

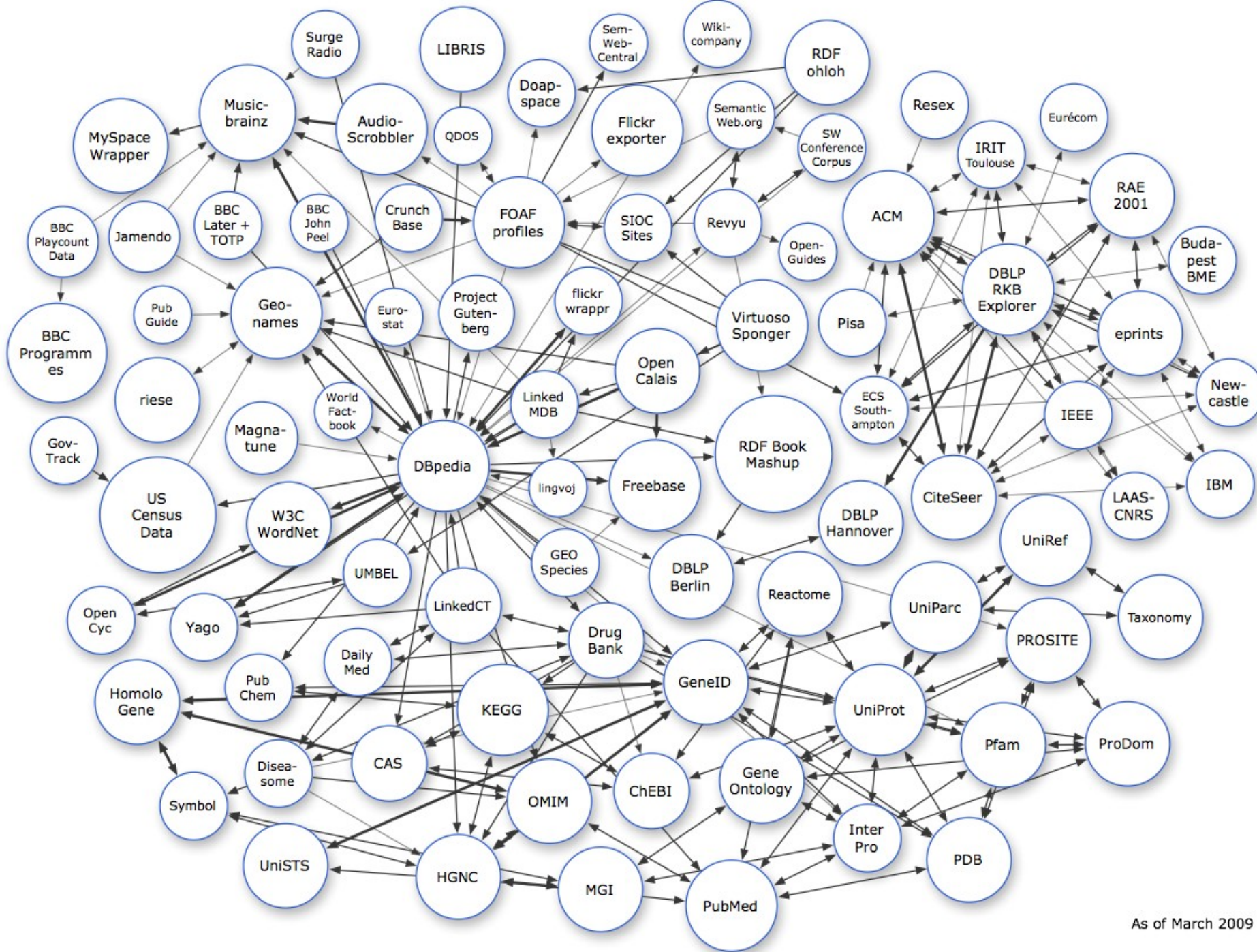


How to Publish Linked Data on the Web

Dr. Tom Heath
Platform Division
Talis Information Ltd

tom.heath@talis.com
<http://tomheath.com/id/me>

9 July 2009
SSSW2009, Cercedilla, Spain





Overview

- Linked Data: What and Why
- How to Publish Linked Data on the Web
- Linked Data Toolbox



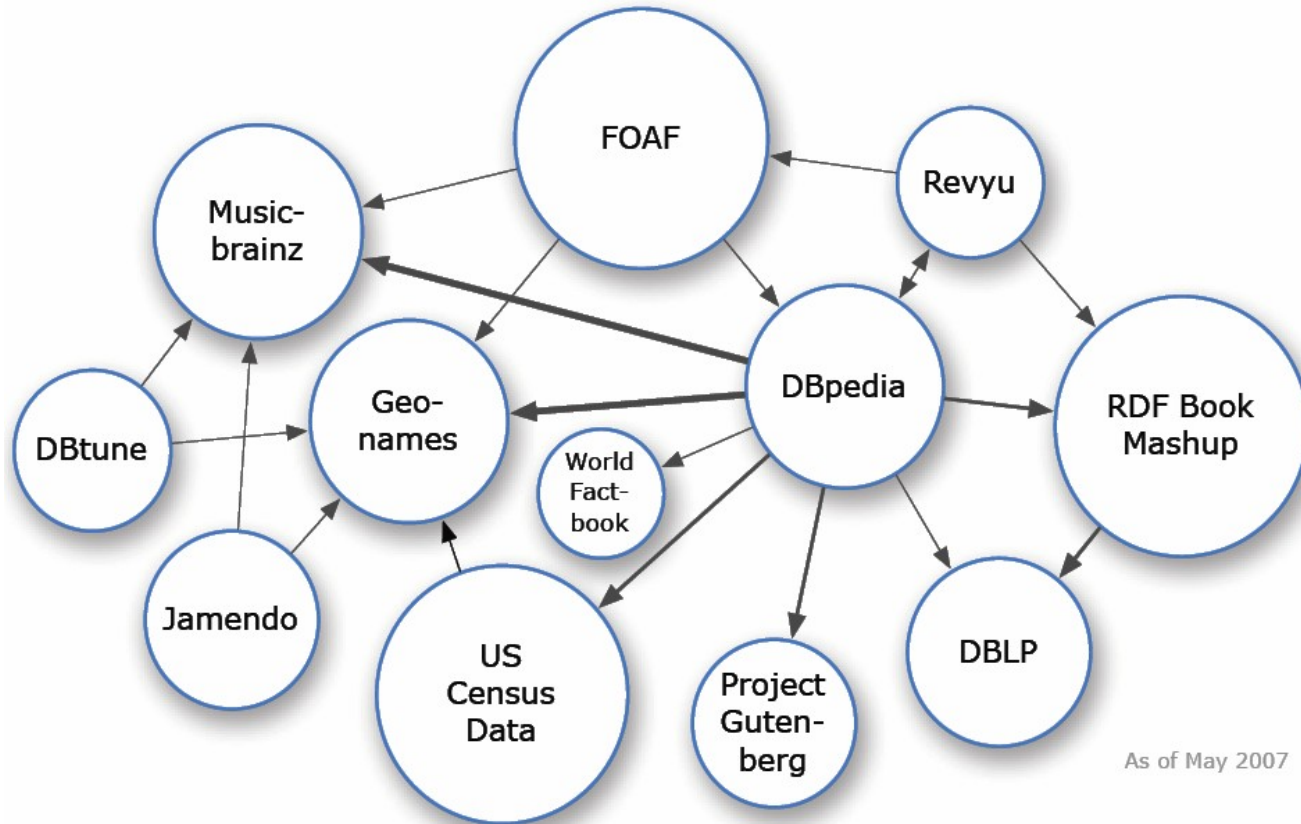
Linked Data: What and Why



Linked Data is...

- ...a way of publishing data on the Web that:
 - exploits the Web architecture and technology stack
 - reduces redundancy
 - facilitates reuse
 - enables discovery
 - maximises inter-connectedness of related things
 - enables network effects that add value to data
 - is experiencing rapid adoption (BBC, UK Gov, US Gov...)

The LOD "Cloud" - May 2007



As of May 2007



Linked Data Technology Stack

- URIs
- HTTP
- RDF
- (RDFS/OWL)



URIs – Not Just for Web Pages

- “A Uniform Resource Identifier (URI) provides a simple and extensible means for identifying a resource.” -- RFC 3986
- Many different schemes: *http://*, *ftp://*, *tel:*, *urn:*, *mailto:*
- Some URIs for “real world” things:
 - <http://tomheath.com/id/me>
 - http://dbpedia.org/resource/Talis_Group
 - <http://sws.geonames.org/4671654/>



HTTP

- Data access mechanism
- Using *http://* URIs to identify things allows people to look these things up



RDF: Resource Description Framework

- Generic data format for describing things and their interrelations



“Talis is Based Near Birmingham”

<http://dbpedia.org/resource/Talis_Group>

<http://xmlns.com/foaf/0.1/Person#based_near>

<<http://sws.geonames.org/3333125/>>



Linked Data Principles (TimBL, 2006)

- Use URIs as names for things
 - anything, not just documents
 - you are not your homepage
 - information resources and non-information resources
- Use HTTP URIs
 - globally unique names, distributed ownership
 - allows people to look up those names
- Provide useful information in RDF
 - when someone looks up a URI
- Include RDF links to other URIs
 - to enable discovery of related information



Why Publish Linked Data?

- For all the reasons stated before!



How to Publish Linked Data on the Web



Scenario

- Online whisky shop: *Wiskii.com*
- New business venture, founded by *Jeff*
- For the whisky connoisseur
- Detailed background information from experts
- Contributions from customers
- Custom web app, relational backend
- Simultaneous publication in HTML and RDF



6 Steps to Publishing Linked Data

1. Understand the Principles
2. Understand your Data
3. Choose URIs for Things in your Data
4. Setup Your Infrastructure
5. Link to other Data Sets
6. Describe and Publicise your Data



1. Understand the Principles

Linked Data Principles: Redux

- Use URIs as names for things
 - anything, not just documents
 - you are not your homepage
 - information resources and non-information resources
- Use HTTP URIs
 - globally unique names, distributed ownership
 - allows people to look up those names
- Provide useful information in RDF
 - when someone looks up a URI
- Include RDF links to other URIs
 - to enable discovery of related information



2. Understand your Data



2. Understand Your Data

- What are the key things present in your data?
 - People?
 - Places?
 - Books?
 - Films?
 - Musicians?
 - Concepts?
 - Photos?
 - Comments?
 - Reviews?
 - ...

2. Understand Your Data

- Things in the *Wiskii.com* database
 - Distilleries
 - Regions and Locations
 - Founders
 - Owners
 - Brands
 - Products
 - Photos
 - Reviews
 - Comments
 - Prices/Offers



DIAGEO

LVMH

MOËT HENNESSY • LOUIS VUITTON





2. Understand Your Data

- What vocabularies can be used to describe these?
 - Principles
 - Reuse, don't reinvent
 - Mix liberally
 - Potential Ontologies/Vocabularies
 - Geo
 - GoodRelations
 - FOAF
 - Review
 - SIOC
 - Whisky



3. Choose URIs for Things in Your Data

3. Choosing URIs: Principles

- Use HTTP URIs
- Keep out of other peoples' namespaces
 1. <http://www.imdb.com/title/tt0441773/>
 2. <http://www.imdb.com/title/tt0441773/thing>
 3. <http://myfilms.com/tt0441773>
 4. <http://myfilms.com/tt0441773/html>
- Abstract away from implementation details
 1. <http://dbpedia.org/resource/Berlin>
 2. <http://www4.wiwiss.fu-berlin.de:2020/demos/dbpedia/cgi-bin/resources.php?id=Berlin>
- Hash or Slash
 1. <http://mydomain.com/foaf.rdf#me>
 2. <http://mydomain.com/id/me>

3. Choosing URIs: Common Patterns

- http://dbpedia.org/resource/New_York_City ← Thing
- http://dbpedia.org/data/New_York_City ← RDF data
- http://dbpedia.org/page/New_York_City ← HTML page

- <http://revyu.com/people/tom> ← Thing
- <http://revyu.com/people/tom/about/rdf> ← RDF data
- <http://revyu.com/people/tom/about/html> ← HTML page

- <http://kmi.open.ac.uk/people/tom/> ← Thing
- <http://kmi.open.ac.uk/people/tom/rdf> ← RDF data
- <http://kmi.open.ac.uk/people/tom/html> ← HTML page

- <http://mydomain.com/thing> ← Thing
- <http://mydomain.com/thing.rdf> ← RDF data
- <http://mydomain.com/thing.html> ← HTML page



3. Choosing URIs: Wiskii.com

- <http://wiskii.com/regions/speyside>
- <http://wiskii.com/distilleries/talisker>
- <http://wiskii.com/brands/talisker>
- <http://wiskii.com/products/talisker-10-yo>
- <http://wiskii.com/products/glenmorangie-lasanta>
- <http://wiskii.com/people/william-matheson>
- <http://wiskii.com/photos/58>
- <http://wiskii.com/reviews/271>



3. Choosing URIs: Wiskii.com

- <http://wiskii.com/distilleries/talisker>
- <http://wiskii.com/distilleries/talisker/rdf>
- <http://wiskii.com/distilleries/talisker/html>

- <http://wiskii.com/brands/talisker>
- <http://wiskii.com/brands/talisker/rdf>
- <http://wiskii.com/brands/talisker/html>

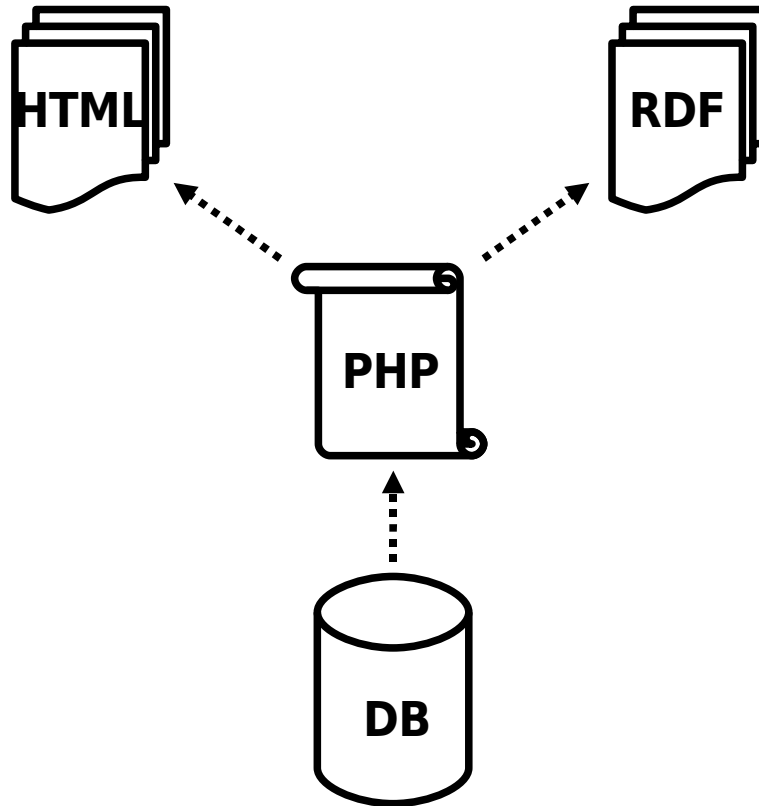
- <http://wiskii.com/people/william-matheson>
- <http://wiskii.com/people/william-matheson/rdf>
- <http://wiskii.com/people/william-matheson/html>

- <http://wiskii.com/photos/58>

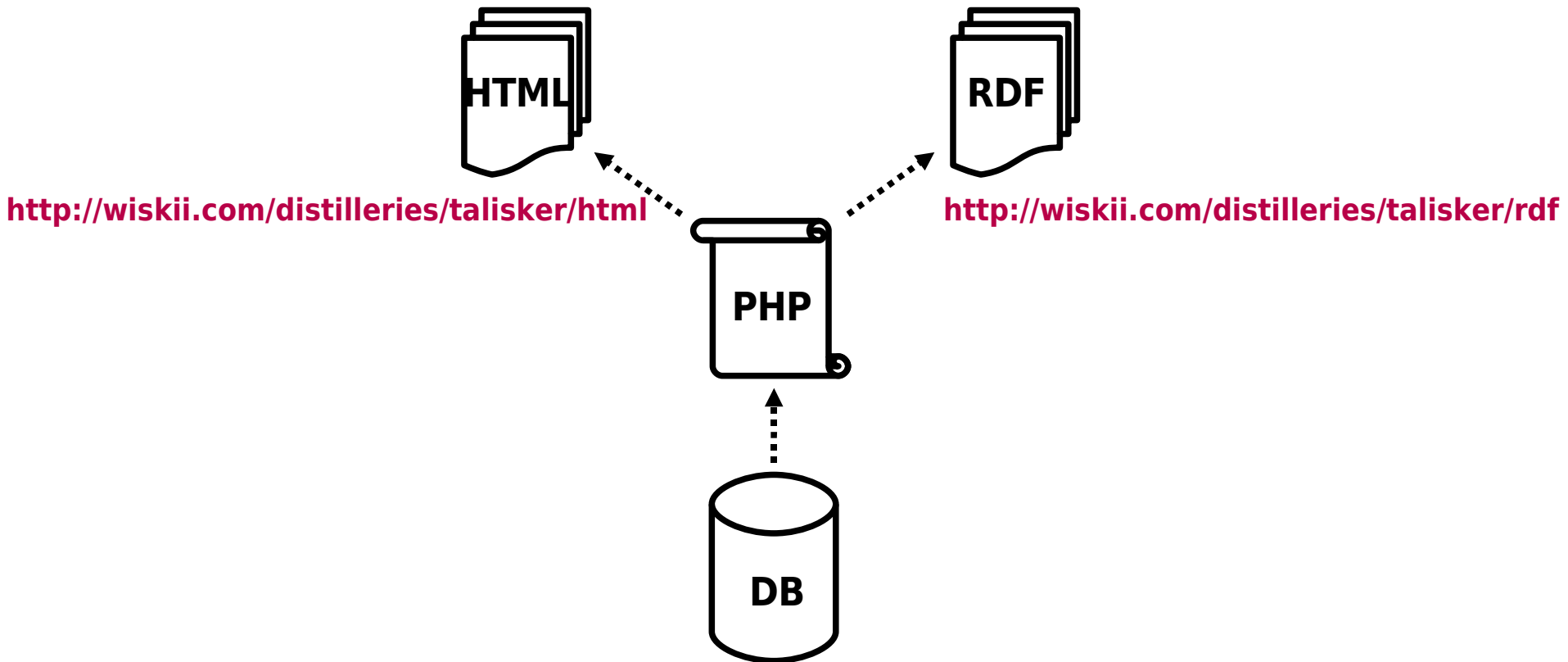


4. Setup Your Infrastructure

4. Setup Your Infrastructure

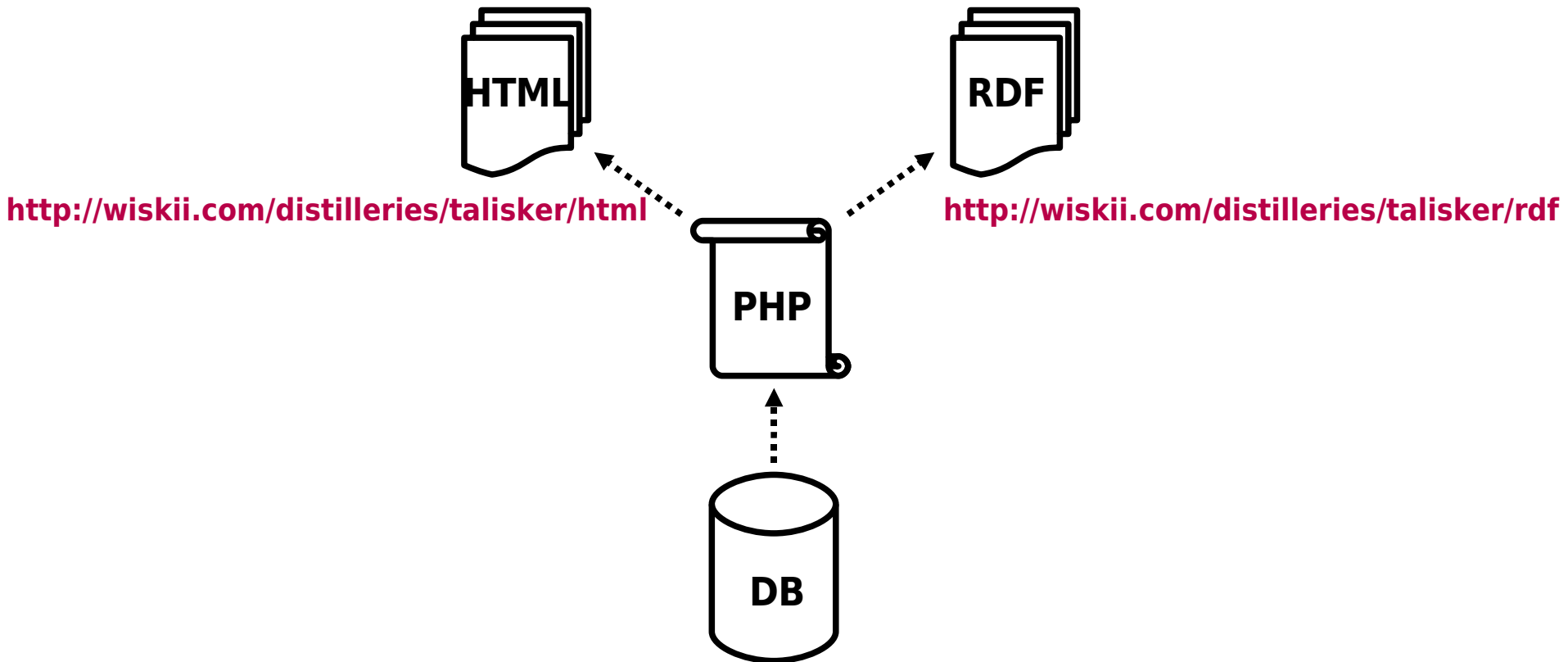


4. Setup Your Infrastructure



4. Setup Your Infrastructure

<http://wiskii.com/distilleries/talisker>



4. Setup Your Infrastructure

HTTP GET



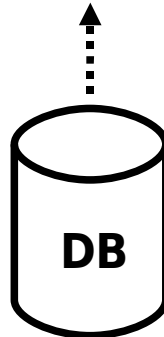
<http://wiskii.com/distilleries/talisker>



<http://wiskii.com/distilleries/talisker/html>

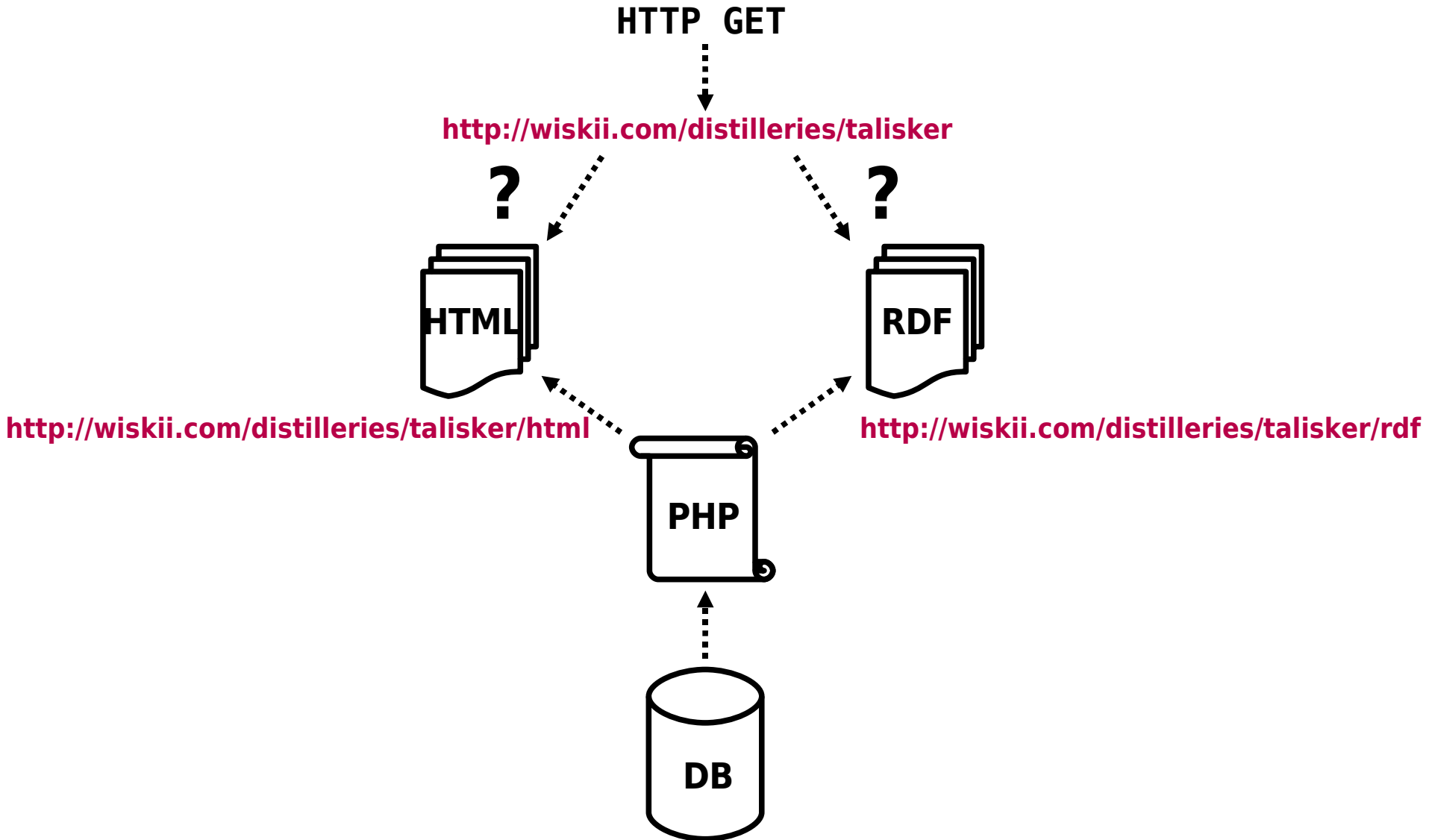


<http://wiskii.com/distilleries/talisker/rdf>

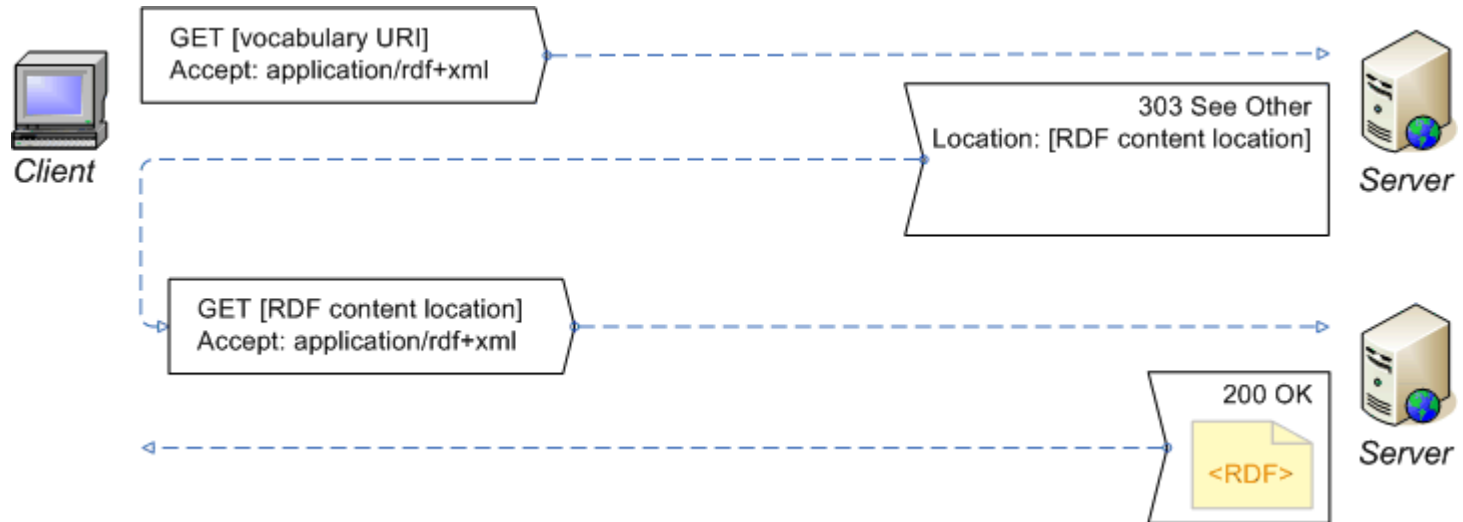


DB

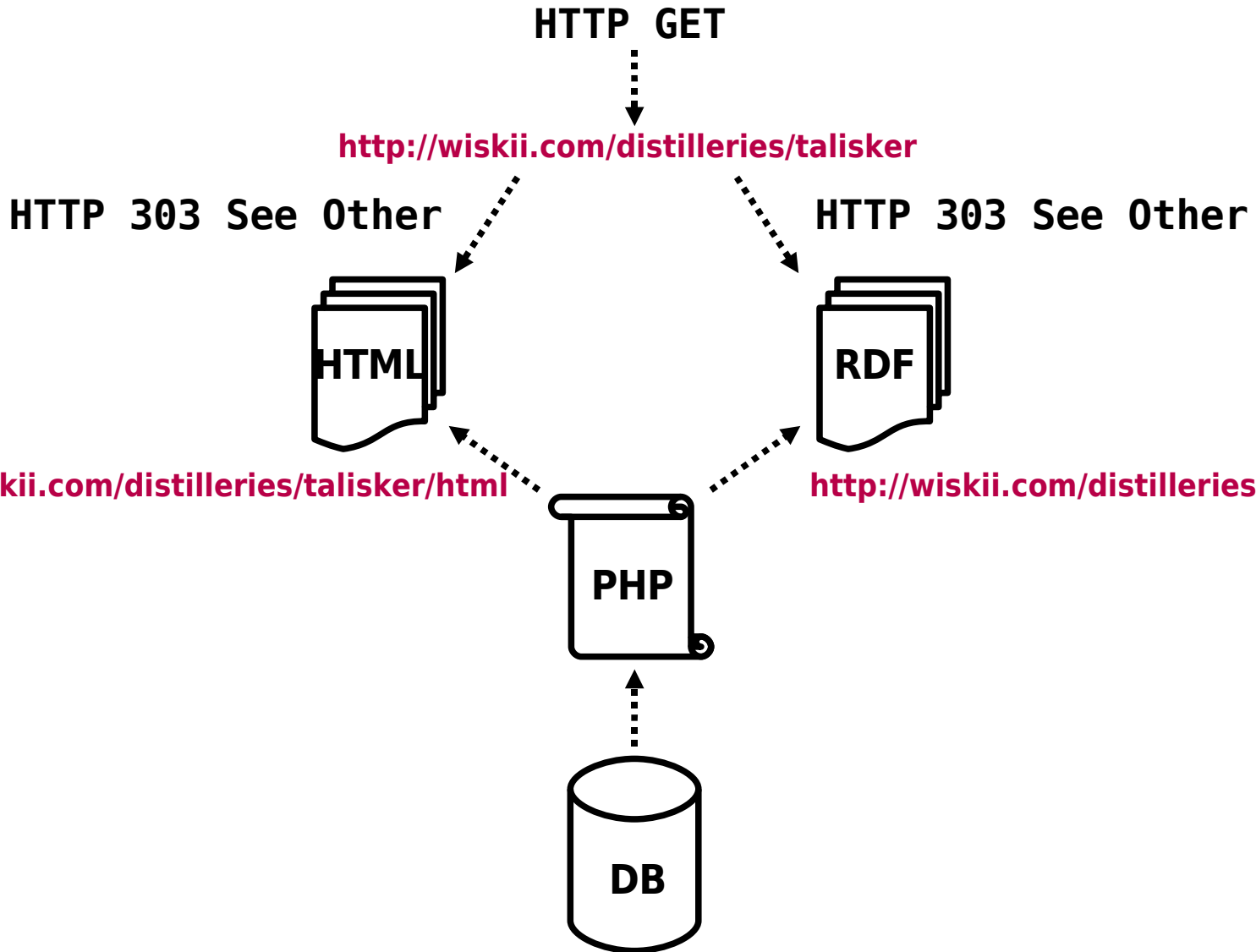
4. Setup Your Infrastructure



Content Negotiation



4. Setup Your Infrastructure



<http://wiskii.com/distilleries/talisker/html>

<http://wiskii.com/distilleries/talisker/rdf>

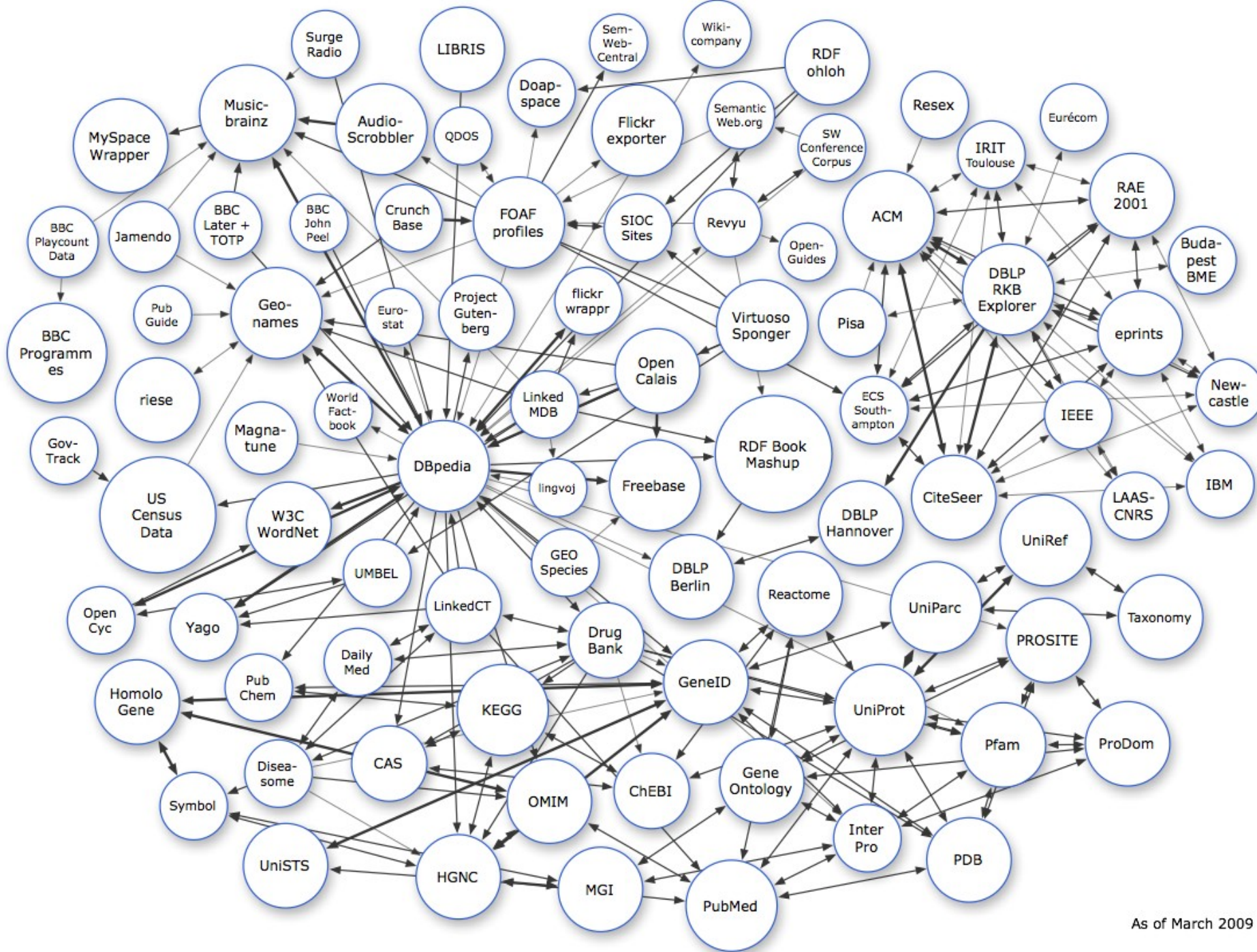


4. Setup Your Infrastructure

- Code samples for ConNeg and 303 Redirects
 - <http://linkeddata.org/tools>
- Useful tools for debugging
 - Firefox Extensions
 - Modify Headers, LiveHTTPHeaders
 - cURL
 - <http://dowhatimean.net/2007/02/debugging-semantic-web-sites-with-curl>
- You don't have to roll your own!
 - See Toolbox section below and <http://linkeddata.org/tools>



5. Link to Other Data Sets

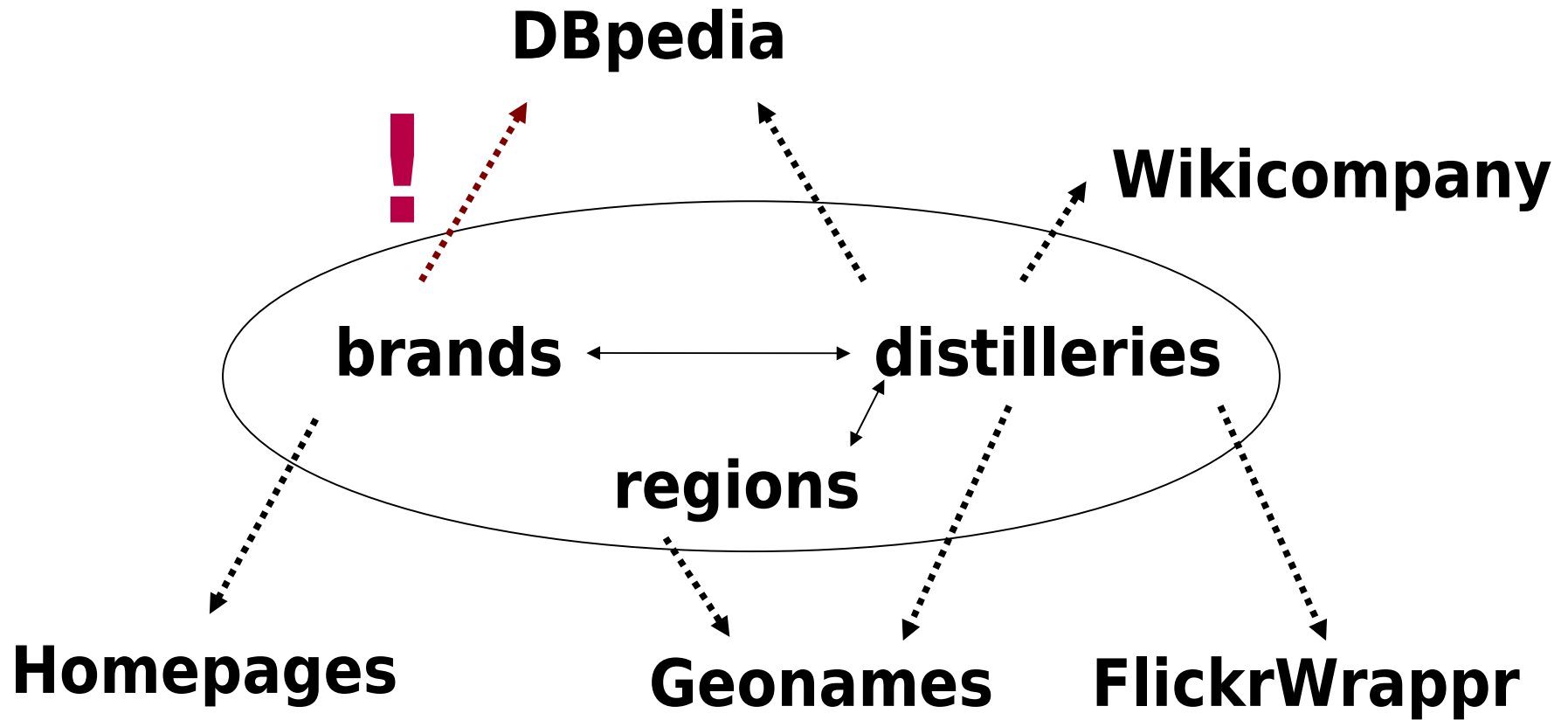




5. Link to other Data Sets

- Popular Generic Predicates for Linking
 - owl:sameAs
 - foaf:homepage
 - foaf:topic
 - foaf:based_near
 - foaf:maker/foaf:made
 - foaf:depiction
 - foaf:page
 - foaf:primaryTopic
 - rdfs:seeAlso

5. Link to other Data Sets



5. Link to other Data Sets

- Basic Linking Approaches
 - String Matching
 - e.g. comparing labels using similarity metrics
 - Common Key Matching
 - e.g. ISBN, Musicbrainz IDs
 - Graph Matching
 - Do these two things have the same label, type and coordinates
- Linking Frameworks
 - Silk: Volz et al., LDOW2009
 - LinQL: Hassanzadeh et al., LDOW2009
- Aim for reciprocal links

6. Describe and Publicise your Data

- Help others discover and index your data
 - Send pings to Sindice and pingthesemanticweb.com
 - Provide a Semantic Sitemap for your Data Set
 - Provide a void description of your Data Set
- Apply a license or waiver to your data set
 - Protects consumers of your data => encourages reuse
 - Creative Commons is probably **not** applicable
 - Use the Open Database License (ODbL) or release into the public domain by applying PDDL or CC0 waivers
 - <http://opendatacommons.org/>



Summary

1. Understand the Principles
2. Understand your Data
3. Choose URIs for Things in your Data
4. Setup Your Infrastructure
5. Link to other Data Sets
6. Describe and Publicise your Data



Linked Data Toolbox



Linked Data Storage/Publishing Layers

- D2R Server
 - Relational Database to RDF Middleware
 - SPARQL access to RDB
 - <http://www4.wiwiss.fu-berlin.de/bizer/d2r-server/>
 - Example:
 - LinkedMDB <http://linkedmdb.org/>



Linked Data Storage/Publishing Layers

- Virtuoso
 - Many things, including RDF triplestore
 - SPARQL access to data
 - Commercial and Open source editions
 - <http://virtuoso.openlinksw.com/>



Linked Data Storage/Publishing Layers

- Talis Platform
 - SaaS, cloud-based storage for RDF data and binary objects
 - SPARQL access
 - REST APIs to additional services
 - Faceting, Augmentation
 - Linked Data compatible out of the box
 - <http://www.talis.com/platform>
 - Connected Commons
 - Free hosting scheme for public domain data
 - <http://www.talis.com/platform/cc>



Linked Data Storage/Publishing Layers

- Paget Framework
 - publishing framework for Linked Data
 - serves up RDF according to Linked Data principles
 - reduces configuration overhead
 - can serve up data from static files or the Talis Platform
 - <http://code.google.com/p/paget>

Consuming Linked Data

- RDF Frameworks
 - ARC (PHP) <http://arc.semsol.org/>
 - RAP (PHP) <http://www4.wiwiss.fu-berlin.de/bizer/rdfapi/>
 - Jena (Java) <http://jena.sourceforge.net/>
 - Summary
 - <http://www.semanticscripting.org/SFSW2005/SFSW-Toolkits.pdf>
- Discovering more data
 - Watson: <http://watson.kmi.open.ac.uk/>
 - Sindice: <http://sindice.com/>
 - Squin: <http://squin.org/>

Outlook

- *Overview article*: Bizer, Heath and Berners-Lee (to appear) Linked Data – The Story So Far, IJSWIS
 - preprint available from <http://tomheath.com/publications>
- Synthesis e-Book on Linked Data
 - coming later this year
- LDOW2010 workshop at WWW2010?
- (Hopefully) very large amounts of Linked Data from UK Government



Questions?

- Contact Details
 - tom.heath@talis.com
 - <http://tomheath.com/>
 - <http://www.talis.com/>
 - @tomheath (identica) / @tommyh (twitter)
- Slides
 - <http://tomheath.com/slides/2009-07-cercedilla-how-to-publish-linked-data.pdf>
- Tutorial
 - <http://linkeddata.org/docs/how-to-publish>

shared innovation