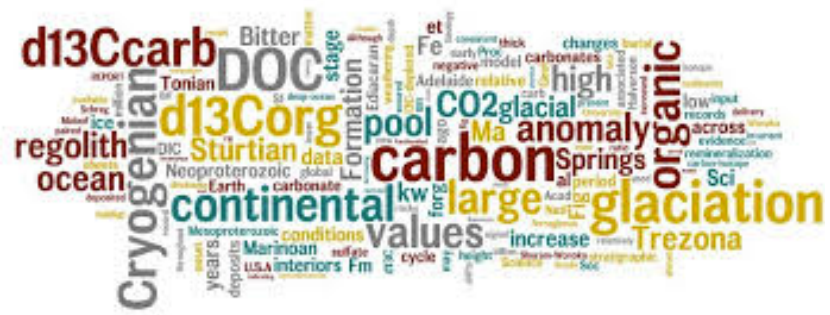
Digital by design: science and innovation policymaking in an era of digital government

Semantic technologies and ontologies in the KNOWMAK project

Dr. Diana Maynard
University of Sheffield, UK

THE KNOWMAK PROJECT

- 3-year EU H2020 project since January 2017
- Develop a web-based tool providing interactive visualizations and indicators on knowledge co-creation in the European research area
- Based around:
  - **Research Actors** (organisations)
  - **Research topics** (projects, publications, patents, social innovation, based on SGC and KET)
  - **Geographical spaces** (based on NUTS and FUA)



SEMANTIC TECHNOLOGIES IN THE S&T LANDSCAPE

- Opportunities:
  - Ability to link different kinds of data sources to provide a richer view of knowledge production

- Challenges
  - Need for a robust approach to identify and model relevant topics
    - **Language** (connect different kinds of data due to terminology differences)
    - **Commensurability** (cannot connect different kinds of classifications)
    - **Flexibility** (model changes over time and space)

Data

SEMANTIC APPROACH

What is the innovation performance of France on climate change compared with Germany?

- climate
  - air_quality_management
  - biomaterials
  - forests_flora_fauna
  - noise
  - soil
  - waste_management_and_recycl
  - water_and_wastewater

6687 2007 0
LED module with gold bonding.
Processes or apparatus specially adapted for the manufacture or treatment of semiconductor

SC5-20-2014
H2020
Zero Emission Robot-Boat for Coastal and Inland Water Monitoring

ghg

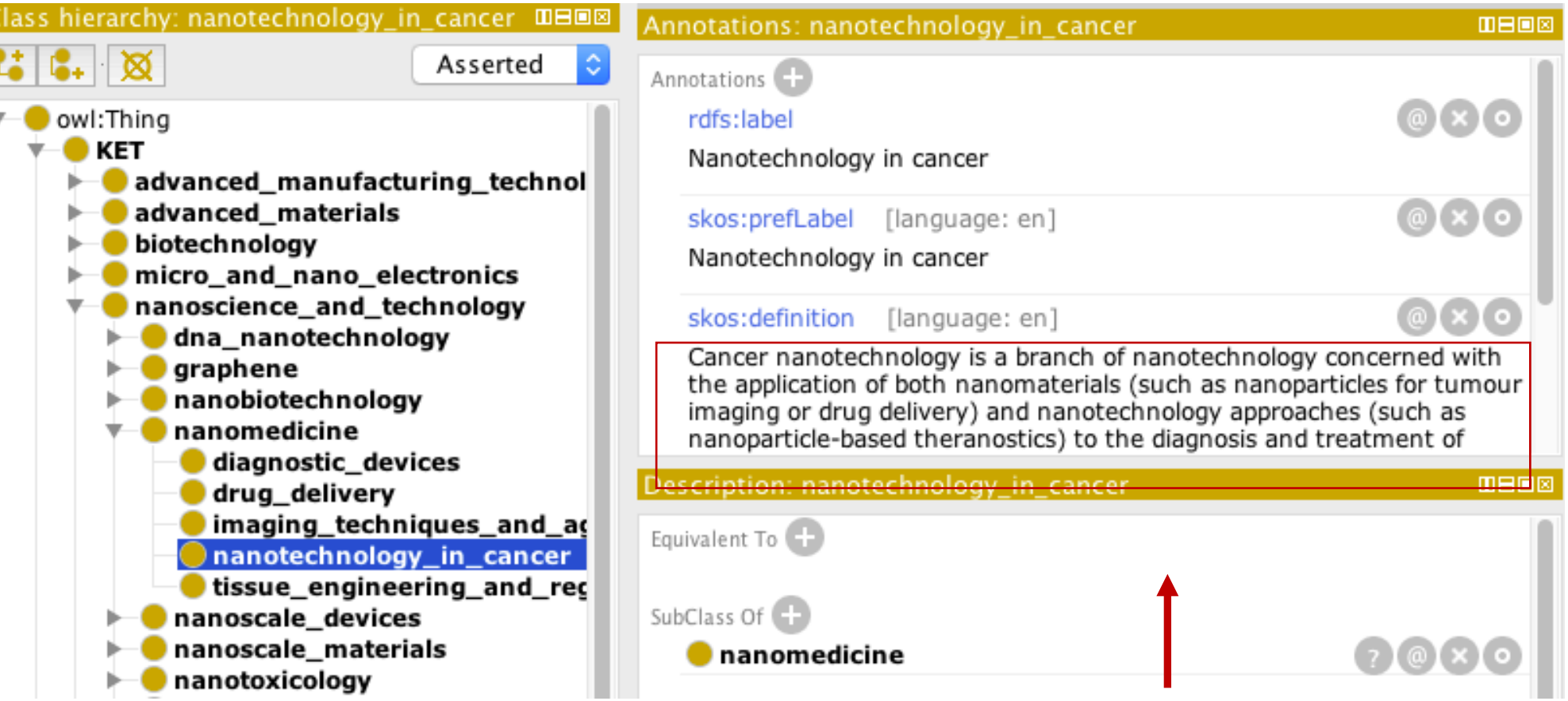
Perspectives on CO2 capture and storage
Filipp Johnsson
Published 14-04-11

Policy ⟷ Ontology ⟵ Data

ONTOLOGIES

- Translate generic user queries related to policy-making into a formal structure of classes and keywords linked to data sources
- Offer a flexible solution allowing
  - variations of language and terminology
  - connections between concepts (at both the topic and keyword level)
  - adaptability over time and topics of interest
  - different levels of aggregation
  - minimal user input when changes are required

ONTOLOGIES CONNECT INFORMATION



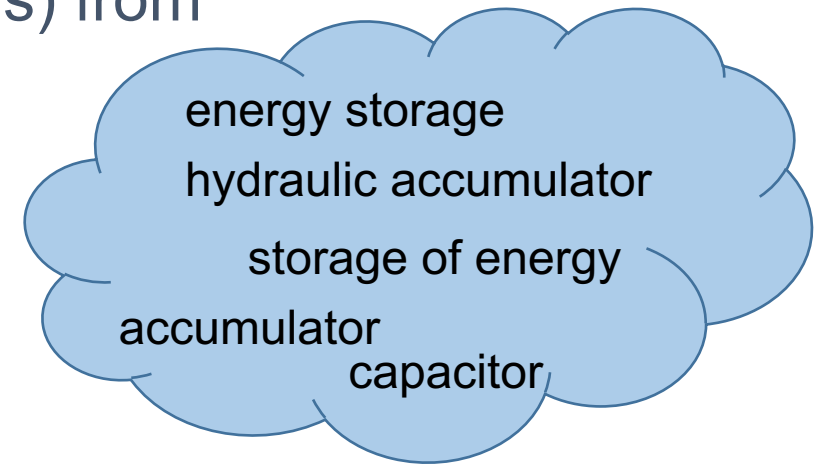
Link with other sources (Nature.com, skos, DBpedia…)

Link related topics

Find more information about the topic

FROM ONTOLOGY TO DATA

1. Create ontology of topics representing KET and SGC
   - From existing classifications, policy documents, expert users, and data
2. Automatically generate collections of keywords
   - NLP techniques (term extraction, word embeddings) from large training dataset
   - Ranking and scoring algorithms to decide:
     - Which topic(s) to match the keywords to?
     - Which are the best keywords?
     - Which are the best keyword combinations?



3. For each document, decide which topic best fits it
   - based on keywords and scoring algorithms

EXAMPLE OF PATENT ANNOTATION

Protein stabilized pharmacologically active agents, methods for the preparation thereof and methods for the use thereof

In accordance with the present invention, there are provided compositions and methods useful for the in vivo delivery of substantially water-insoluble pharmacologically active agents (such as the anti-cancer drug paclitaxel) in which the pharmacologically active agent is delivered in the form of suspended particles coated with protein (which acts as a stabilizing agent)…..

- RNA vaccines: (agent, protein, vaccine)
- anti-viral agents: (protein, anti-cancer, drug)
- protein vaccines: (protein, vaccine, antimicrobial)

KET: Industrial biotechnology
SGC: Health

ONGOING CHALLENGES

- Inconsistencies
  - ontology design has to be tailored to user needs, but these are not uniform
- Automation
  - keyword-based approach still requires some manual intervention for best results
- Accuracy
  - language processing is never 100% accurate
- Evaluation
  - how do we know if/when it's good enough?
  - Determination of weighting mechanisms; cut-off thresholds…
- The future?
  - integration of existing classification and modelling approaches with semantics

EUROSCIENCE OPEN FORUM

SHARING SCIENCE: TOWARDS NEW HORIZONS

9-14 JULY 2018

TOULOUSE, FRANCE

#ESOF2018 @ESOF_eu ESOF.eu