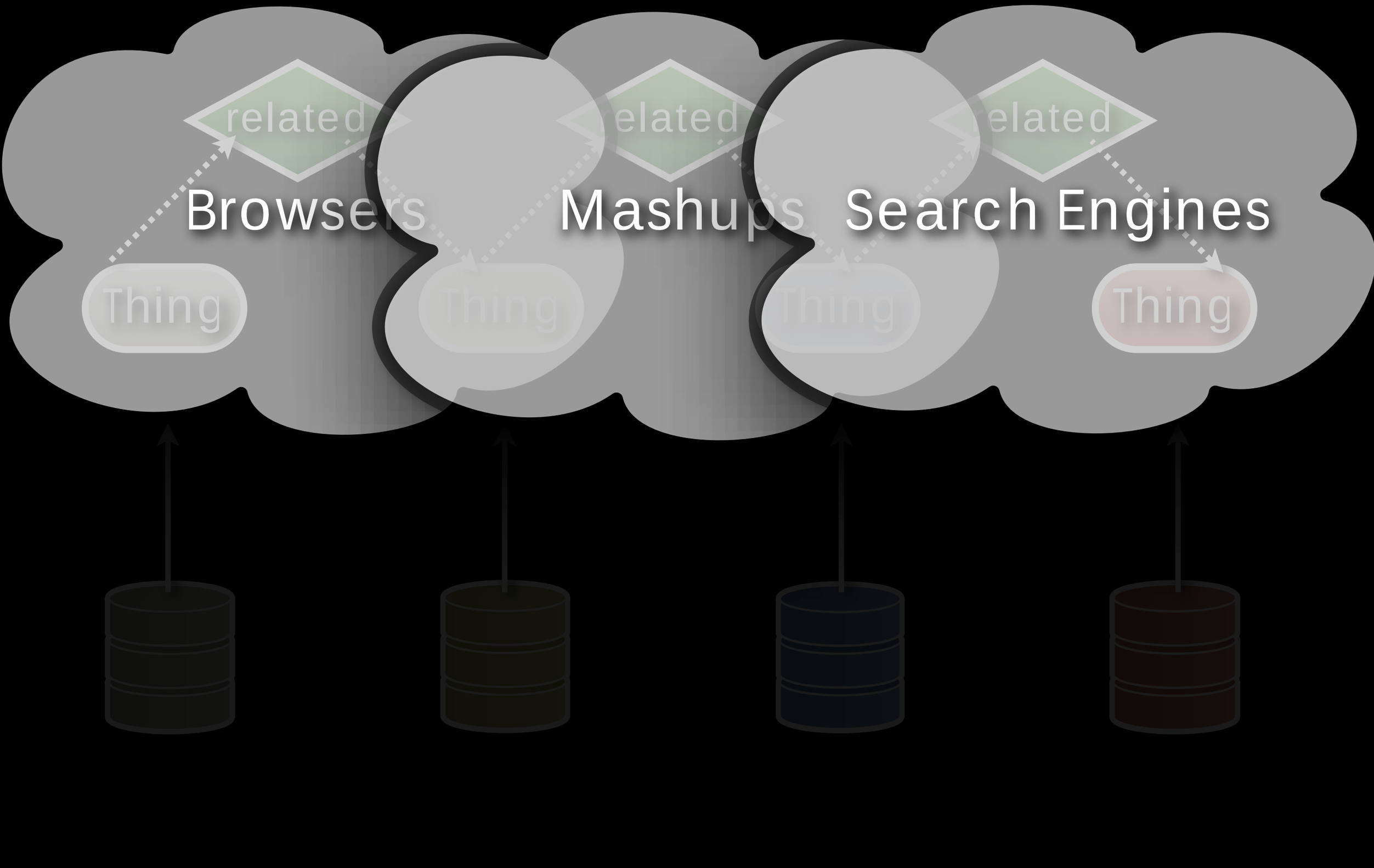
The Web of Data



The Web of Data



The Web of Data



The Web of Data

Linked data opens silos,
enabling data integration,
network effects,
and interoperability.

Linking Open Data Project

- Started in February 2007 by Chris Bizer and Richard Cyganiak
- Project of the W3C SWEO
  - Publish datasets as Linked Data
  - Interlink all the data
  - Develop apps that consume it
- A grassroots effort to bootstrap the emerging web of data

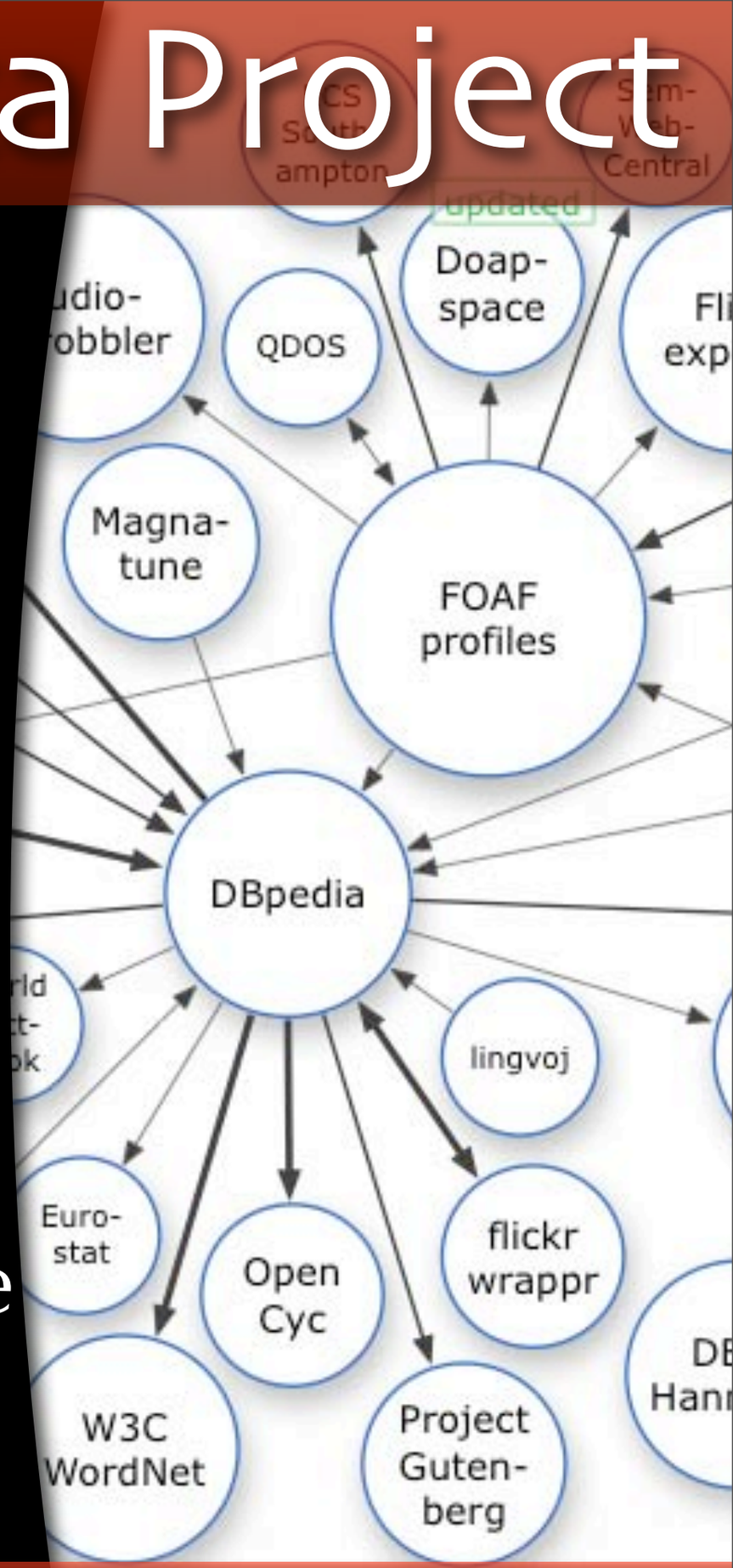


Image: Richard Cyganiak, http://richard.cyganiak.de/2007/10/lod/

Linking Open Data Project

Participating organizations:

- MIT
- University of Southampton
- DERI
- U Penn
- BBC
- OpenLink
- Talis

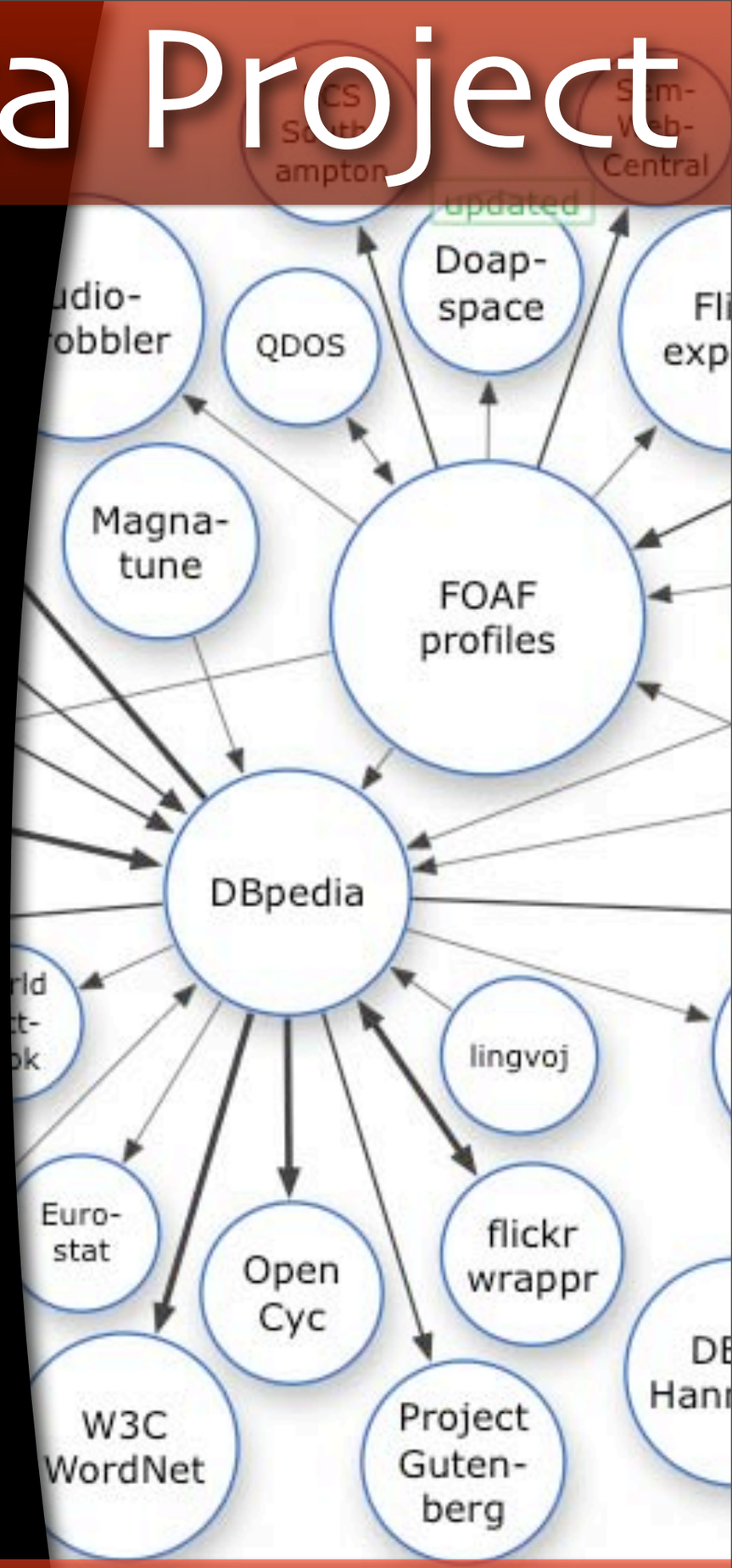
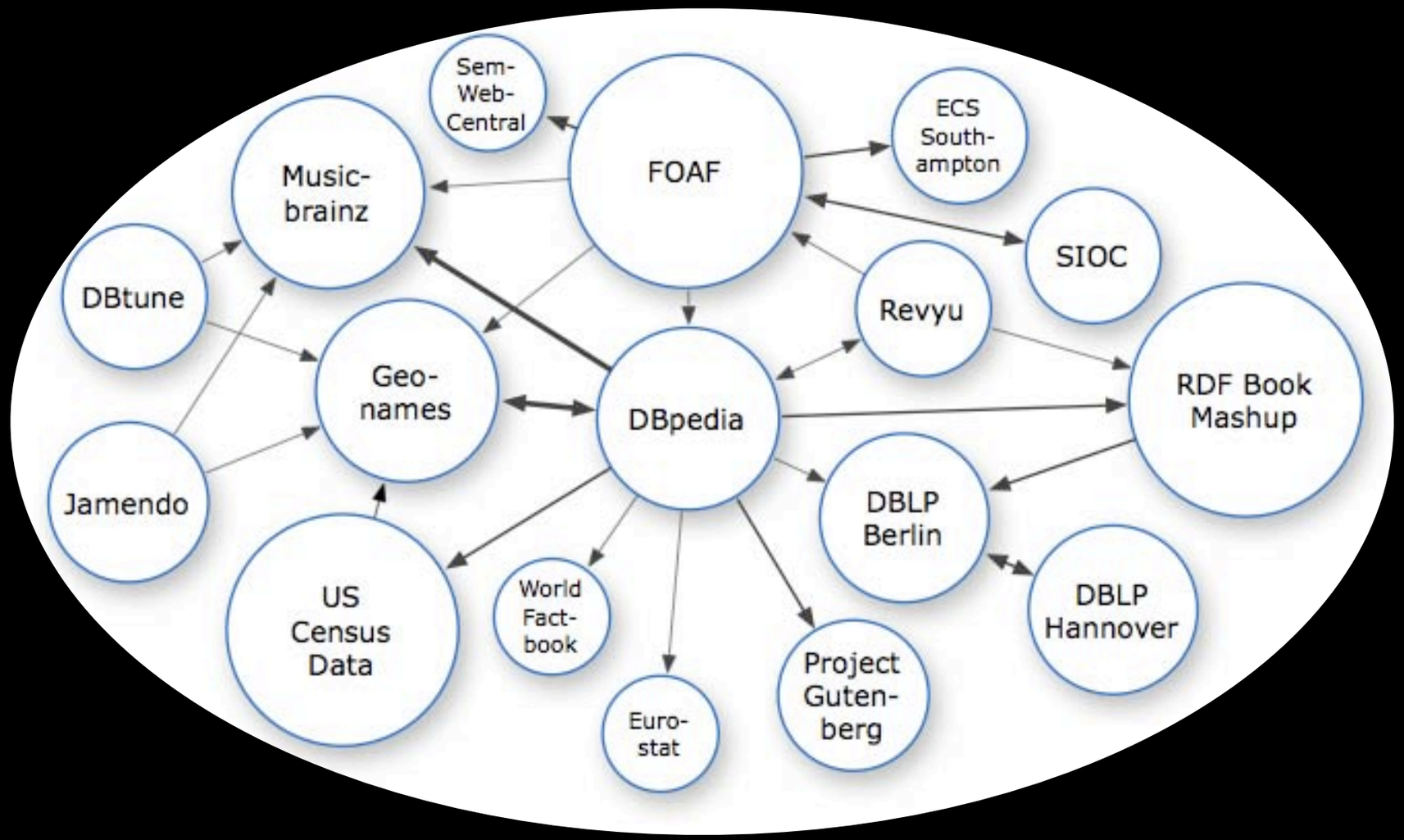


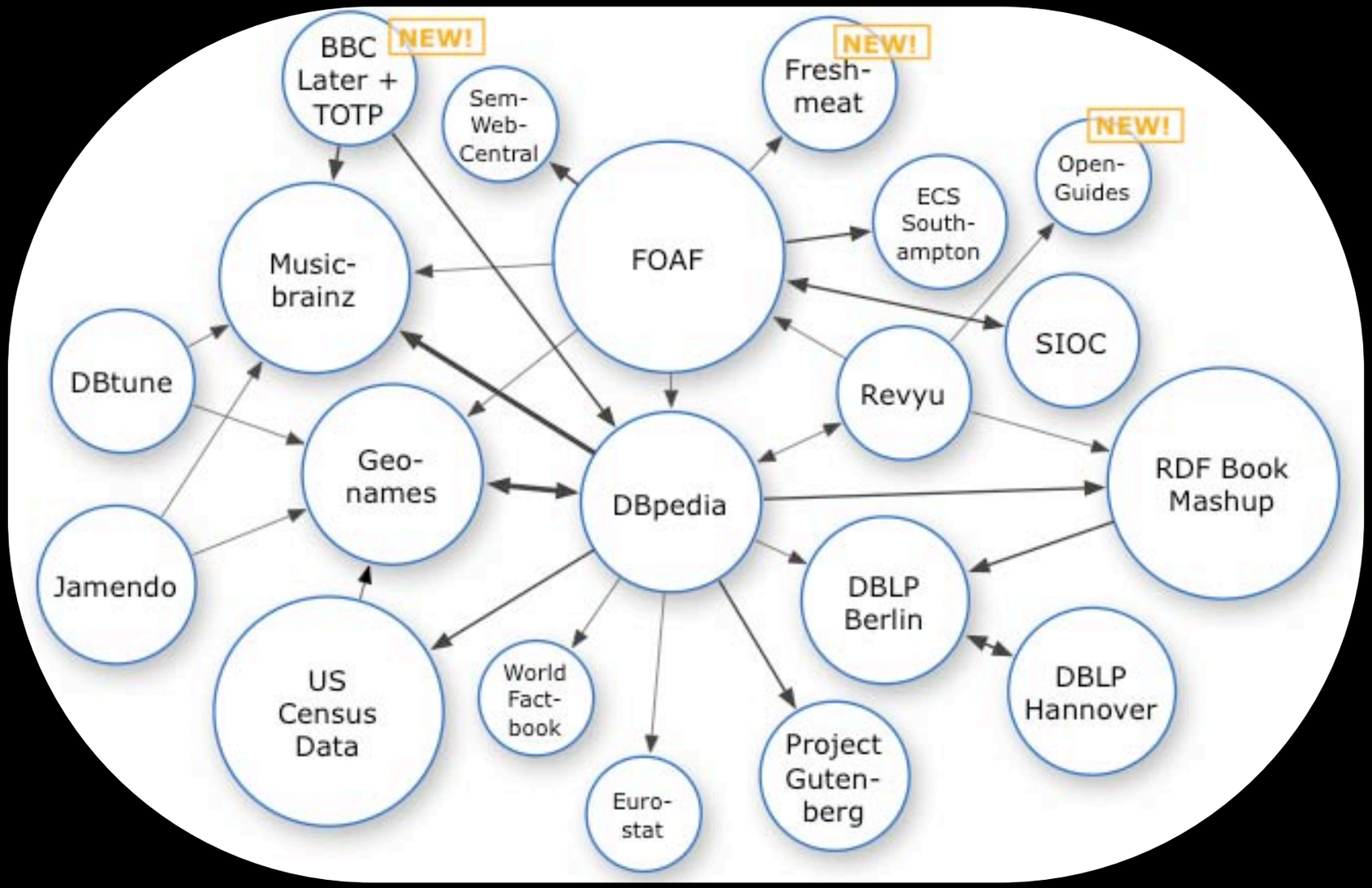
Image: Richard Cyganiak, http://richard.cyganiak.de/2007/10/lod/

LOD Growth



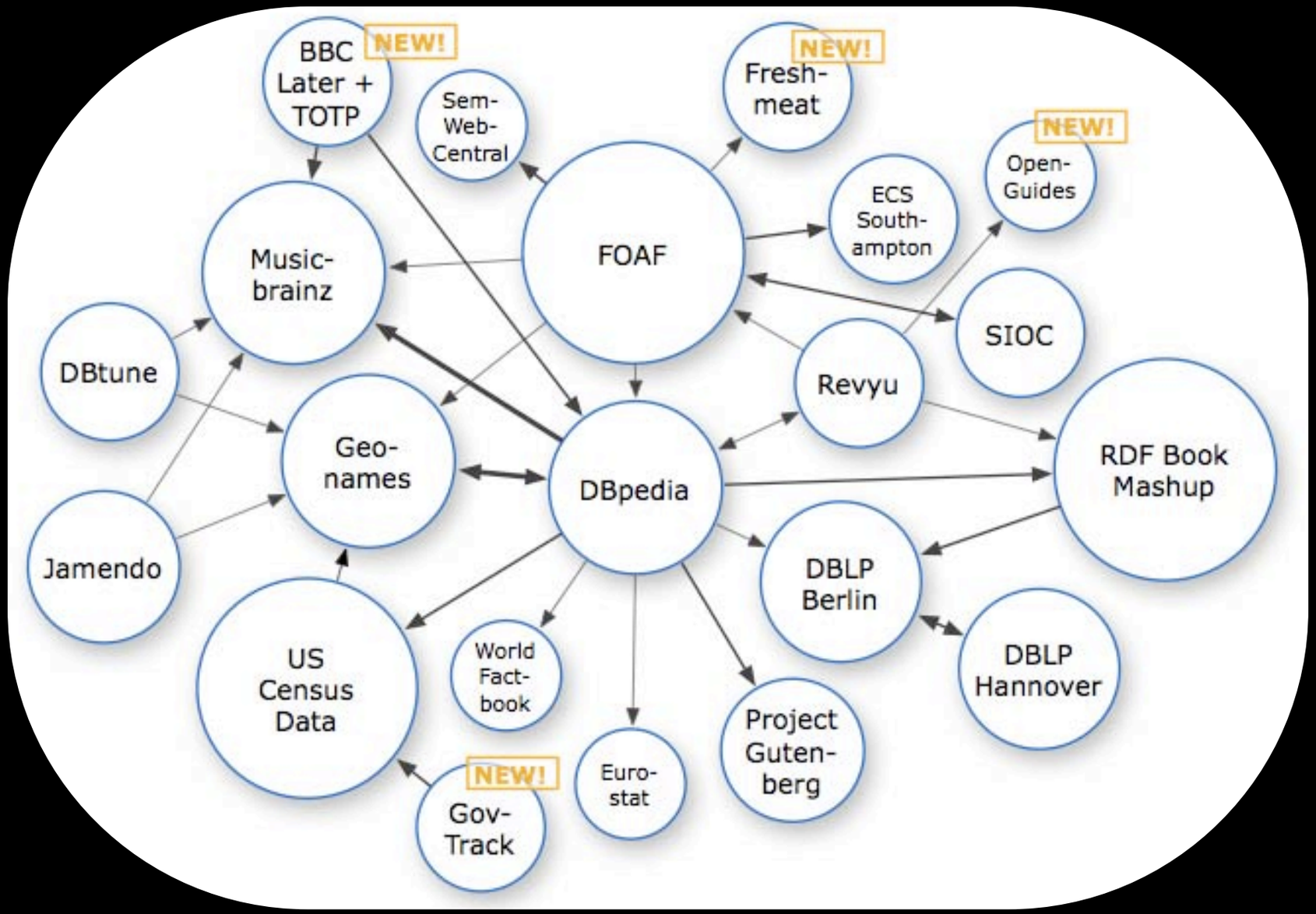
May 2007

LOD Growth



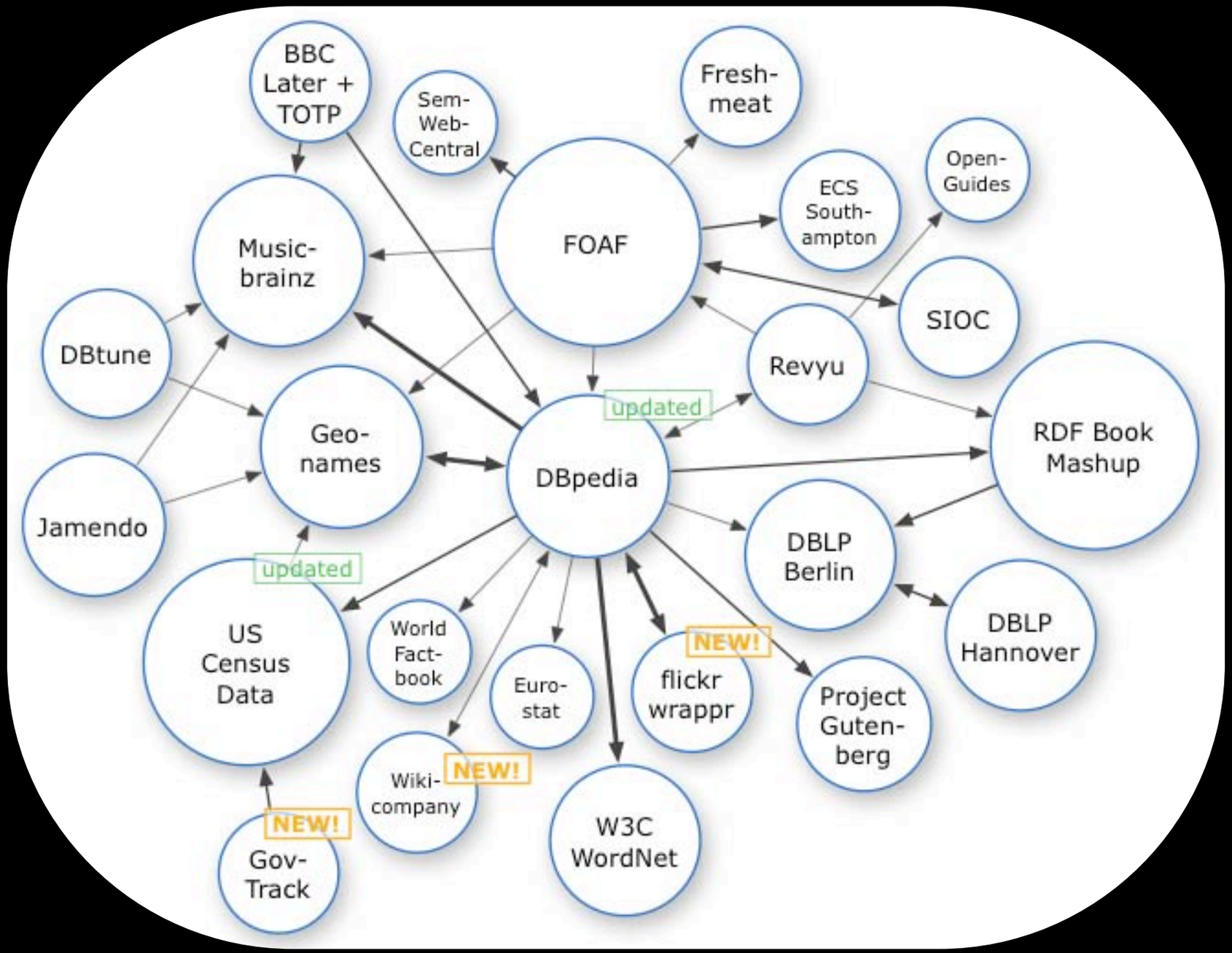
July 2007

LOD Growth



August 2007

LOD Growth



September 2007

LOD Growth

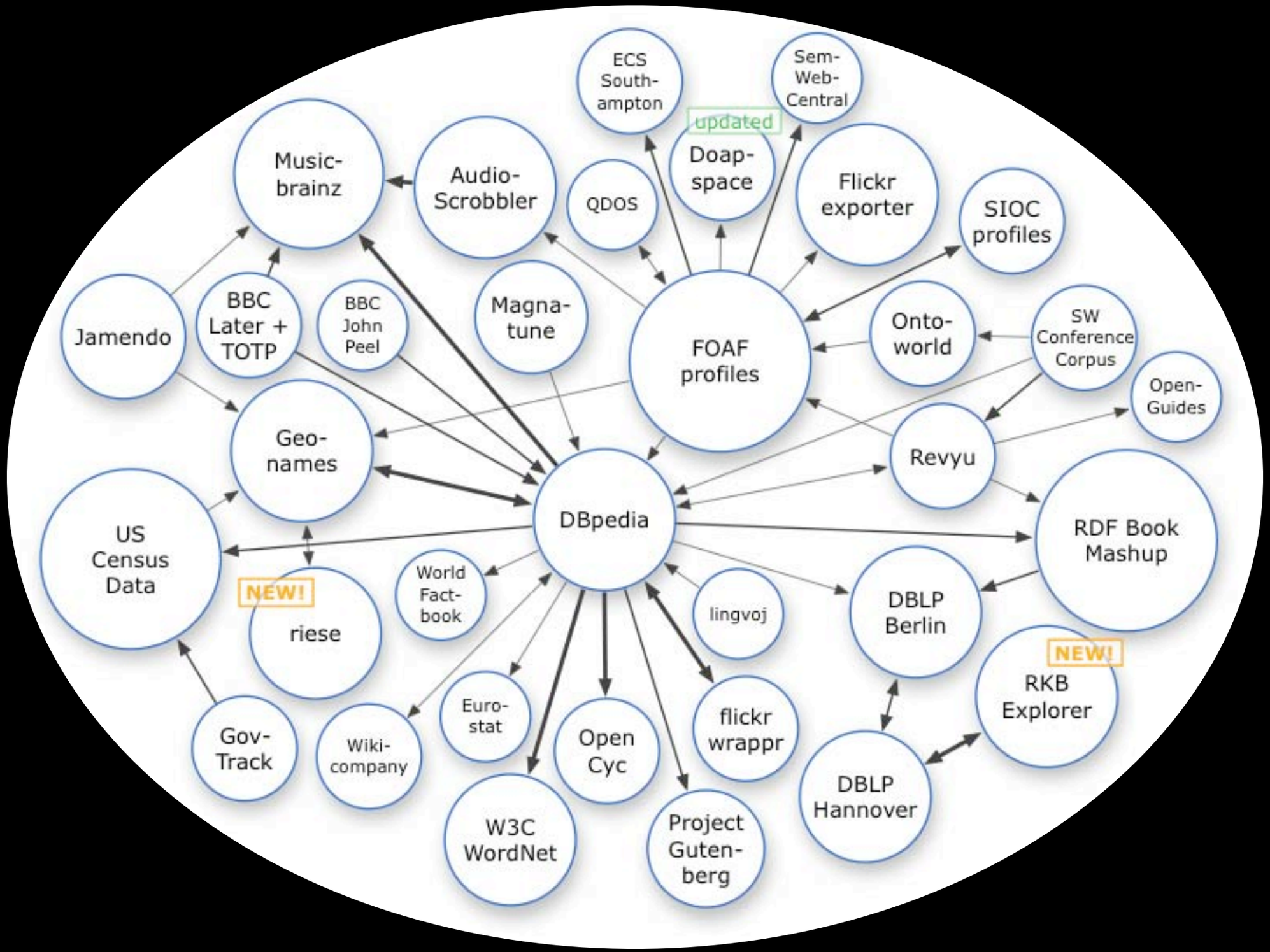


Image: Richard Cyganiak, http://richard.cyganiak.de/2007/10/lod/

LOD Growth

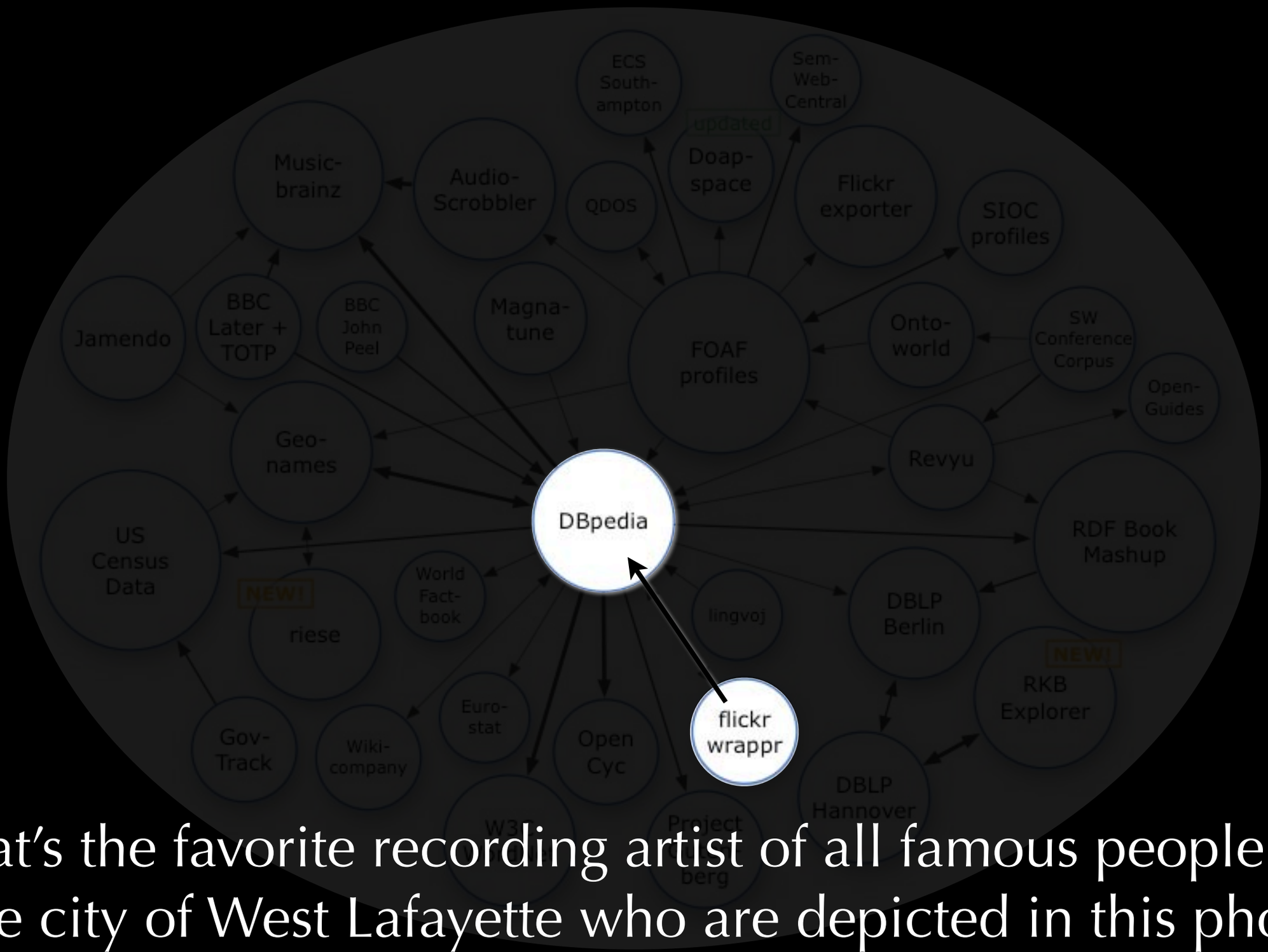
"What's the favorite recording artist of all famous people born in the city of West Lafayette who are depicted in this photo?"

LOD Growth

"What's the favorite recording artist of all famous people born in the city of West Lafayette who are depicted in this photo?"

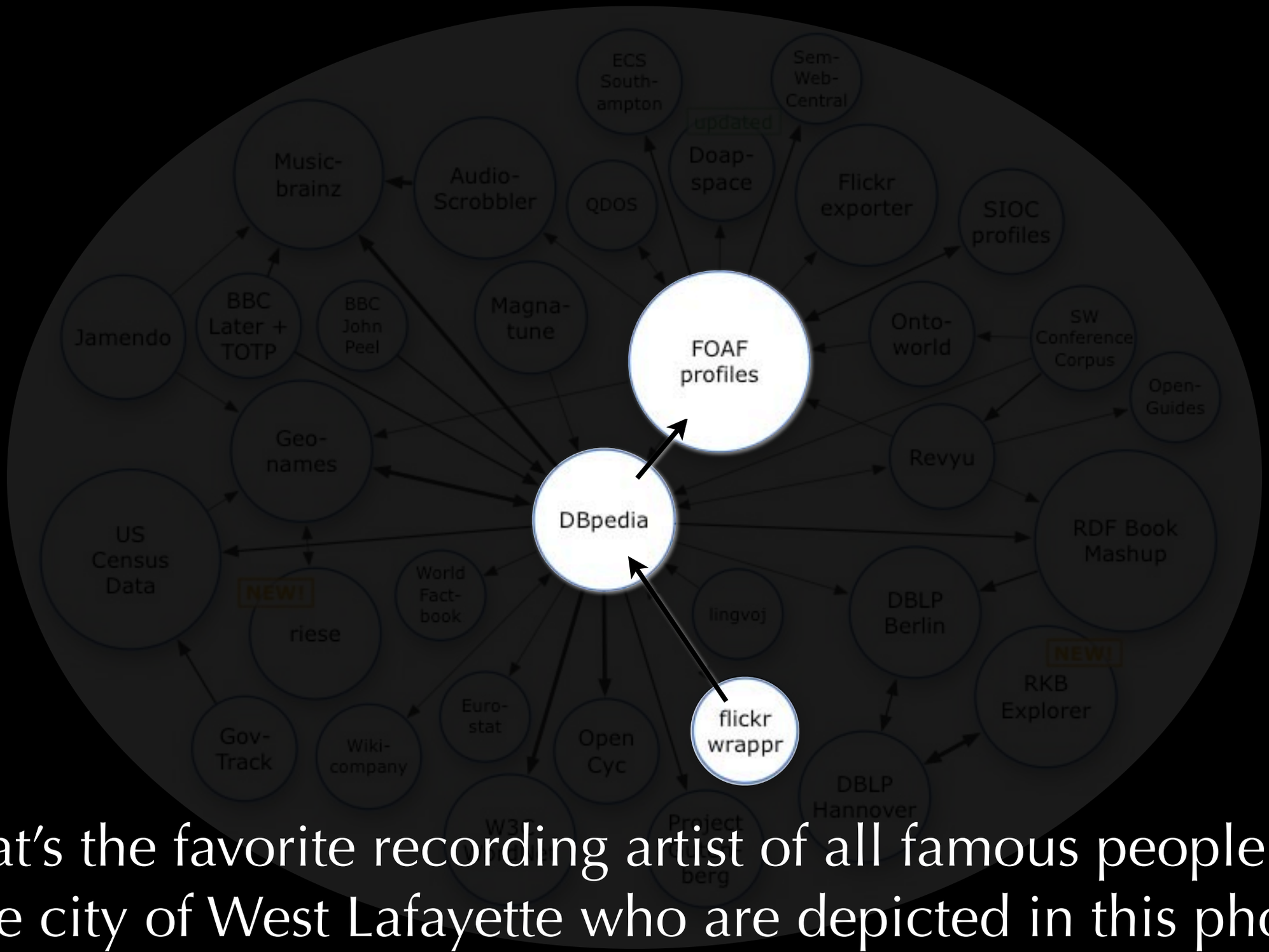
Image: Richard Cyganiak, http://richard.cyganiak.de/2007/10/lod/

LOD Growth



"What's the favorite recording artist of all famous people born in the city of West Lafayette who are depicted in this photo?"

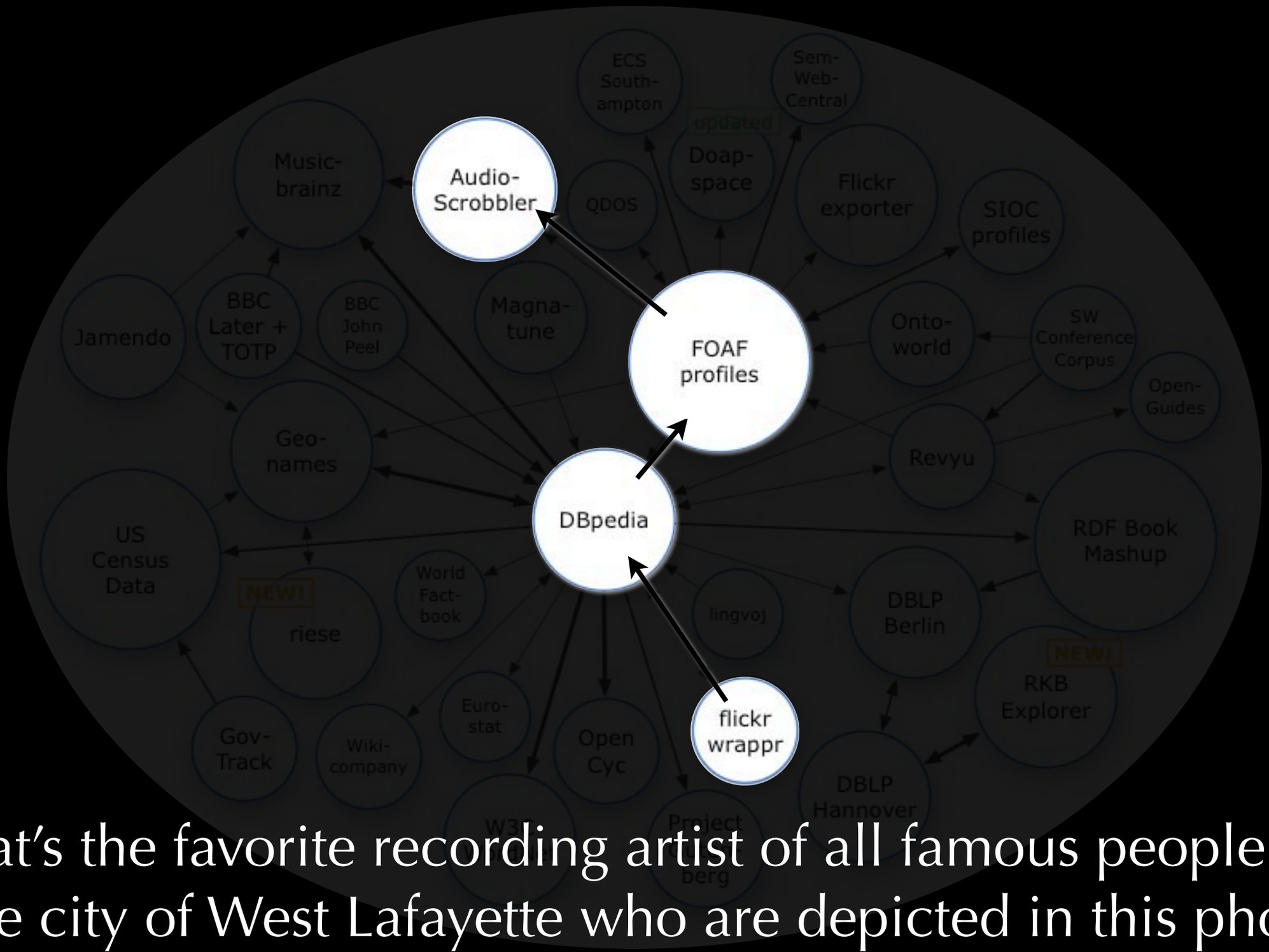
LOD Growth



"What's the favorite recording artist of all famous people born in the city of West Lafayette who are depicted in this photo?"

Image: Richard Cyganiak, http://richard.cyganiak.de/2007/10/lod/

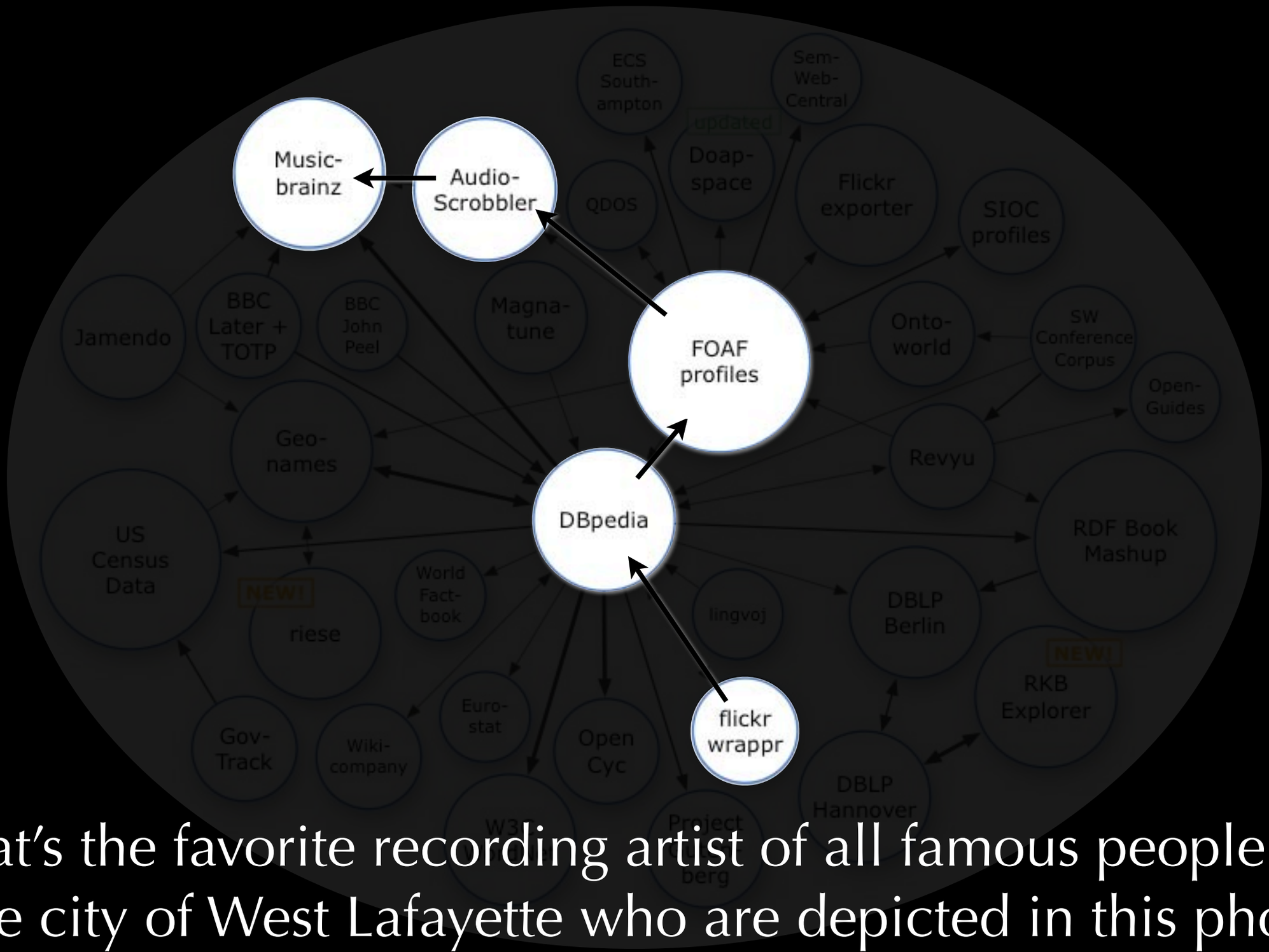
LOD Growth



"What's the favorite recording artist of all famous people born in the city of West Lafayette who are depicted in this photo?"

Image: Richard Cyganiak, http://richard.cyganiak.de/2007/10/lod/

LOD Growth



"What's the favorite recording artist of all famous people born in the city of West Lafayette who are depicted in this photo?"

Image: Richard Cyganiak, http://richard.cyganiak.de/2007/10/lod/

Linked Data & Science

- Ability to query across disparate datasets as if they were integrated & locally available
- Ability to construct arbitrarily complex search queries
- Ability to “follow your nose” through links to data you may not have known existed

Image: frankz, http://flickr.com/photos/frankz/

Queries of the Future

- Find data representing the state of the neutral atmosphere anywhere above 100km and toward the Arctic circle (above 45° North) at times of high geomagnetic activity.
- ... peer-reviewed papers that incorporate those data?
- ... other datasets used by authors of those papers?
- ... by authors of all cited papers?

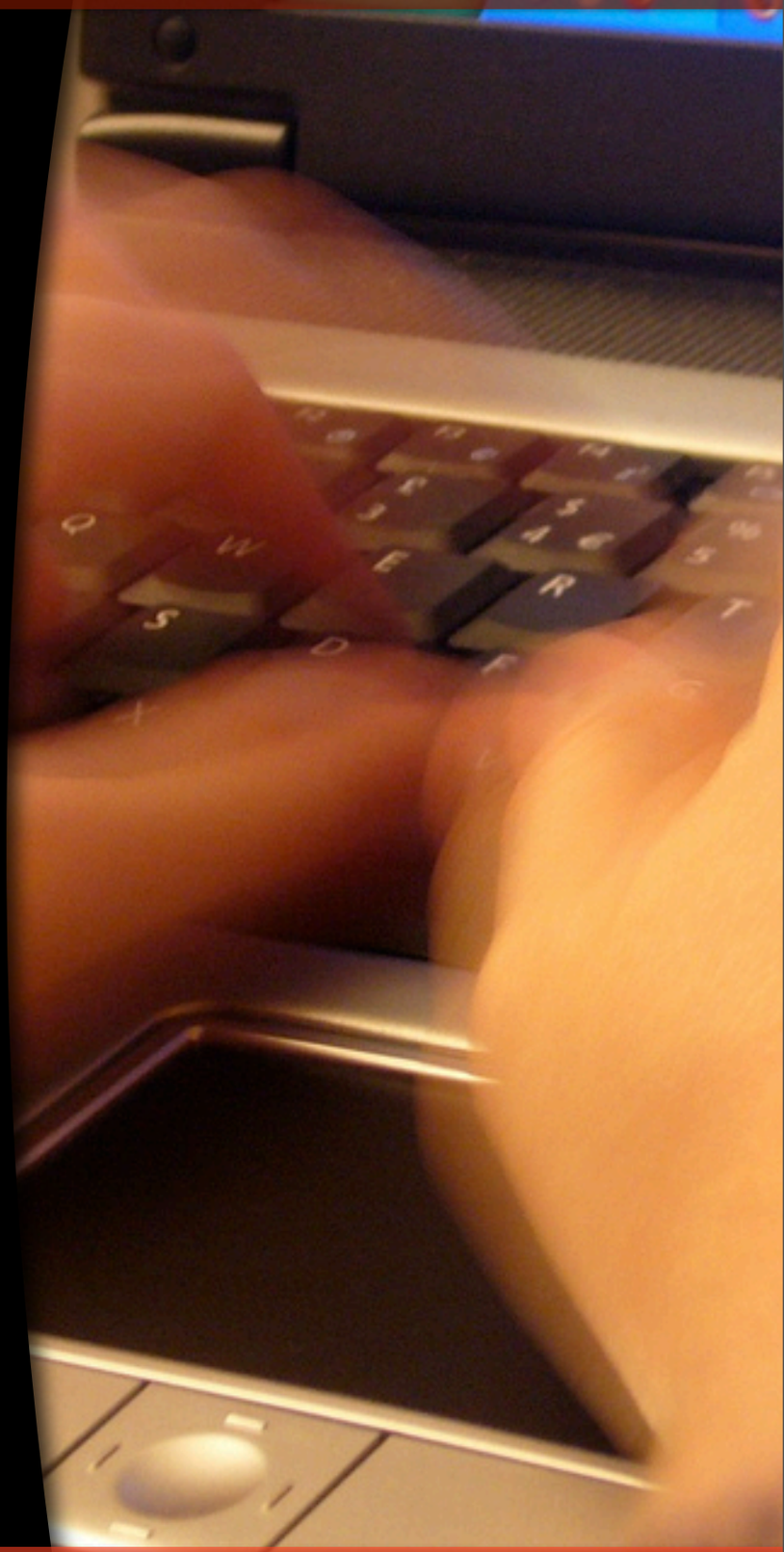
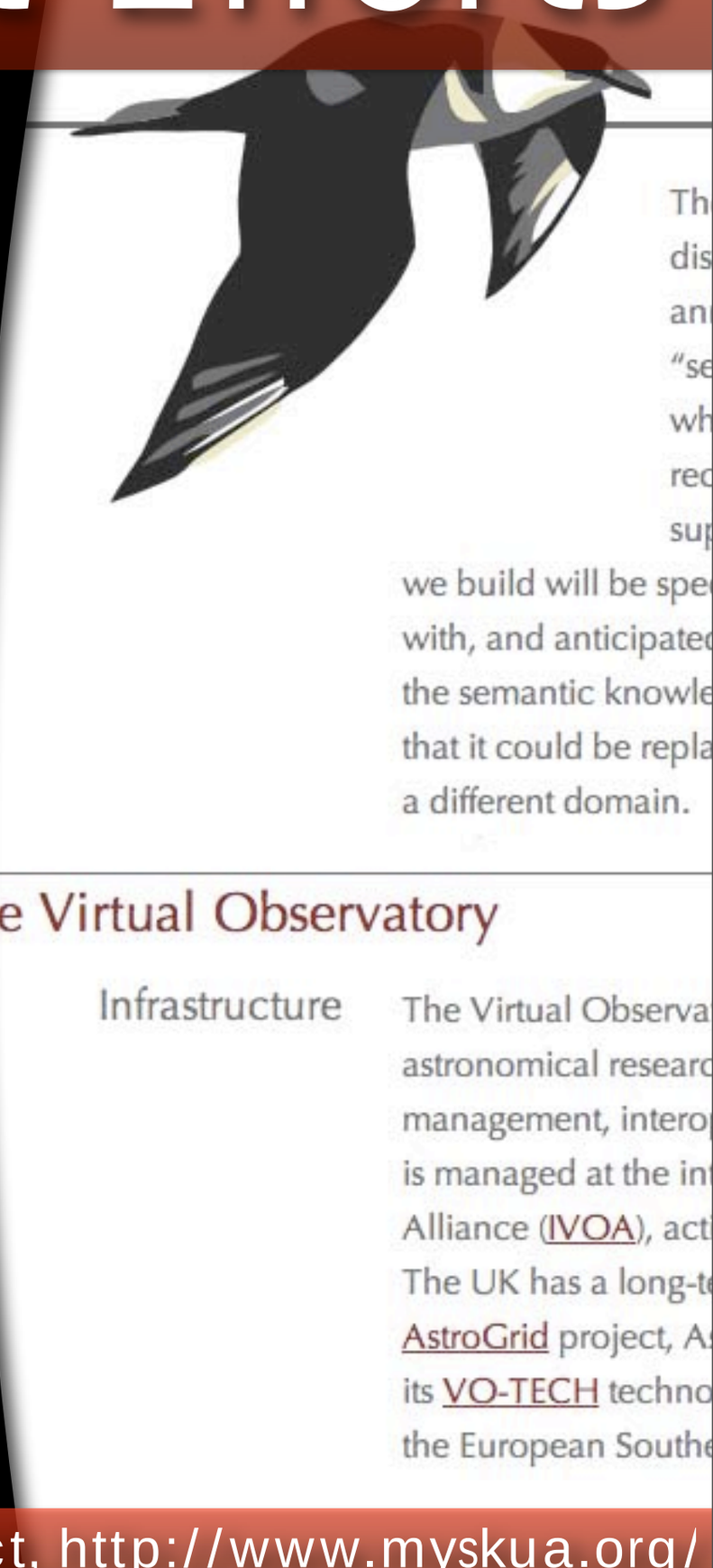


Image: Jim Grant, http://flickr.com/photos/jimgrant/

Semantic e-Science Efforts

- IVOA / SKUA
- VSTO
- SESDI
- SWEET
- GEON
- Lots of biomedical stuff



Screenshot: SKUA project, http://www.myskua.org/

A Slight Problem

- The graph of linked science data is sparse
- Violates Linked Data principle #4
- Using RDF improves interoperability, but datasets are still not well connected
- This makes discovery a challenge

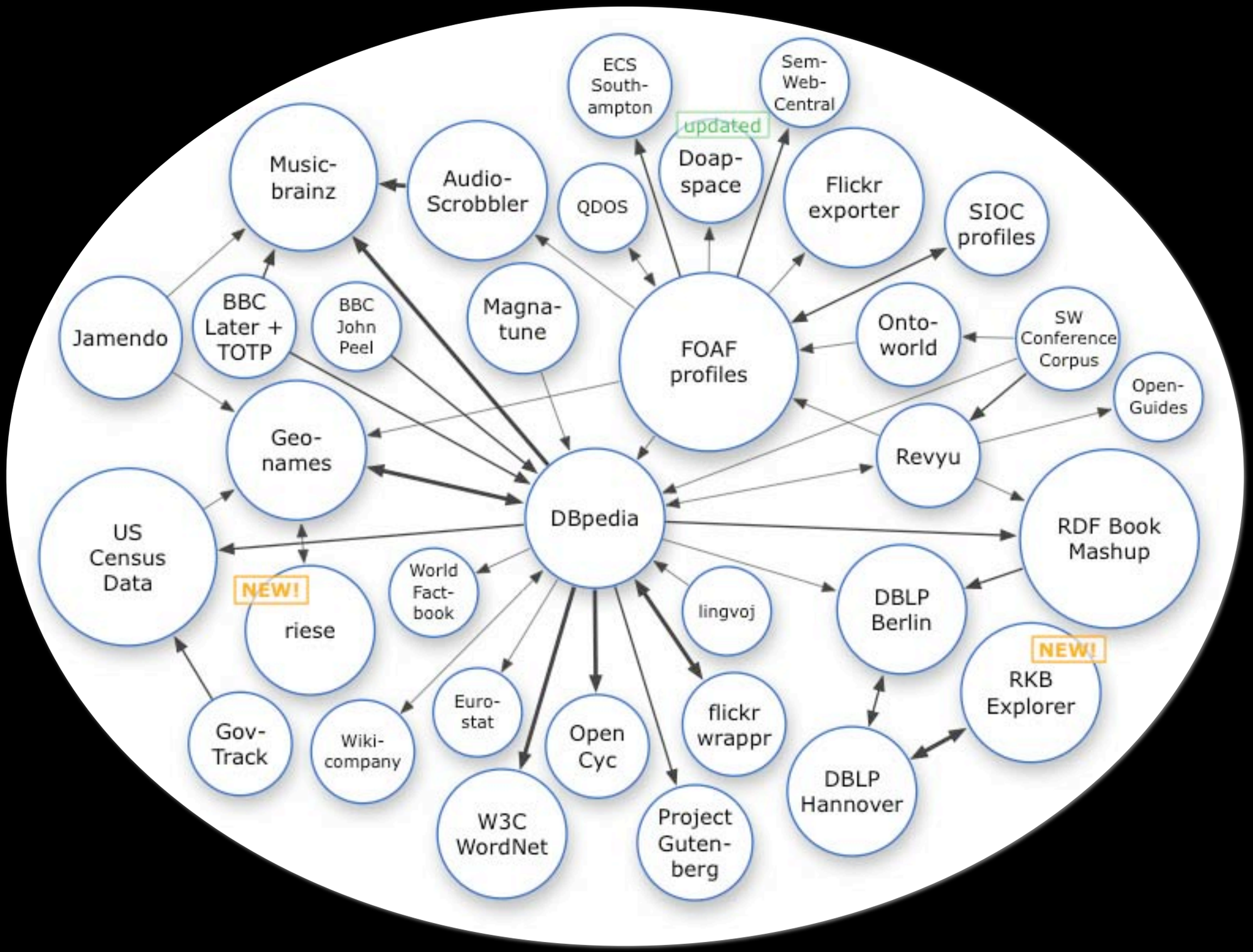
Image: Mike Giovinazzo, http://flickr.com/photos/mike_giovinazzo/

e-Science Needs Hubs

- Like many network structures, linked data graphs tend to develop a few “hubs” that are well-connected
- These hubs create links between data that would otherwise remain disconnected
- In LOD: DBpedia, FOAF profiles, Geonames

Image: Cobalt, http://flickr.com/photos/cobalt/

e-Science Needs Hubs



e-Science Needs Hubs

e-Science Needs Hubs

The Point

Library data
can provide the hubs
for a network
of linked scientific data.

Image: dulcelife, http://flickr.com/photos/dulcelife/

The Point

- Catalog records, authority files, controlled vocab, publisher data
- Links exist, but not yet machine readable
- Reference librarians know this from experience
- Not a new concept (Semantic Association Networks)

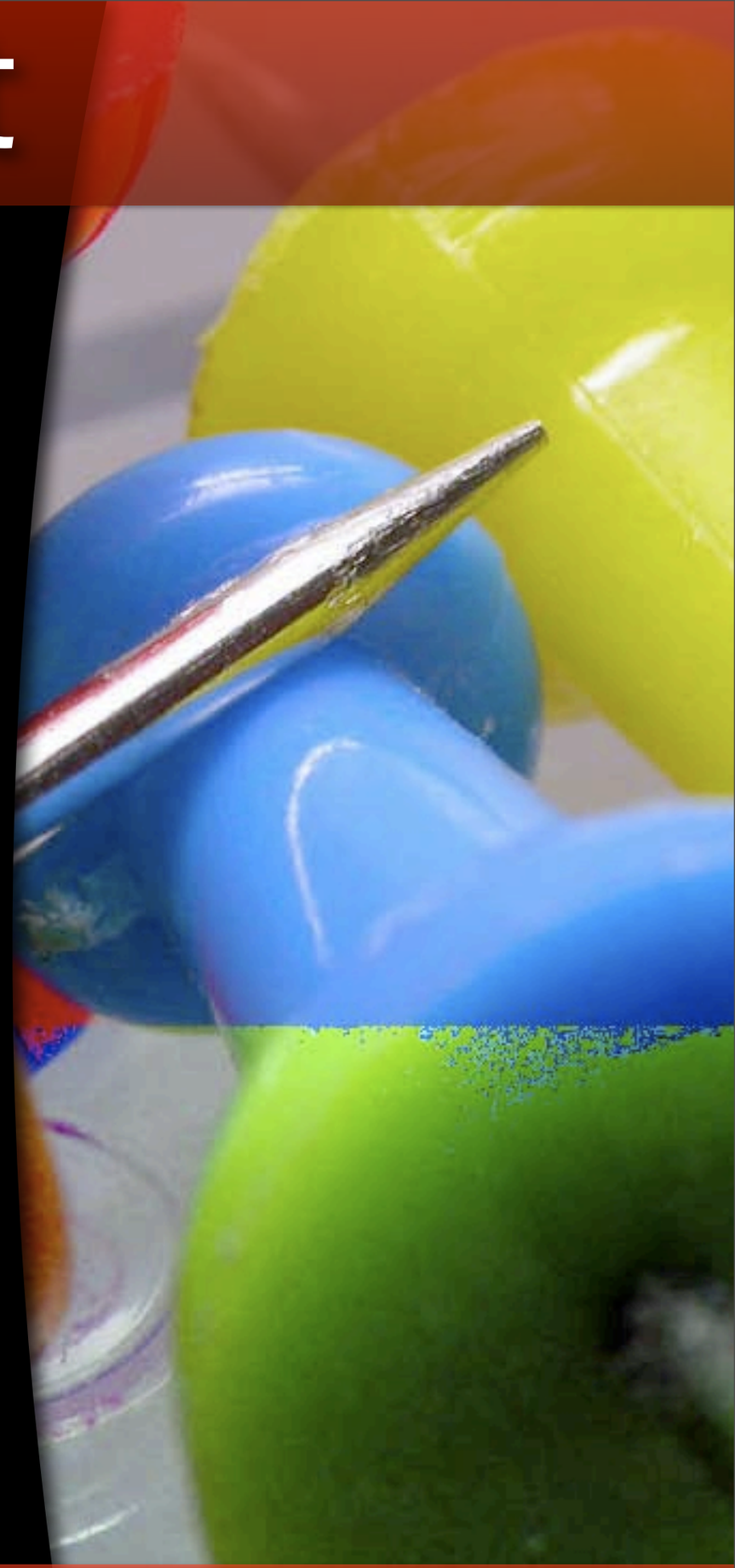


Image: dulcelife, http://flickr.com/photos/dulcelife/

The Point

- Catalog records, authority files, controlled vocab, publisher data
- Links exist, but not yet machine readable
- Reference librarians know this from experience
- Not a new concept (Semantic Association Networks)

Image: triplefivedrew, http://flickr.com/photos/triplefivedrew/

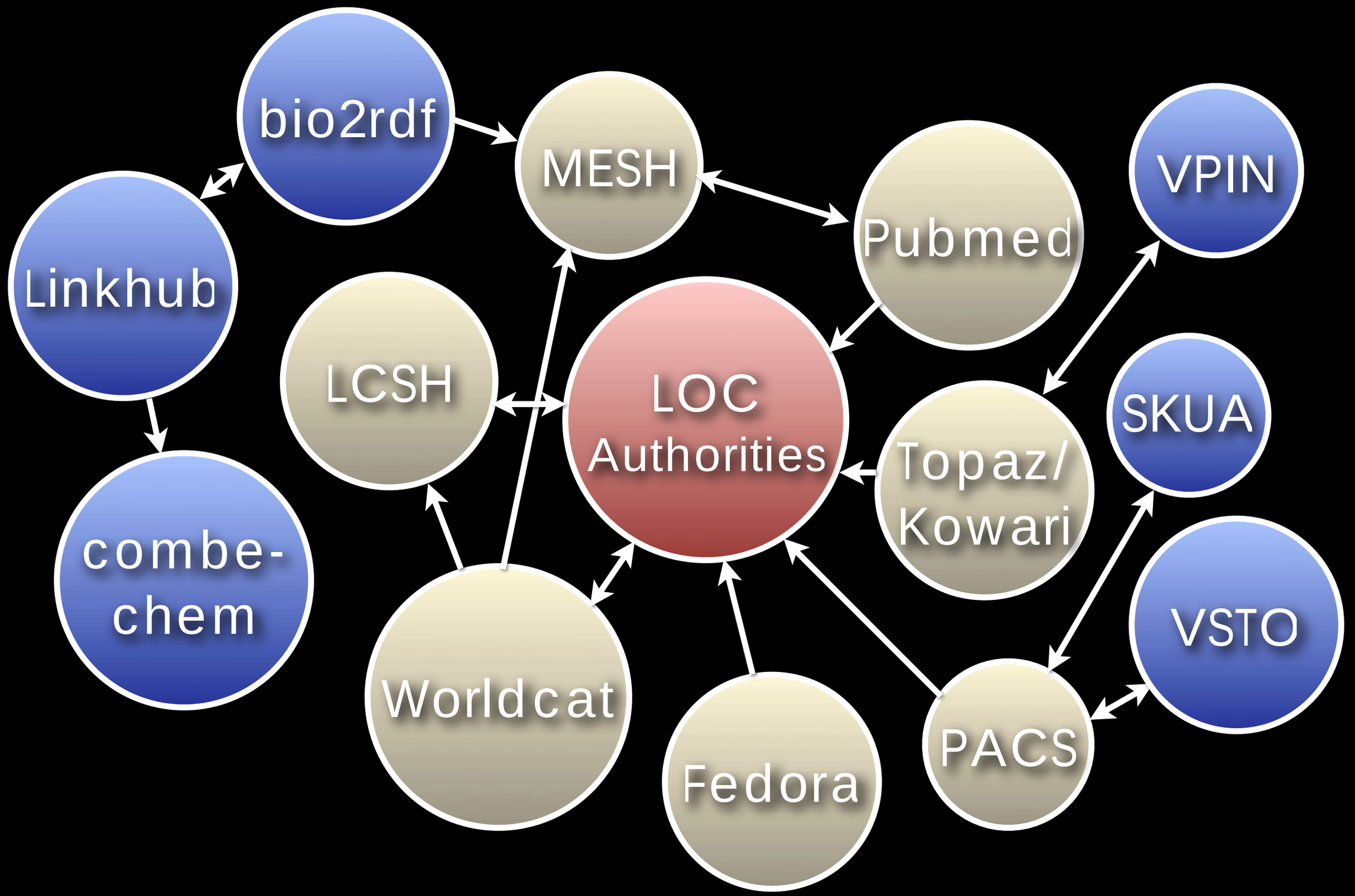
The Point

- Catalog records, authority files, controlled vocab, publisher data
- Links exist, but not yet machine readable
- Reference librarians know this from experience
- Not a new concept (Semantic Association Networks)

Image: metimbers2000, http://flickr.com/photos/metimbers2000/

Library as Linked Data Hub

Library as Linked Data Hub

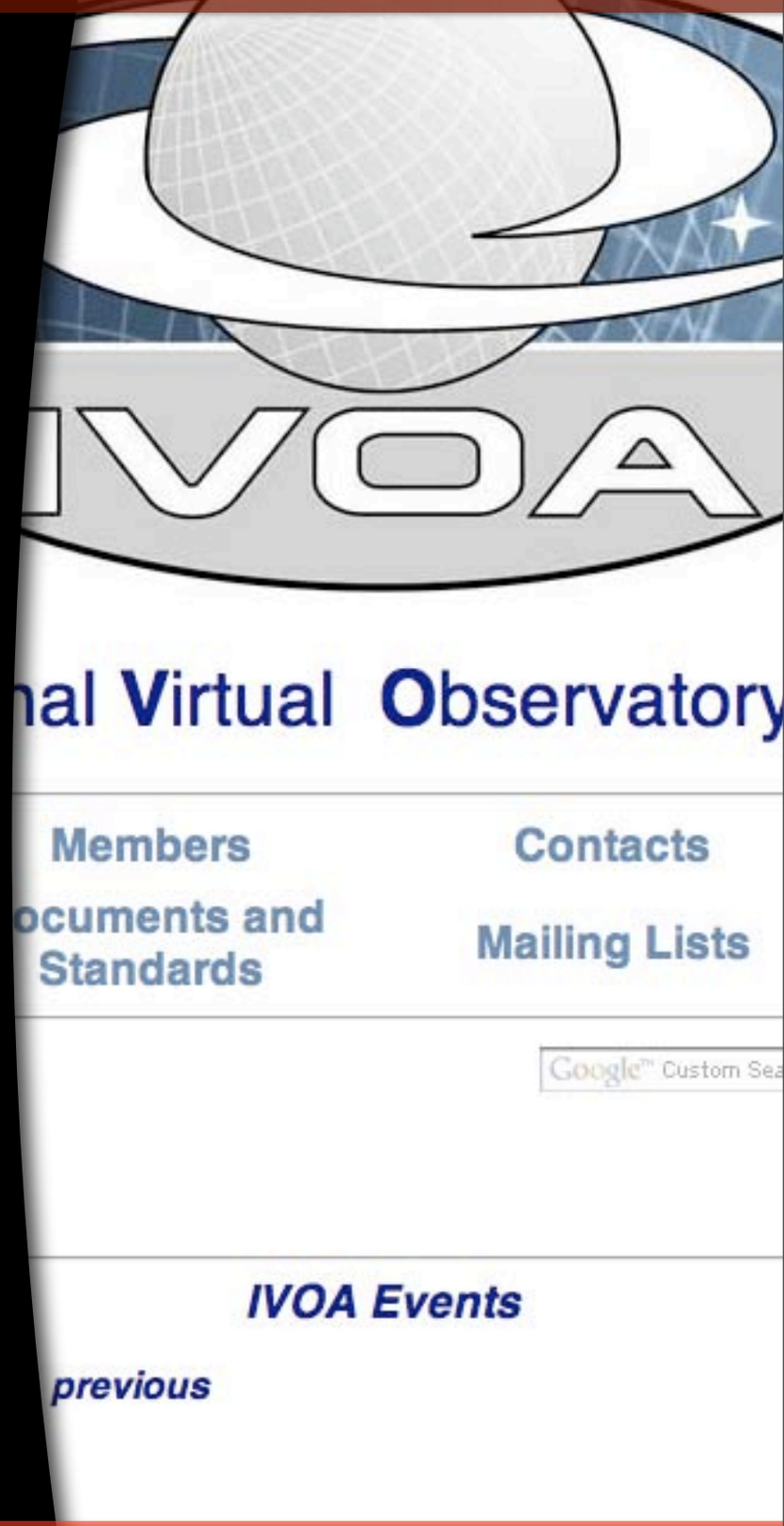


Library as Linked Data Hub

We need a Linking Open Data project
for library data.

Who's Working On It?

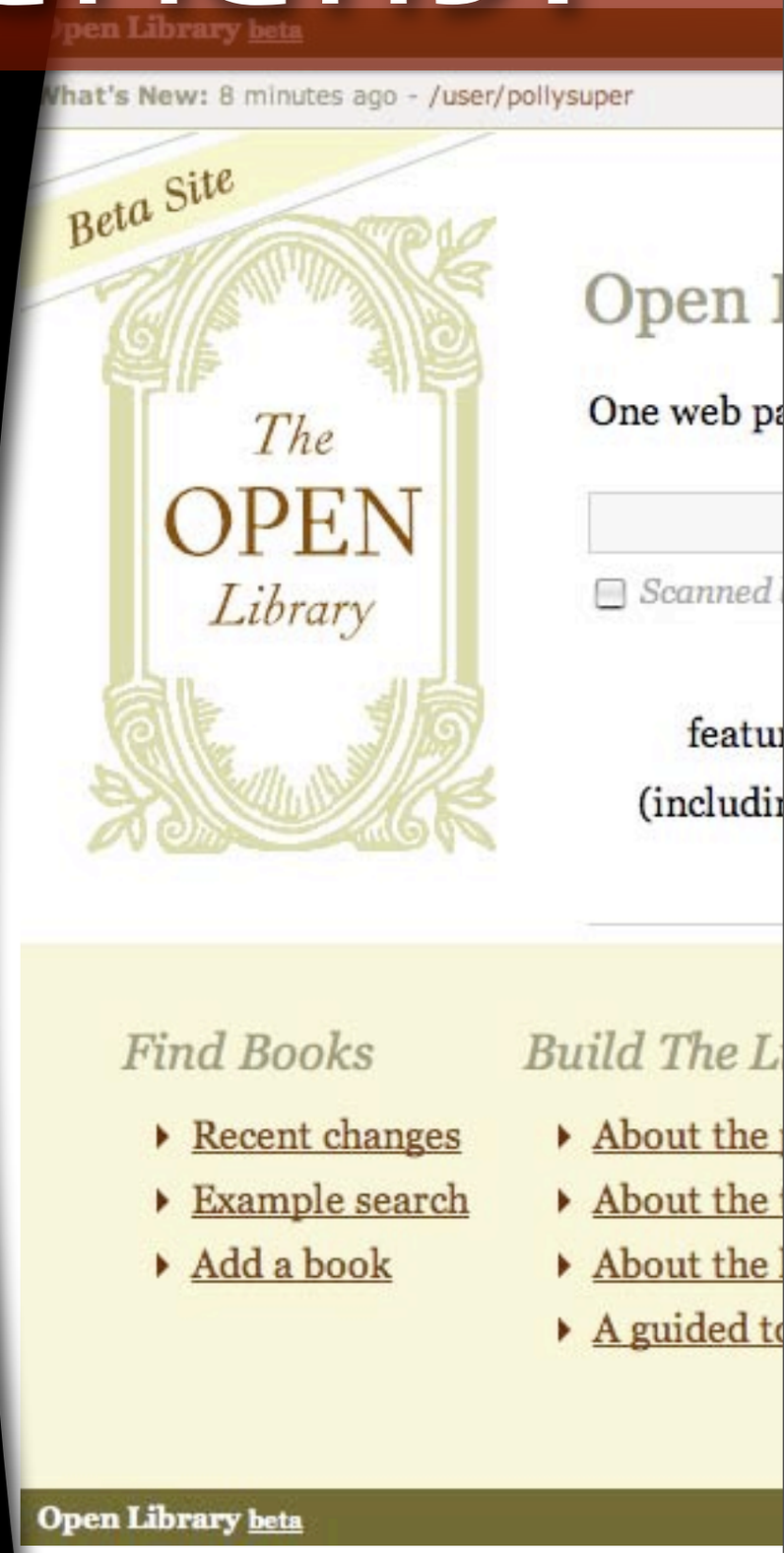
- ADS - data linking
- NLP / text mining
- Manual indexing
- Publisher-supplied metadata
- Microsoft and Google
- Publishers (Nature, Elsevier, PLoS)



Screenshot: International Virtual Observatory Alliance, http://www.ivoa.net/

What About Librarians?

- NSDL Metadata Registry
- DCMI/RDA Joint Task Group
- OAI ORE
- LCSH in SKOS @ LoC
- Bibliographic Ontology
- OpenLibrary
- Semantic repositories
- Semantic MARC @ Talis



Screenshot: Open Library, http://www.openlibrary.org/

Get Involved

- http://metadataregistry.org/
- http://dublincore.org/dcmirdataskgroup/
- http://www.openarchives.org/ore/
- http://lcsh.info/
- http://bibliographicontology.com/
- http://openlibrary.org/
- http://www.fedora-commons.org/
- http://tinyurl.com/639pp4 (Rob Styles)

Thanks!

- http://metadataregistry.org/
- http://dublincore.org/dcmirdataskgroup/
- http://www.openarchives.org/ore/
- http://lcsh.info/
- http://bibliographicontology.com/
- http://openlibrary.org/
- http://www.fedora-commons.org/
- http://tinyurl.com/639pp4 (Rob Styles)

jkglenn@umich.edu