Towards an Infrastructure for Understanding and Interlinking Knowledge Co-Creation in European research

Diana Maynard, Adam Funk
University of Sheffield, UK

Benedetto Lepori
UPEM, France
Aims of the KNOWMAK project

• Develop a web-based tool providing interactive visualizations and indicators on knowledge co-creation in the European research area

• Structured around 3 integrative elements:
  – **Research topics**: relevant to Societal Grand Challenges (SGC) and Key Enabling Technologies (KET)
  – **Research Actors**: both “conventional” and social actors
  – **Geographical spaces**

• Combines 5 data sources:
  – Established: *publications, patents, projects*
  – New: *social innovation projects* and user attention based on *social media.*
# Topics: Societal Grand Challenges

<table>
<thead>
<tr>
<th>Category</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Health</td>
<td>Health, demographic change and wellbeing</td>
</tr>
<tr>
<td>Bioeconomy</td>
<td>Food security, sustainable agriculture and forestry, marine and maritime and inland water research, and the bio-economy</td>
</tr>
<tr>
<td>Energy</td>
<td>Secure, clean and efficient energy</td>
</tr>
<tr>
<td>Transport</td>
<td>Smart, green and integrated transport</td>
</tr>
<tr>
<td>Climate</td>
<td>Climate action, environment, resource efficiency and raw materials</td>
</tr>
<tr>
<td>Security</td>
<td>Secure societies - protecting freedom and security of Europe and its citizens</td>
</tr>
<tr>
<td>Society</td>
<td>Europe in a changing world - inclusive, innovative and reflective societies</td>
</tr>
</tbody>
</table>
### Topics: Key Emerging Technologies

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>IB</td>
<td>Industrial Biotechnology</td>
</tr>
<tr>
<td>NANO</td>
<td>Nanotechnologies</td>
</tr>
<tr>
<td>PHOT</td>
<td>Photonics</td>
</tr>
<tr>
<td>AMT</td>
<td>Advanced Manufacturing Technology</td>
</tr>
<tr>
<td>NME</td>
<td>Micro- and Nano-Electronics</td>
</tr>
<tr>
<td>AM</td>
<td>Advanced Materials</td>
</tr>
</tbody>
</table>

- Clearly these 6 KETs have overlap
- In particular, AMT is designed to be cross-cutting over the other 5
- This causes problems for ontology design (and topic assignment)
The datasets

• 3 existing data sources on knowledge production:
  – scientific publications derived from the Web of Science database (CWTS-WoS database)
  – patents derived from PATSTAT (UPEM-PATSTAT database)
  – European projects derived from CORDIS (AIT-EUPRO database)

• 2 new data sources:
  • social innovation projects
  • user attention based on social media.
Communities of attention around publications in Twitter
European Nanotechnology Clusters
The problem

• Increasing complexity, dynamicity and multi-disciplinarity of emerging scientific and technological research
• Knowledge production comes from multiple sources and crosses disciplines, institutional and geographical borders
• Mapping European state-of-the-art in research in key technologies is difficult
• Traditional indicators use rigid and coarse classification systems, and do not cater for new places of social innovation or new knowledge domains and technologies
• They are backward-looking and difficult to use for policy decisions
• Terms in different kinds of data vary widely – policy makers do not use the same language as patents or publications
The solution: ontologies

- Ontologies enable mapping between user queries, indicators and topics
- The ontologies give us a way to handle searching by topic
- The user can search by either topic(s) or by associated keyword(s) that can be linked with these topics
- Ontologies allow user exploration of knowledge around topics
- Enable creation of indicators around topics
- Act as a bridge between user queries and information in the databases
Topic-based Scenario

Query: new cancer therapy

System returns 5 different topics about cancer therapy
(new diagnostic techniques; new methods in drug delivery; nanotechnology in cancer, ...)

Choose a topic          Choose aggregated result (all topics)
Show result              Show related topics / more information
KNOWMAK Filter Search

This is a simple example of how a filtered search using the KNOWMAK ontology might look.

Match all drug cancer

Nanotechnology in cancer

Selected Class: http://www.gate.ac.uk/ns/ontologies/knowmak/nanotechnology_in_cancer

Cancer nanotechnology is a branch of nanotechnology concerned with the application of both nanomaterials (such as nanoparticles for tumour imaging or drug delivery) and nanotechnology approaches (such as nanoparticle-based theranostics) to the diagnosis and treatment of cancer. Nanotechnology in cancer.

Related Keywords: application, approach, branch, cancer, concerned, delivery, diagnosis, drug, imaging, nanomaterials, nanoparticle-based, nanoparticles, nanotechnology, such, theranostics, treatment, tumour
This is a simple example of how a faceted search using the KNOWMAK ontology might look.

### Key Emergent Technologies
- Advanced Manufacturing Technology
- Advanced Materials
- Biotechnology
- Micro and Nano electronics
- Nanoscience and technology
- Optics and photonics

### Societal Grand Challenges
- Bioeconomy
- Climate

### Applied Immunology
- Assay systems
- Biologics
- Biomaterials
- Biomimetics
- Cell delivery
- Environmental biotechnology
- Expression systems
- Functional genomics
- Gene delivery

### Antagonism and RNA sponge
- Antibody fragment therapy
- Antibody therapy
- Antisense oligonucleotide therapy
- Cell therapies
- DNA vaccines
- Gene therapy
- Locked nucleic acid
- Meganucleases
- Nucleic acid therapeutics

Selected Class: [http://www.gate.ac.uk/ns/ontologies/knowmak/dna_vaccines](http://www.gate.ac.uk/ns/ontologies/knowmak/dna_vaccines)

Dna vaccines. A DNA vaccine is a substance that is composed of deoxyribonucleic acid (DNA) and encodes antigens. After administration of the DNA, antigens are produced and stimulate an immune response. DNA vaccines.

Related Keywords: acid, administration, antigen, composed, deoxyribonucleic, dna, immune, produced, response, substance, vaccine
Ontologies connect information

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

Link related topics

Find more information about the topic
Topics can belong to multiple classes

We can now look at both **biomaterials** and **nanomedicine** to find related information
A possible user interface
Potential user queries

• What kinds of research topic does a region specialise in?
• Who are the main actors on a particular topic in a particular region?
• How are they connected?
• How diversified is a region’s knowledge base?
• What is the innovation performance of a region compared to other regions?
• How diversified is a region’s knowledge base?
Ontology Design

• The ontologies are built around the KET and SGC
• Topics are based on existing principled classifications
  • KET/SGC subclasses in policy documents
  • Nature.com ontology
• Linked to the primary data sources (patents, publications, projects)
  • Mappings to topics in patents/publications/projects
• Keywords associated with topics
  • created from a combination of policy documents and primary data sources
Annotating Data with Ontologies

• The data sources are annotated against the ontologies, i.e. each document is associated with one or more topics
• Sophisticated NLP matching of keywords in the documents (from titles, abstracts etc) with ontology
• Based on linguistic pre-processing, term recognition, frequency and some weighting mechanisms
• Multi-word terms are more important than single-word terms, e.g. “vapor deposition” is more useful than “vapor”
• These annotated data sources are then used to build indicator, e.g. for each topic, how many publications are there and in which region?
I propose to investigate a new research frontier on spin physics at the boundaries (surfaces) of materials with strong spin-orbit interaction (SOI). Although the properties of these materials have been studied for more than half a century, researchers are just starting to grasp the richness of SOI phenomena that occur at them. SOI leads to surface and boundary states with unusually large spin splitting in simple heavy elements. It can also produce a nontrivial topology in band insulators that brings about metallic surface states with exotic spin textures that are protected by time reversal symmetry.

I plan to use our cutting-edge expertise on all-electrical lateral spin injection and detection methods to unravel the spin dynamics in them, providing a wealth of information that could not be otherwise obtained. A comprehensive set of objectives will include material integration with ferromagnets and insulators, and innovative devices and measurement protocols.

- Optical data storage
- Biotechnology
- Advanced materials
Summary

• Project only started 6 months ago, so very early stages
• Ontologies and topics are the core of the system, but the hardest to develop
• Many problems with ontology population, annotation, ambiguity, and different use of language in different data sources
• Continuing development of ontologies and annotation methodologies
• Next phase will integrate elements from social media and new actor types
• High-risk but highly exciting!
More information

- Main project website
- Sheffield's KNOWMAK work
- RISIS project
- GATE tools
- KNOWMAK poster in the ESWC Poster Session
- Learn more about the project in the EU Project Networking Session

The KNOWMAK project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 726992.