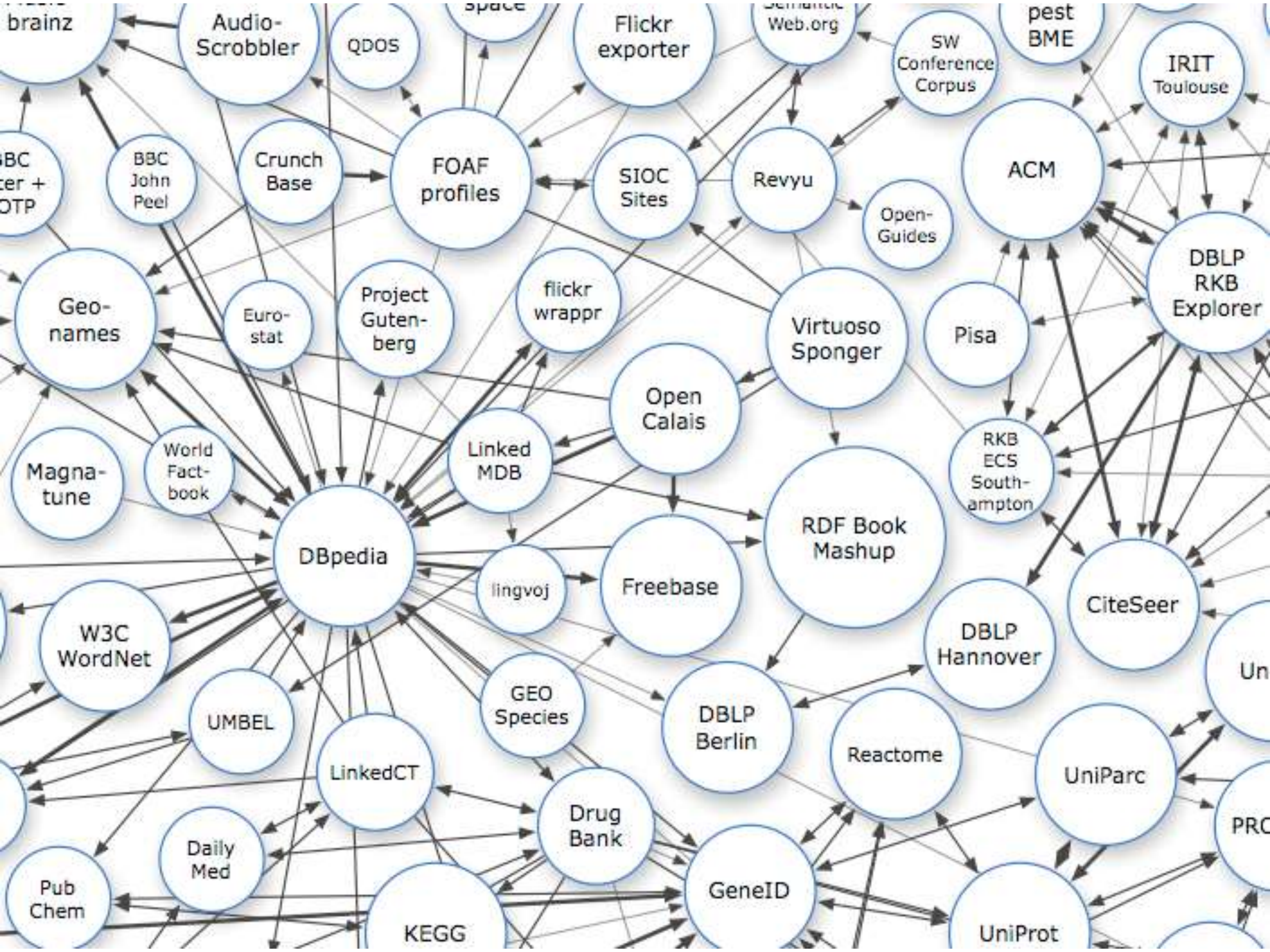
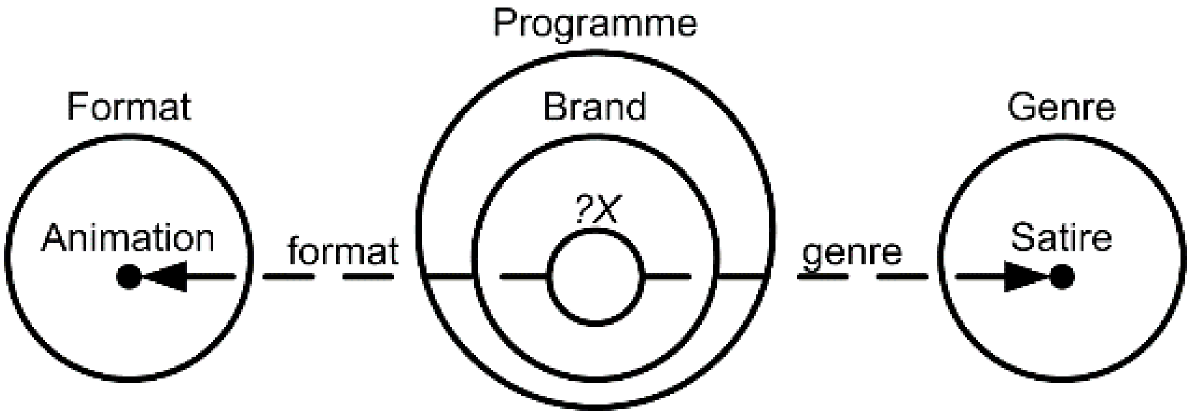

brainz
Audio-Scrobbler
QDOS
space
Flickr exporter
Web.org
pest BME
SW Conference Corpus
IRIT Toulouse
BBC
ter +
OTP
BBC John Peel
Crunch Base
FOAF profiles
SIOC Sites
Revyu
ACM
Open-Guides
DBLP RKB Explorer
Geo-names
Euro-stat
Project Guten-berg
flickr wrappr
Virtuoso Sponger
Pisa
Open Calais
Magna-tune
World Fact-book
Linked MDB
RKB ECS South-ampton
DBpedia
lingvoj
Freebase
RDF Book Mashup
CiteSeer
W3C WordNet
DBLP Hannover
Un
UMBEL
GEO Species
DBLP Berlin
Reactome
UniParc
LinkedCT
Drug Bank
PRO
Daily Med
Pub Chem
KEGG
GeneID
UniProt



Querying Ontologies with Diagrams

A overview of a submitted EPSRC First Grant proposal

Dr Jon Nicholson

School of Computing, Engineering and Mathematics

Summary

- Background (linked data, ontologies, SPARQL, etc.)

Why?

- The project summary

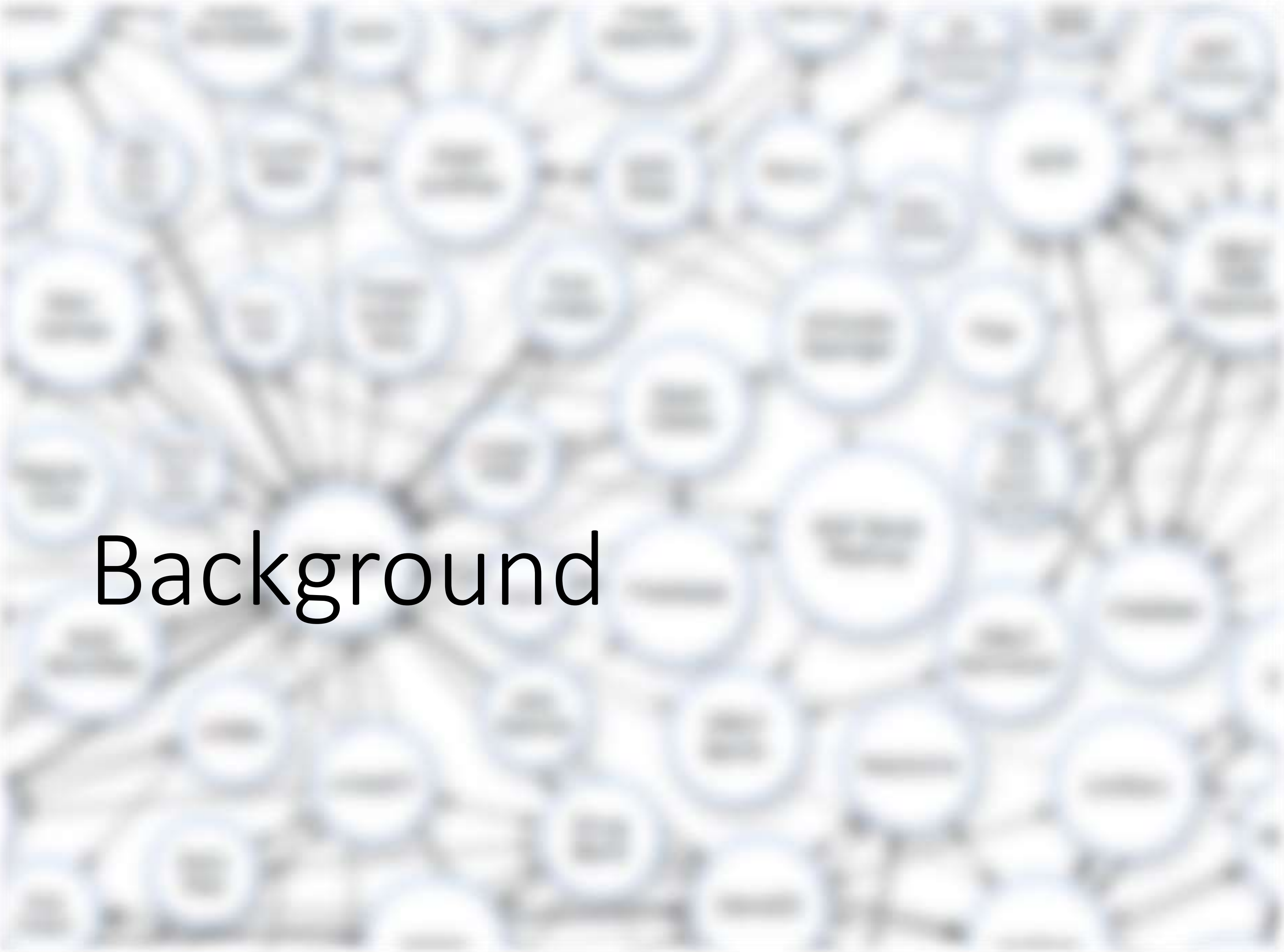
What?

- The notation

How?

- Work so far

When?

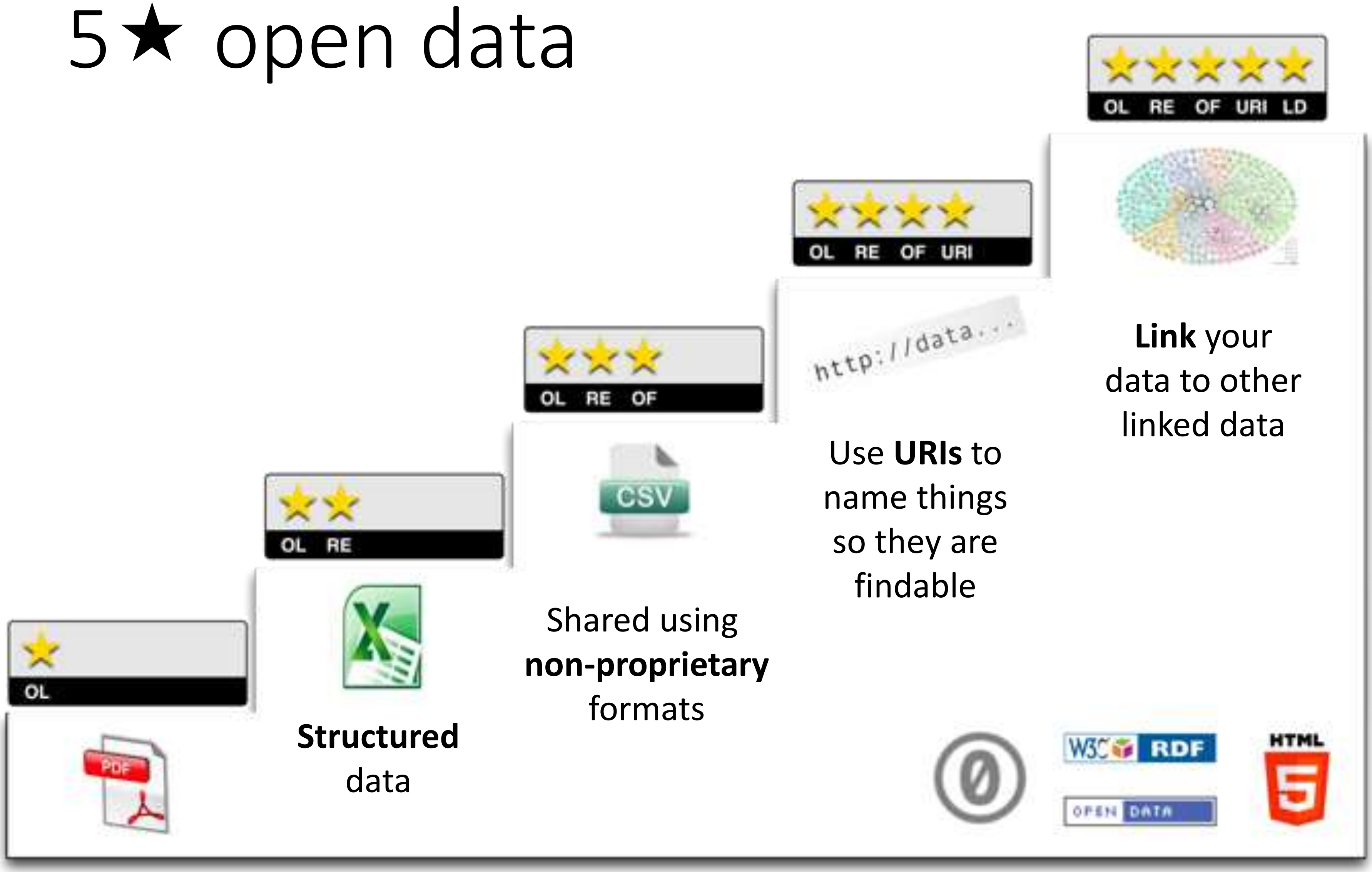


Background

Document- or Data-oriented

- We are all now familiar with the World Wide Web...?
- The WWW is a shared set of interconnected **documents** accessible over a global network (the internet)
    - A **document-oriented** approach, i.e. a "Web of Documents"
    - Designed for human consumption
    - Data not directly exposed
    - Problematic for machine processing
- Wouldn't it be great to reason over data from multiple sources?
    - To join data that is otherwise isolated
    - E.g. identify environmental and health factors that impact on education
- Tim Berners-Lee proposes the **Semantic Web**
    - See his 2009 TED talk: On the next web
    - A **data-oriented** approach, i.e. a "Web of Data"
    - Data linked to other data, i.e. "Linked Data"
    - Enables computers to do more useful work

5★ open data



Online data

Structured data

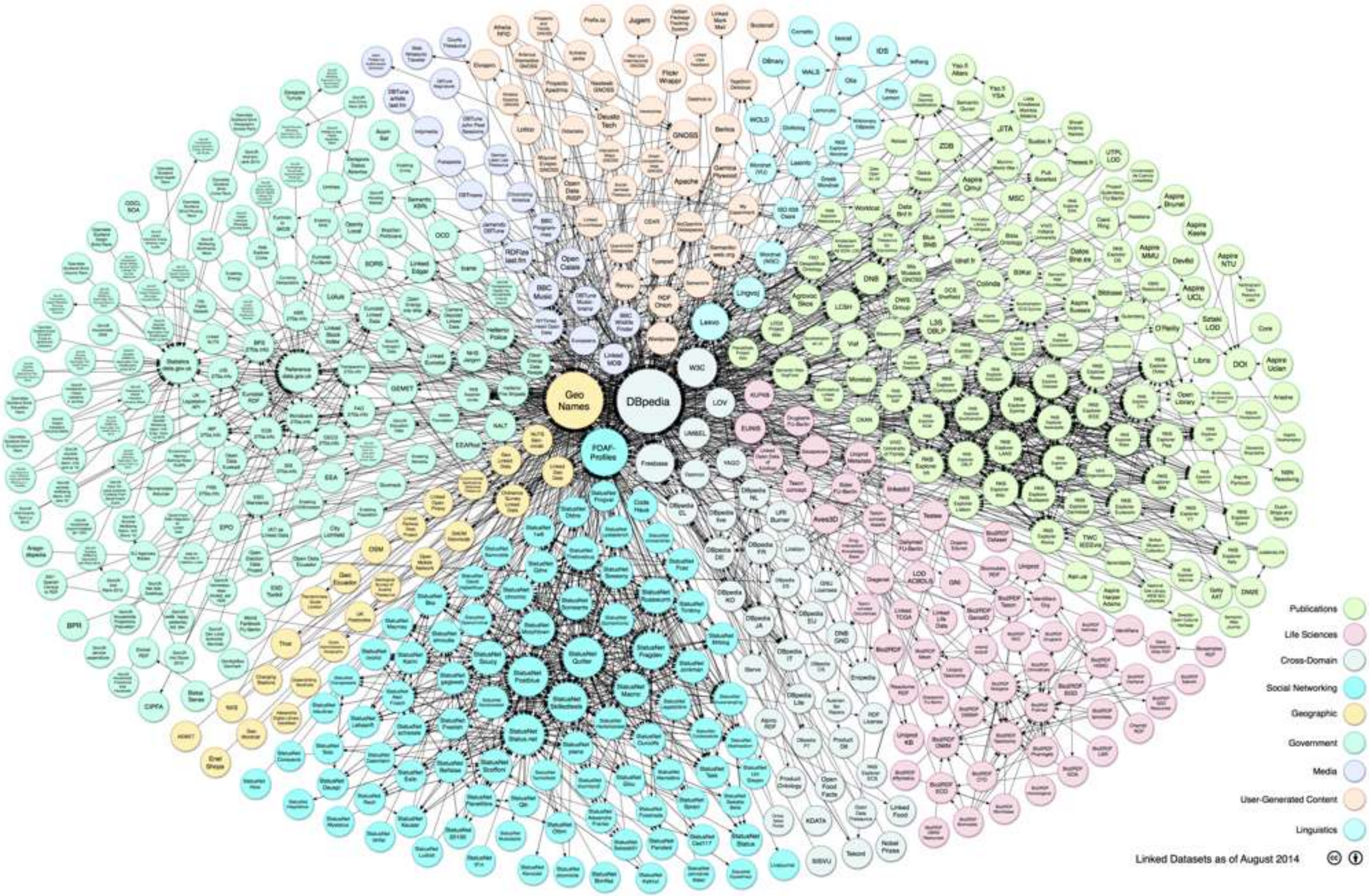
Shared using **non-proprietary** formats

Use **URIs** to name things so they are findable

Link your data to other linked data

Linked/open data examples

- **Data.gov.uk: Opening up government**
  - http://data.gov.uk/
  - Over 19,000 datasets (not all linked) published by the UK government
  - The EU, USA, etc., all committed to opening up their data (within reasonable limits)
- **Ordnance Survey**
  - http://data.ordnancesurvey.co.uk/
  - Great Britain's national mapping agency, providing the most accurate and up-to-date geographic data
- **BBC Programmes Ontology**
  - http://www.bbc.co.uk/ontologies/po
  - A vocabulary for programme (episodes, series, brands, etc.) data
- **DBPedia**
  - http://dbpedia.org/
  - Scrapes all the info boxes on Wikipedia pages into RDF format
- …And many more…



Source: http://lod-cloud.net/

Example uses

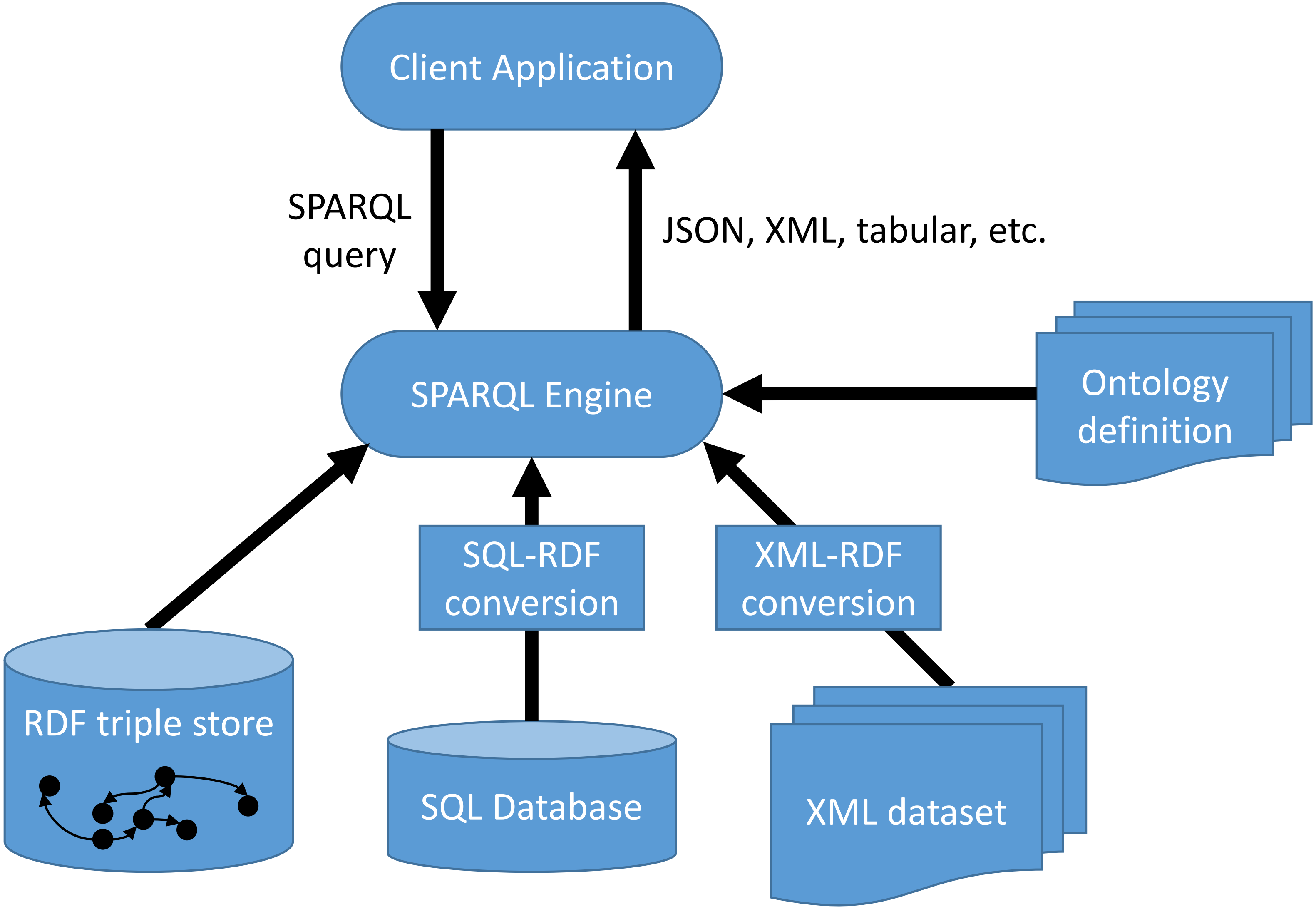
- Tim Berners-Lee's 2010 TED talk: The year open data went worldwide

- Lots available from data.gov.uk and others
- Police.uk
  - Lets you see crime statistics, street-by-street
- Local Authority Profiles
  - Provides a dashboard of local authority statistics, including deprivation, wellbeing, etc.
- Unistats
  - Statistics (including NSS and KIS data) of courses at university

How it works…

- **URI**: Uniform Resource Identifier
  - Web addresses (URLs) are common examples of URIs
  - A unique name that tells us where to find something
- **RDF**: Resource Description Framework
  - A very flexible notation for serializing data
  - Encodes data as triples forming a directed labelled graph
  - E.g. `Jon isa Lecturer`
- **Ontologies**
  - A way of describing data, often defined in formal mathematical notations such as description logics
  - Provides a vocabulary, constraints and rules
- **SPARQL**: **S**PARQL **P**rotocol **A**nd **R**DF **Q**uery **L**anguage
  - A query language for retrieving and manipulating RDF
  - Some implementations can infer information from the ontology def.

How it works…


Client Application
SPARQL query
JSON, XML, tabular, etc.
SPARQL Engine
Ontology definition
SQL-RDF conversion
XML-RDF conversion
RDF triple store
SQL Database
XML dataset

SPARQL example

- Get BBC satirical quiz/panel shows:

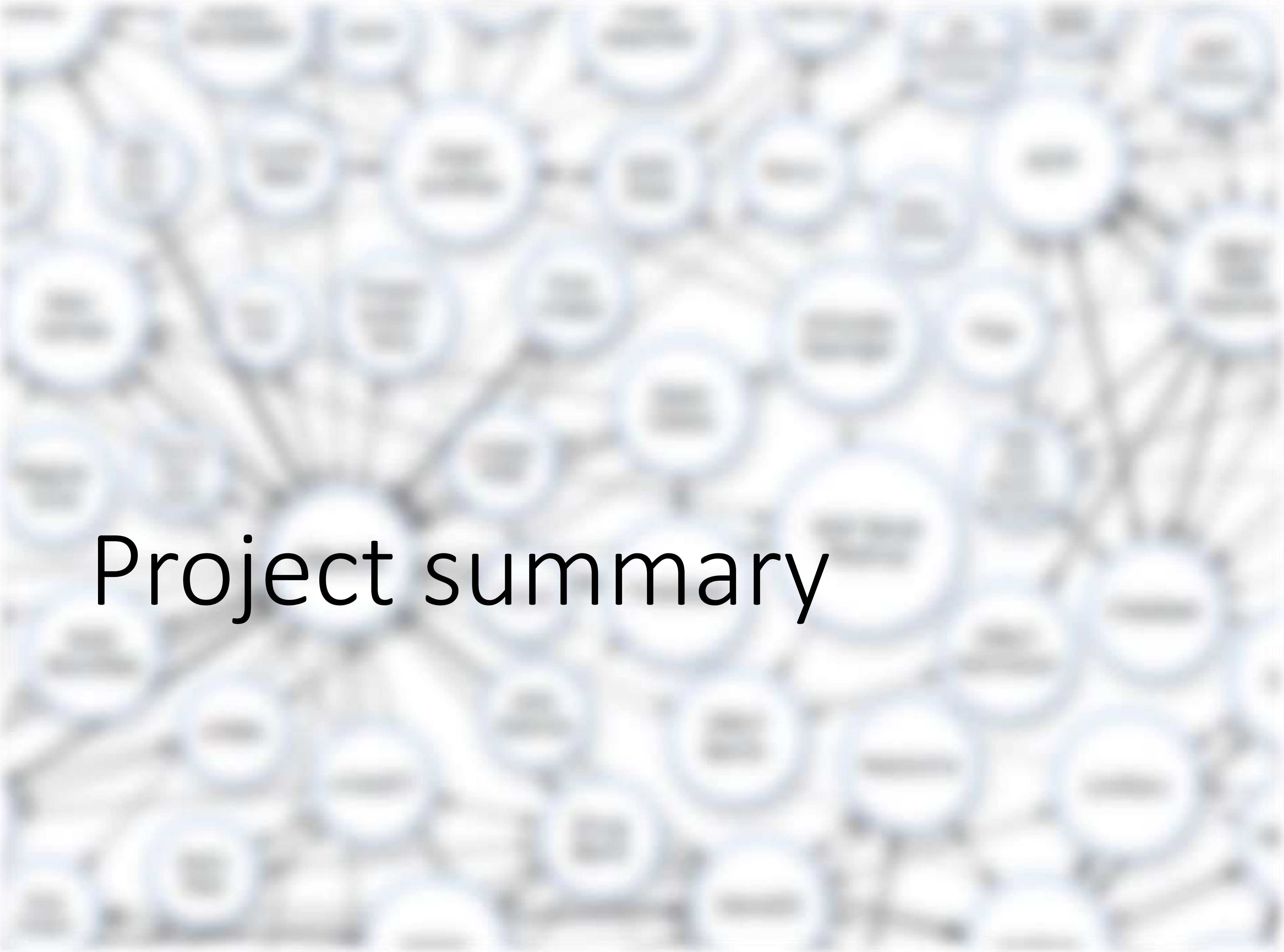
```
PREFIX po: <http://purl.org/ontology/po/>
PREFIX dc: <http://purl.org/dc/elements/1.1/>
SELECT ?uri ?title
WHERE {
  ?uri po:genre <bbc/genres/comedy/satire#genre>.
  ?uri po:format <bbc/formats/gamesandquizzes#format>.
  dc:title ?title
}
```

- Result (e.g.):

```
bbc/b006mkw3 "Have I Got News for You"
```

bbc/… URLs, replace with `http://www.bbc.co.uk/programmes`

Project summary



Proposal details

- 125k project (100k to be funded by the EPSRC)

- Submitted September 2014
  - Have just received reviews and am preparing a response

- This funds:
  - My time as principle investigator, 1 year, 1 day a week
  - 1x post-doctoral research fellow, 1 year, full time
  - 2x visits to project partner (Nokia, in Helsinki)
  - 3x conferences

- Nokia is a project partner
  - Providing staff time supporting the project
  - Helps to ensure relevance with industry

NOKIA

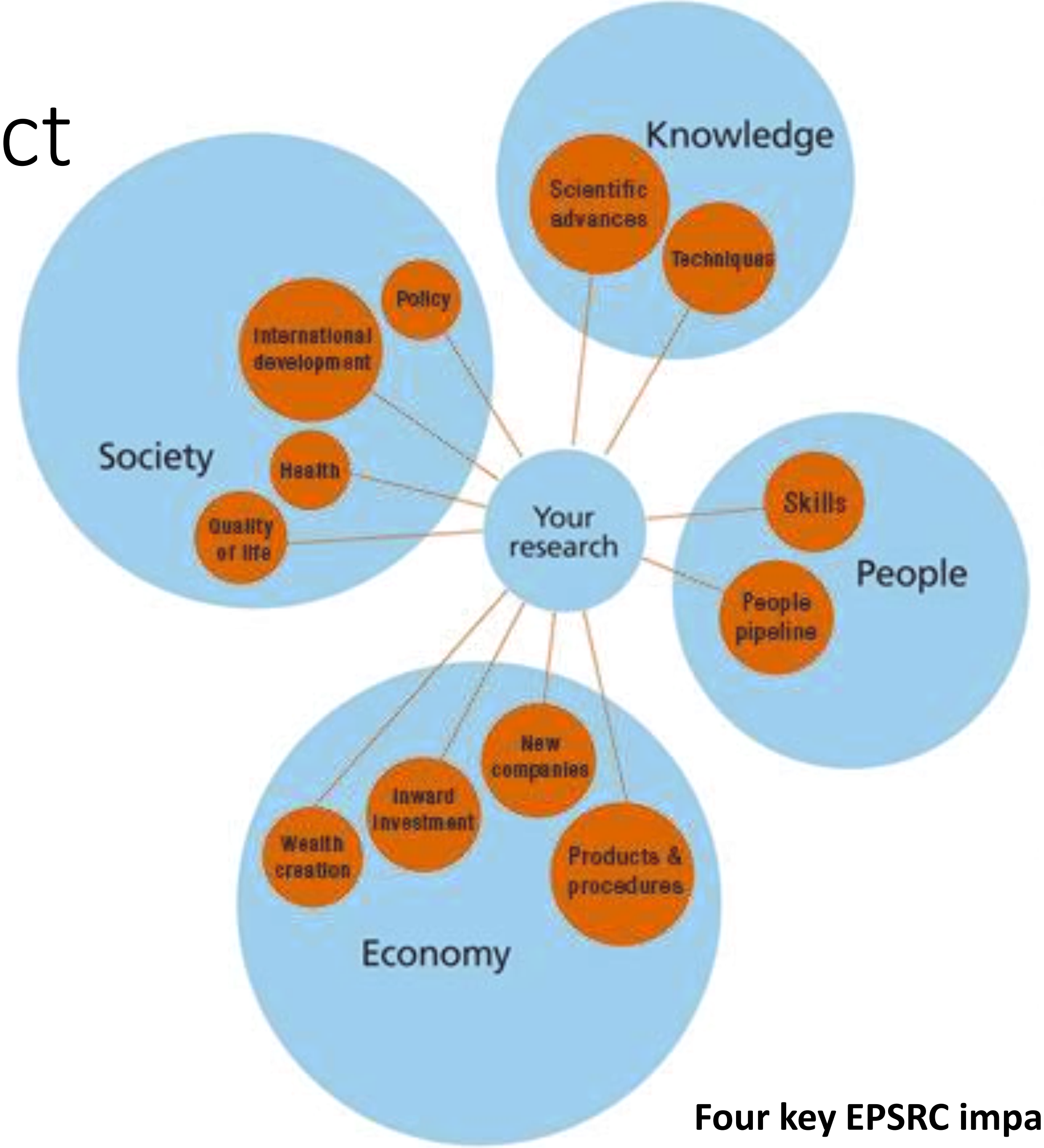
What is the project going to do?

- Manipulating populated ontologies/linked data is not easy, lot of technical symbolic notations and jargon…
- **Aim**
  To make querying populated ontologies (i.e linked data) more accessible to diverse stakeholders such as software engineers, lawyers, data analysts, marketing personnel, and managers
- **Hypothesis**
  We can devise an accessible formal diagrammatic notation for querying ontologies

Stakeholders

- **Ontology engineers**
  - Construct and query ontologies
  - Most likely to be trained in notations (e.g. symbolic/textual) used to define (e.g. OWL2) and query (e.g. SPARQL) ontologies
- **Domain experts**
  - Work with ontologies *in their field*
  - May not be trained in languages like SPARQL
  - Rely on ontology/software engineers for query construction
- **Other end users of ontologies**
  - Those with a need to extract information from populated ontologies for analysis
  - Perhaps the furthest removed from the definition of the underlying ontology and the logics used to define them
  - E.g. data analysts, marketing professionals and managers
- **The project team and other academic researchers**
  - The Visual Modelling Group (VMG)
  - Wider academic community in ontology engineering and diagrammatic logics

Impact



Four key EPSRC impact indicators

Impact: examples

- Proposal shows each key area of impact associated with each identified stakeholder group
- Too big to fit on a slide…

- Examples:
  - People/Ontology engineers
    - Enables increased communication between technical and non-technical staff, promoting efficient use of the skills of both (medium term)
  - Economy/Other users
    - Simplifies training required to interrogate large datasets, reducing costs to business, particularly useful to new and small businesses (long term)

Impact strategy

- Short-term (0-2 years)
  - Focuses on **project team**
  - Primarily through collaboration with project partner
  - Project web site to provide materials
- Medium-term (2-5 years)
  - Focuses on **software and ontology engineers**
  - Develop tools, such as a plugin for Protégé
  - Further publications in journals etc.
- Long-term (5+ years)
  - Focuses on **domain experts** and **general public**
  - Look to develop more tools targeting more general use
    - E.g. a web based tool made available through data.gov.uk
  - But ultimately dependant on interest

So what are the objectives?

1. Design a diagrammatic query notation informed by SPARQL
    - Based on *Concept Diagrams*
    - Informed by queries used in industry, i.e. with project partner Nokia
2. Formalise the diagrammatic query notation
    - Including formal mappings between SPARQL and the new query notation
    - Establish that all of SPARQL can be expressed
3. Evaluate the accessibility of the new query notation
    - Empirical study of the relative accessibility of the new diagrammatic query notation for **formulating queries** as compared to SPARQL
    - Empirical study of the relative accessibility of **representing results** diagrammatically as compared to the textual output from SPARQL
    - Accessibility measured in terms of performance, e.g. time taken and error rate

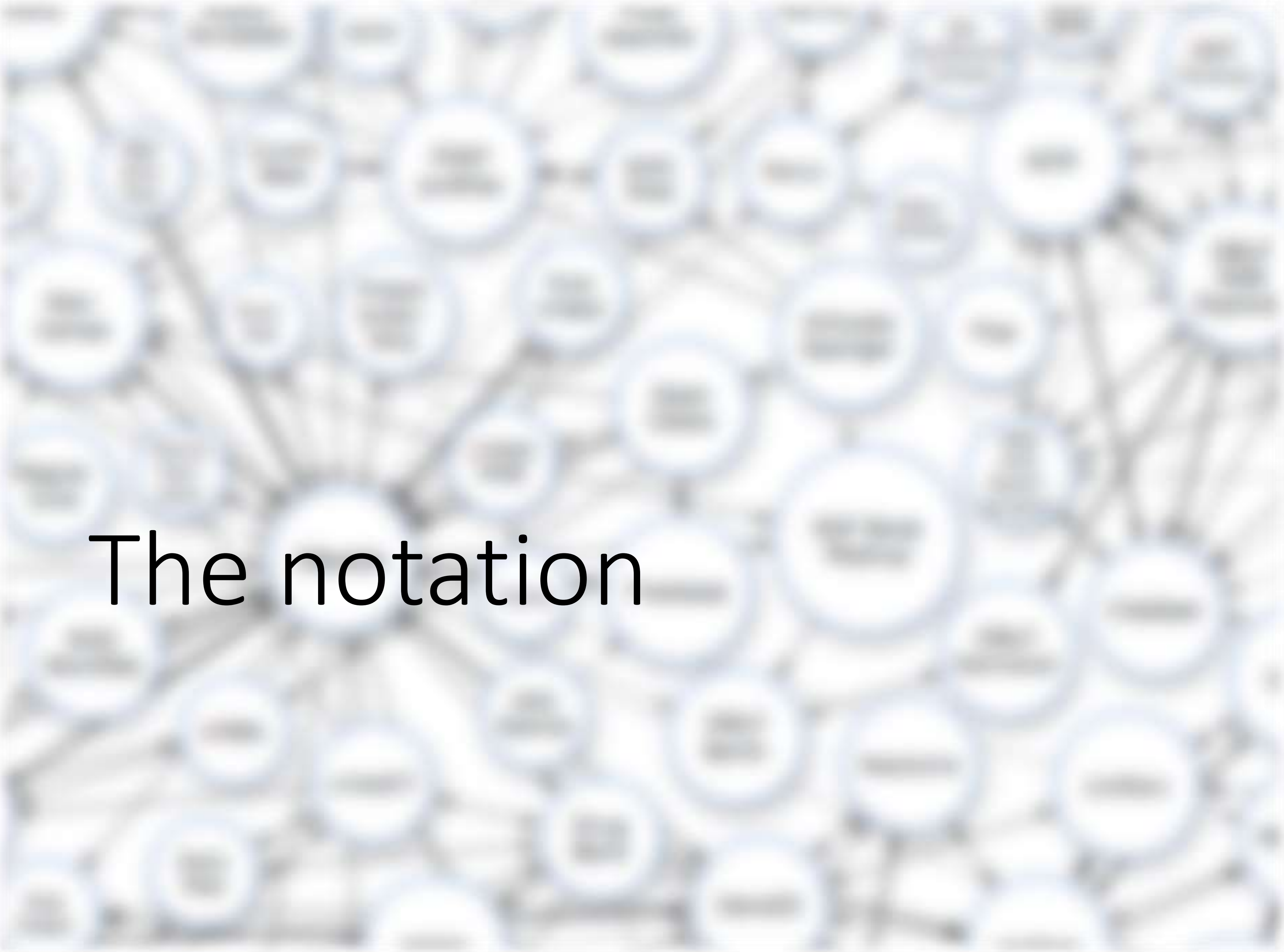
Challenges

- Covering all of SPARQL is very ambitious
  - Several types of queries
    - SELECT queries (tell me specific things about x)
    - DESCRIBE queries (what can I find out about x?)
    - ASK queries (does some property hold for x?)
  - Optionality (give me information if it exists)
  - Several types of conditions
    - FILTERs, regular expressions
  - Combination of data sources (linked data)

- Deciding how to represent these in a formal **and accessible** way is not easy
  - Likely to be trade-offs between expressiveness and clarity
  - But I believe that the basic building blocks are there…

Related work

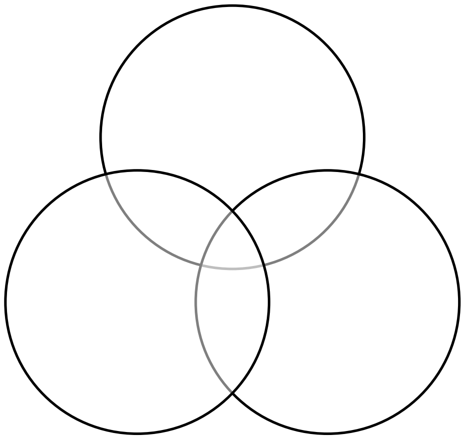
- Diagrammatic query notations / languages / visualizations are not new
  - Kaleidoquery for object databases (OCL)
  - Visionary for relational databases (SQL)
  - XQBE for XML documents (XQuery)
  - Konduit VQB, OptiqueVQS, and MashQL for linked data (SPARQL)
- Important because "domain experts mostly do not possess necessary competences to formulate queries by using structured query languages”
  - A. Soylu, M. Giese, E. Jimenez-Ruiz, E. Kharlamov, D. Zheleznyakov, and I. Horrocks. OptiqueVQS: Towards an ontology-based visual query system for big data. In 5th Int. Conf. on Management of Emergent Digital EcoSystems, pages 119–126. ACM, 2013.
- But they typically focus on building the query only
- This project's novelty is for the query **and** its results to be presented using the same diagrammatic notation
  - I theorise that presenting the query and its results in the same format will help users understand the context of the results making it easier to interpret



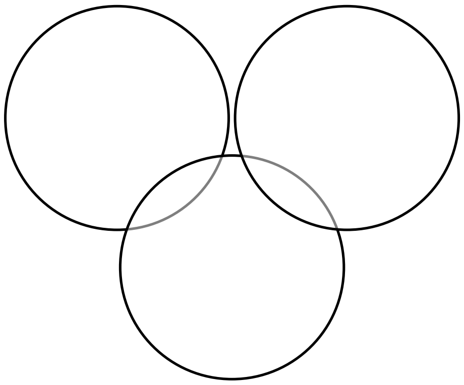
The notation

The notation

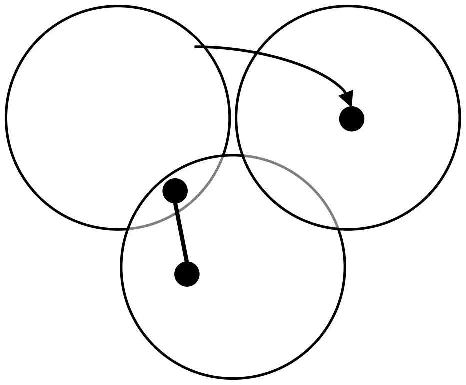
- Will be based on **Concept Diagrams**
  - Developed by J.Howse, G.Stapleton, et al. in the Visual Modelling Group (CEM, UoB)
  - Roughly speaking, Concept diagrams are based on Euler diagrams with the addition of individuals and edges



Venn diagram



Euler diagram



Concept diagram

Concept Diagrams

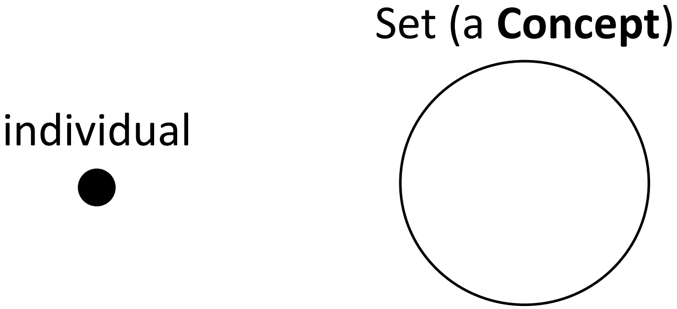
- Concept Diagrams have a formal syntax and semantics

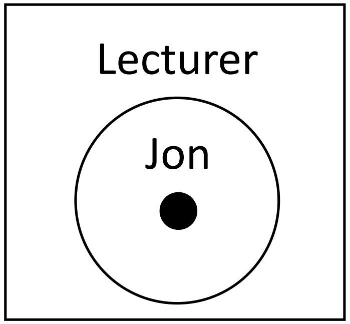
- Have been shown to be useful for describing/defining ontologies
    - Nokia use them for privacy engineering

- Concept diagrams have been said to be the exception, where most languages are either not accessible or are not formal artefacts
    - P. Warren, P. Mulholland, T. Collins, and E. Motta. The usability of description logics: Understanding the cognitive difficulties presented by description logics. In The Semantic Web: Trends and Challenges, LNCS 8465:550–564, Springer, 2014.

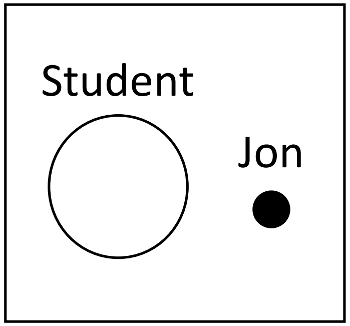
- I believe they have the potential to make an intuitive and accessible diagrammatic query notation

Concept Diagrams: a crash course

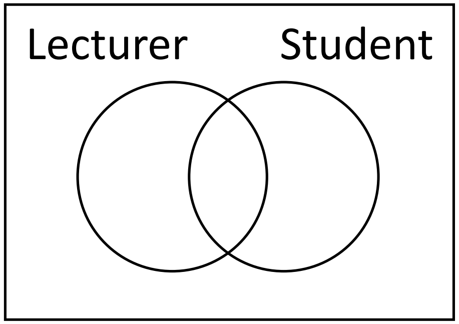




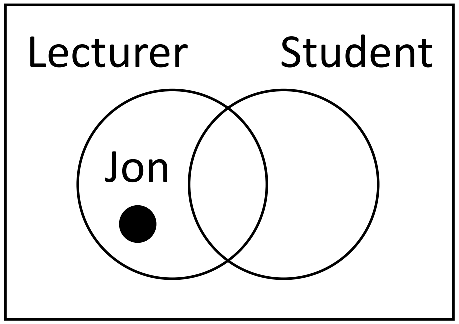
Jon is a Lecturer



Jon is not a Student

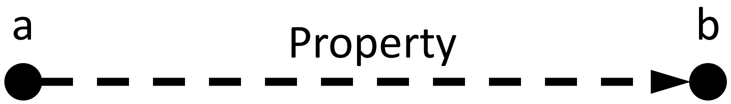


Some Lecturers are Students

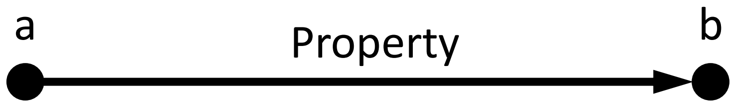


Jon is a Lecturer, but not a Student

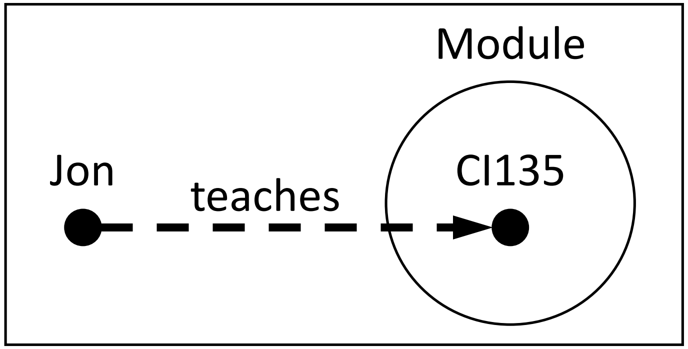
Concept Diagrams: a crash course



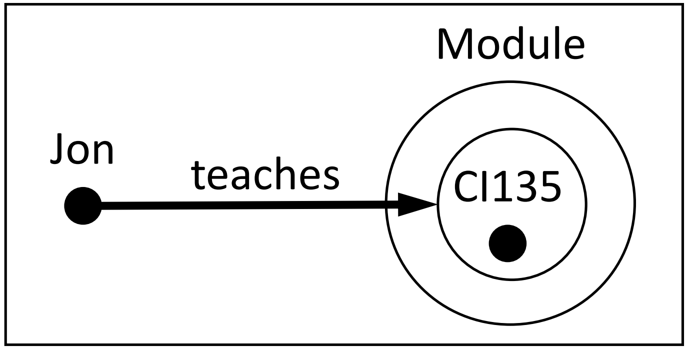
a relates to at least b under property



a relates to exactly b under property



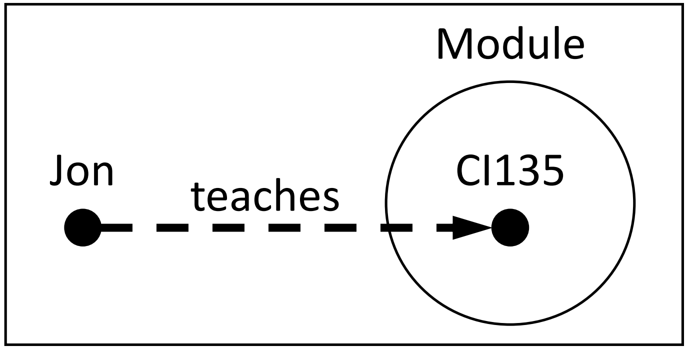
Jon teaches the module CI135, and possibly others things that may not be modules



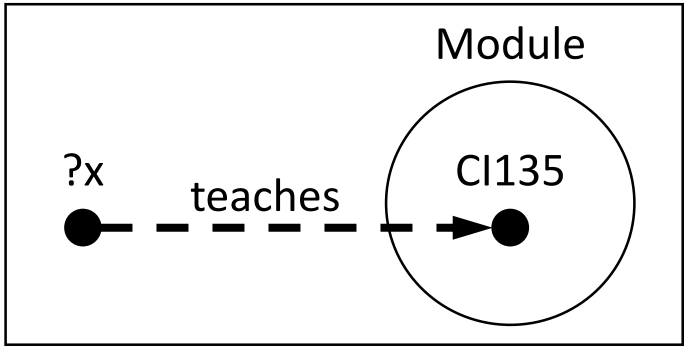
Jon teaches only modules, which includes CI135

A simple example

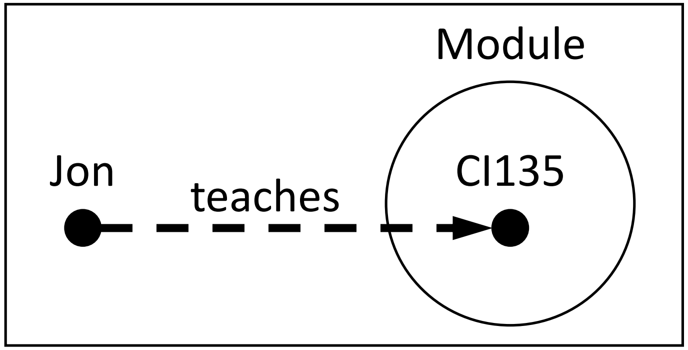
Specification



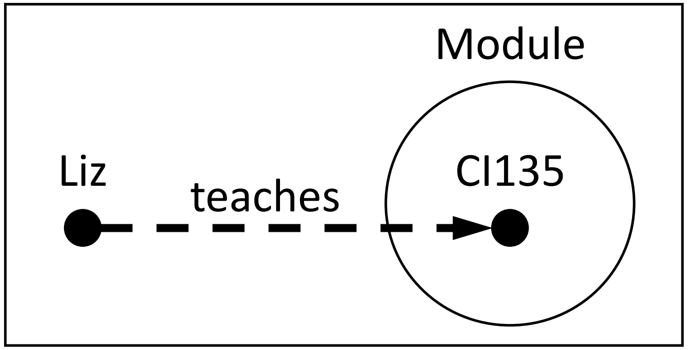
Query



Results

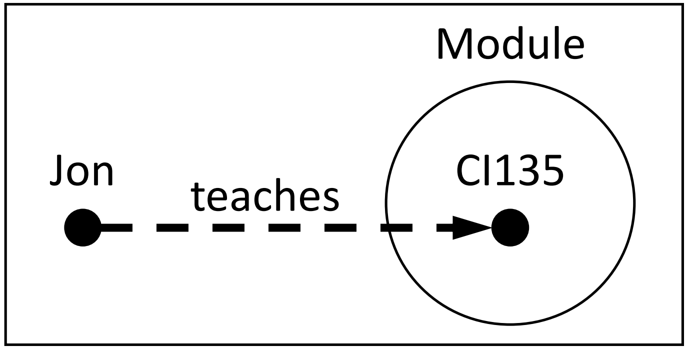


$\wedge$

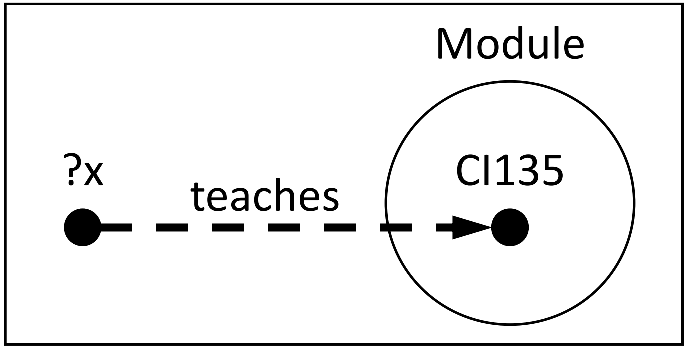


A simple example

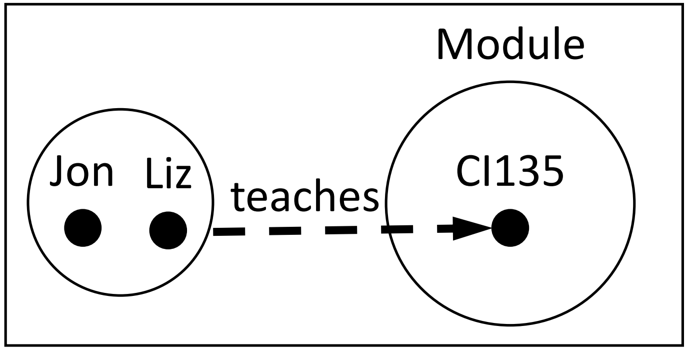
Specification



Query



Results



This is a very simple example

It's not clear yet how other parts of SPARQL will be captured in the notation

Work so far

Progress…

- Submitted the proposal (Sept 2014)
- Received 3 peer reviews (end of Jan 2015)
- Responded to reviews (start of Feb 2015)
- Now awaiting a panel date…

Work on the notation

- The project is only a year long
  - Will have a standing start to some extent

- Trying to get started now..
  - Identifying a small selection of SPARQL examples
  - Drawing diagrams for these
  - Highlights some of the difficulties that need to be addressed

- Have so far identified a need for a class, often only represented only as a concept (i.e. a set of its members), to be treated as an individual in some cases

Thank you

Questions welcome.