

# FutureTDM

Reducing Barriers and Increasing Uptake of Text and Data Mining for Research Environments using a Collaborative Knowledge and Open Information Approach



## FutureTDM: Improving Uptake of Text and Data Mining in the EU

*Maria Eskevich, Centre for Language and Speech Technology (CLST),  
Radboud University Nijmegen, the Netherlands*

@LT-Accelerate 2016

21.11.2016



Horizon 2020

Coordination and Support Action

GARRI-3-2014 Scientific Information in the Digital Age: Text and Data Mining (TDM)

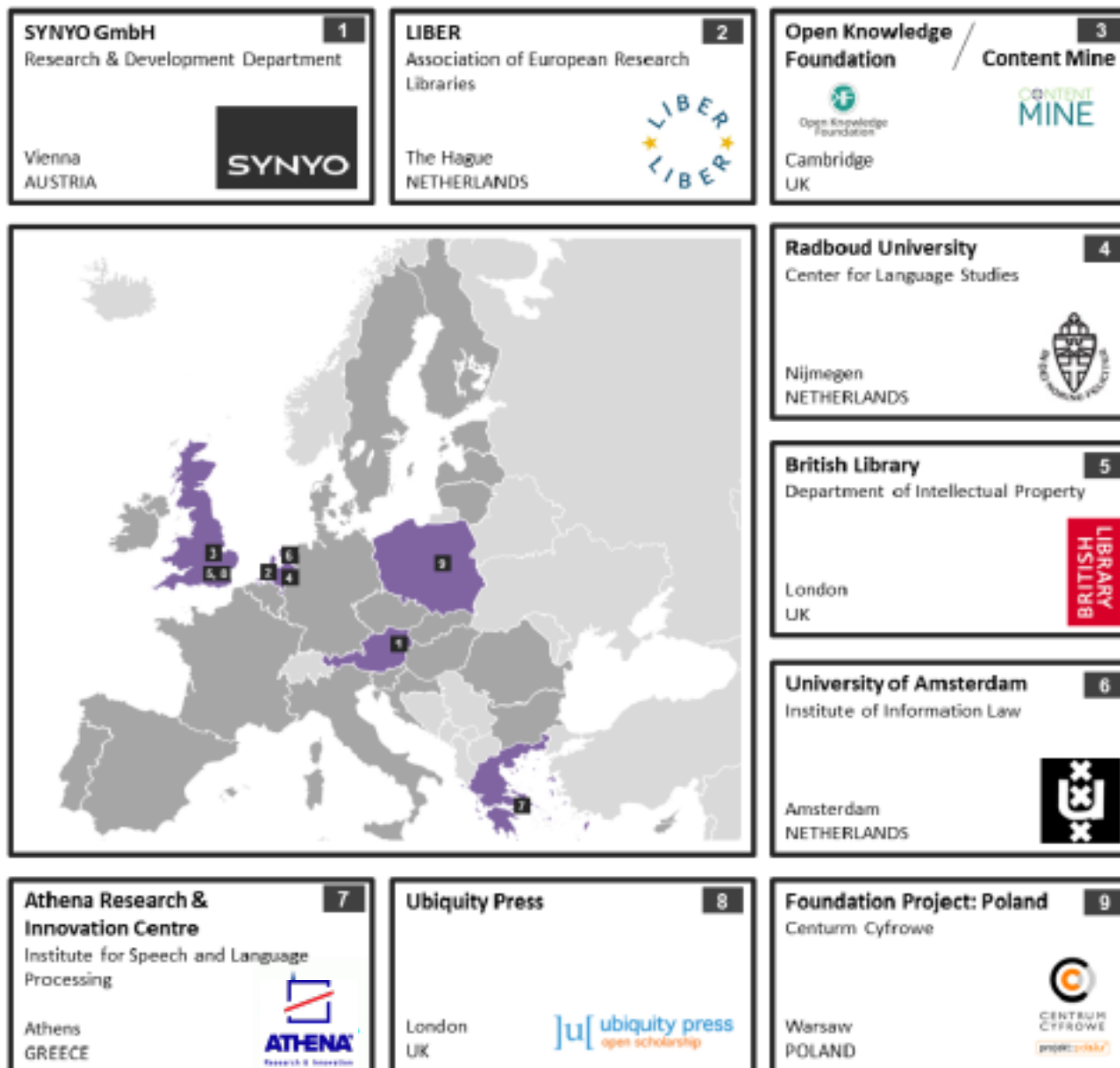
Project number: 665940



# Who are we?



# FutureTDM Consortium



## Focus

- **Policy developments and legal framework** in the EU and its Member States (including copyright exemptions).
- **Increasing awareness of TDM** in institutional settings and among stakeholders, as well as the skills, rewards and support involved in using this technology.

## Expected impact

- This action should **reduce barriers in order to increase the uptake of TDM** by researchers and innovative businesses as a tool in the co-creation of knowledge, which will result in wider and more efficient circulation of knowledge and **ultimately lead to a R&I policy that is more relevant for and responsive to society.**



# FutureTDM Objectives

**INVOLVE** all key stakeholders via targeted stakeholder consultation to identify practices, requirements, and specific challenges in the field of TDM and installation of an advisory board and integration of practitioner groups in local and pan-European workshops

**ANALYSE** current application areas and trends in TDM including statistics and key figures, collect relevant research and industrial projects, derive best practices and provide a structured collection of relevant knowledge and information resources

**BUILD** a Collaborative Knowledge Base and an Open Information Hub combined on a web-based platform including intuitive tools like a TDM Stakeholder Map, a TDM Expert Navigator, a European TDM Research Project Directory, or a TDM Best Practices Library



**ASSESS** existing studies, legal regulations and policies on TDM within the European Union considering generic and domain-specific barriers and national idiosyncrasies anomalies, and compare them with international developments

**ELABORATE** a legal and policy framework for future TDM, define policy priorities, specify a research agenda with future applications in mind to create interdisciplinary, case-driven practitioner guidelines to foster the spread of TDM in various research fields within the EU

**INCREASE** awareness of TDM to especially attract new target groups and science domains by creating a roadmap, run mobilisation and engagement activities and provide information material and modern TDM Visualisations and Infographics

# Text and Data Mining: Definition

- “**Discovery** by computer of new, **previously unknown** information, by automatically extracting and relating information from different (...) resources, to reveal otherwise hidden meanings” (Hearst, 1999), or
- “**Exploratory data analysis** that leads to the discovery of heretofore unknown information, or to answers for questions for which the answer is not currently known” (Hearst, 1999).



# TDM across all economic sectors



# TDM in Europe: Language perspective



## EU case is specific due to the additional language dimension

- Bias towards English language in terms of availability of tools/data/overall focus
- Multilinguality of the population requires additional costs on TDM solutions localisation

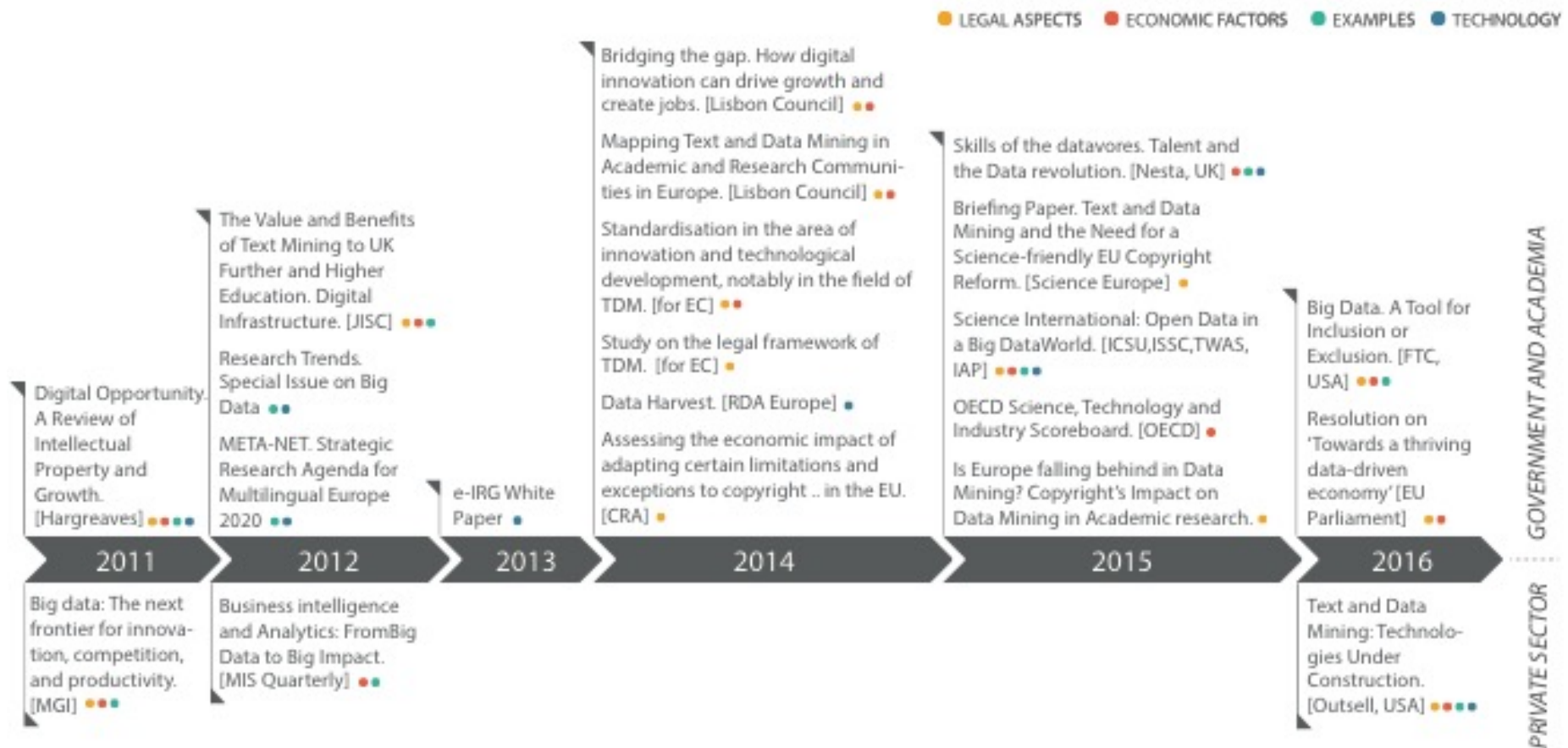
# **FutureTDM Analysis**

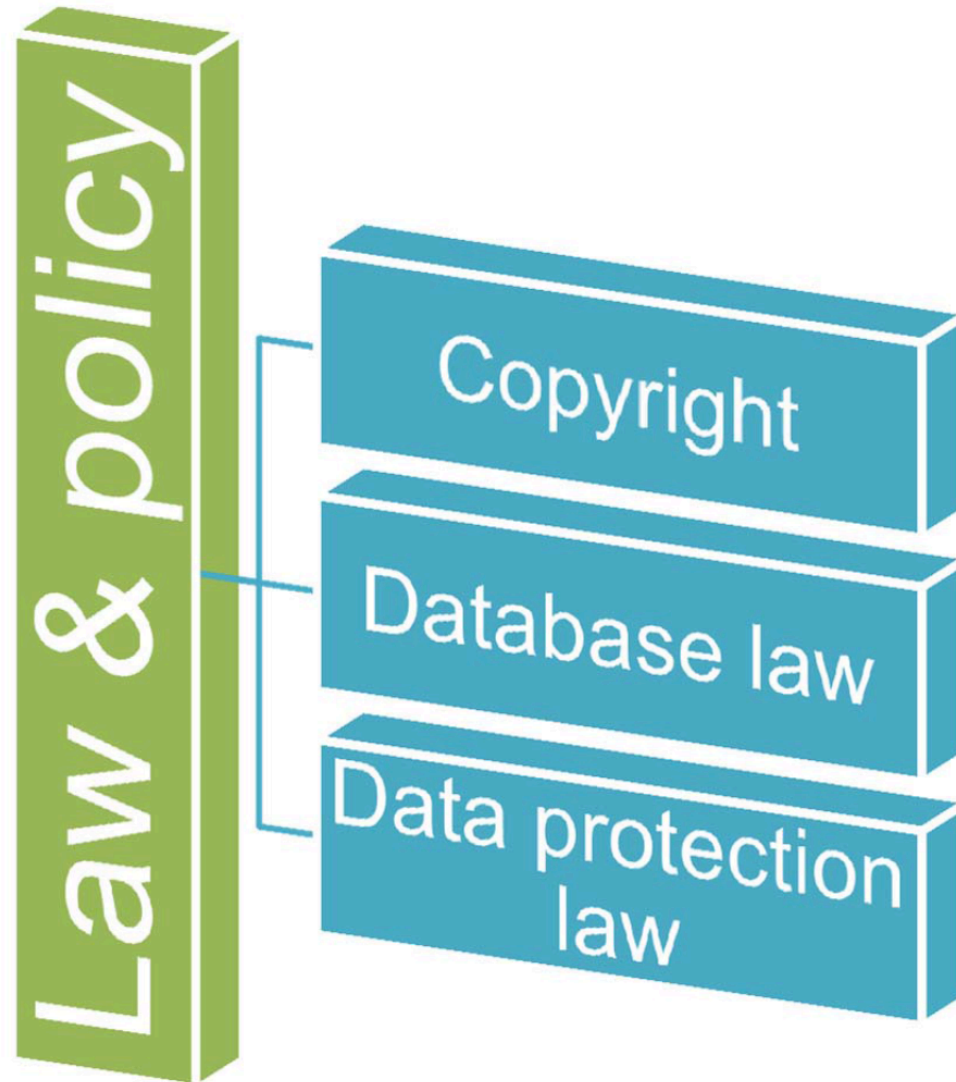




- Technology:
  - General Landscape of the TDM in Europe [[D3.1](#)]
  - European Landscape of TDM Applications [[D4.1](#)]
- Legal aspects:
  - Analysis of the legal barriers within the EU [[D3.3](#)]
- Collection of the stakeholders feedback [[D4.2](#)]
  - Knowledge cafes
  - Interviews

# Reports of TDM





## Benchmark: categories of legal barriers

**Restrictiveness:**  
Strict rules

**Fragmentation:**  
Different rules and  
interpretations

**Uncertainty:**  
Unclear rules

# Knowledge Cafés





# Insights from the practitioners

- Broad spectrum of opinions on existence of barriers and challenges for the TDM uptake

## Legal and content:

- Legal uncertainty:
  - Data protection regulation
  - Copyright exceptions
- Access barrier : API and restrictive licensing
- Use barrier:
  - Quality of data
  - Tools and infrastructure for TDM

## Technical and Infrastructure:

- Access/availability of data
- API use restrictions
- Formats
- Unstructured data
- Lack of tools
- Languages
- Lack of efficient standardisation

## Economy and Incentives:

- No single European market
- Not enough policing on non-European companies
- Investments for infrastructure and localization
- More funding needed for digitization
- Overall mismatch between academic and private sector funding

## Education and skill:

- Lack of knowledge
- Gap between industry needs and academic curriculum
- Lack of awareness on TDM
- Data Management plans
- Rights clearance

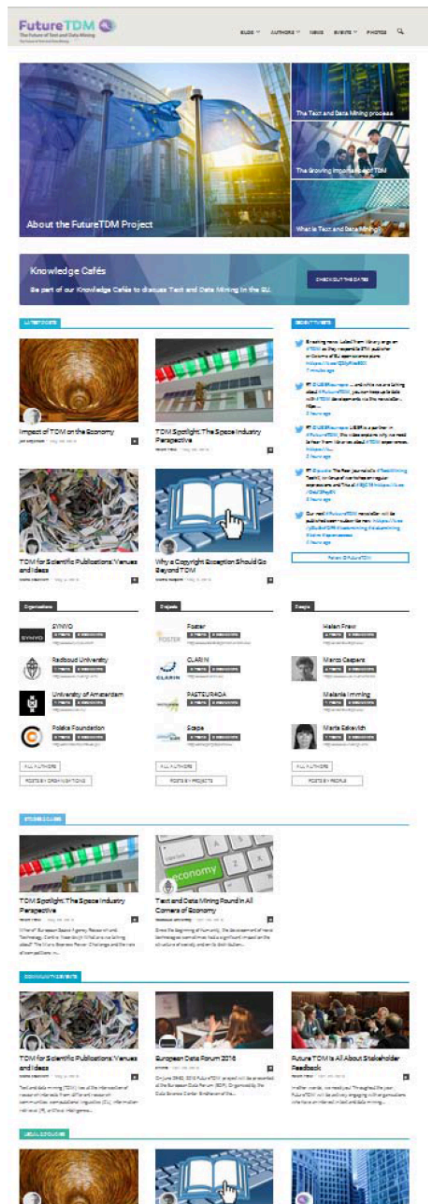


**FutureTDM Contribution**

- Dissemination to raise awareness and community engagement
  - FutureTDM Platform
  - Awareness sheets
  - FutureTDM newsletter
  - Collaboration with H2020 OpenMinTeD project
- Developing policy framework recommendations
  - TDM Value chain
  - Stakeholders categories



# Dissemination: FutureTDM platform



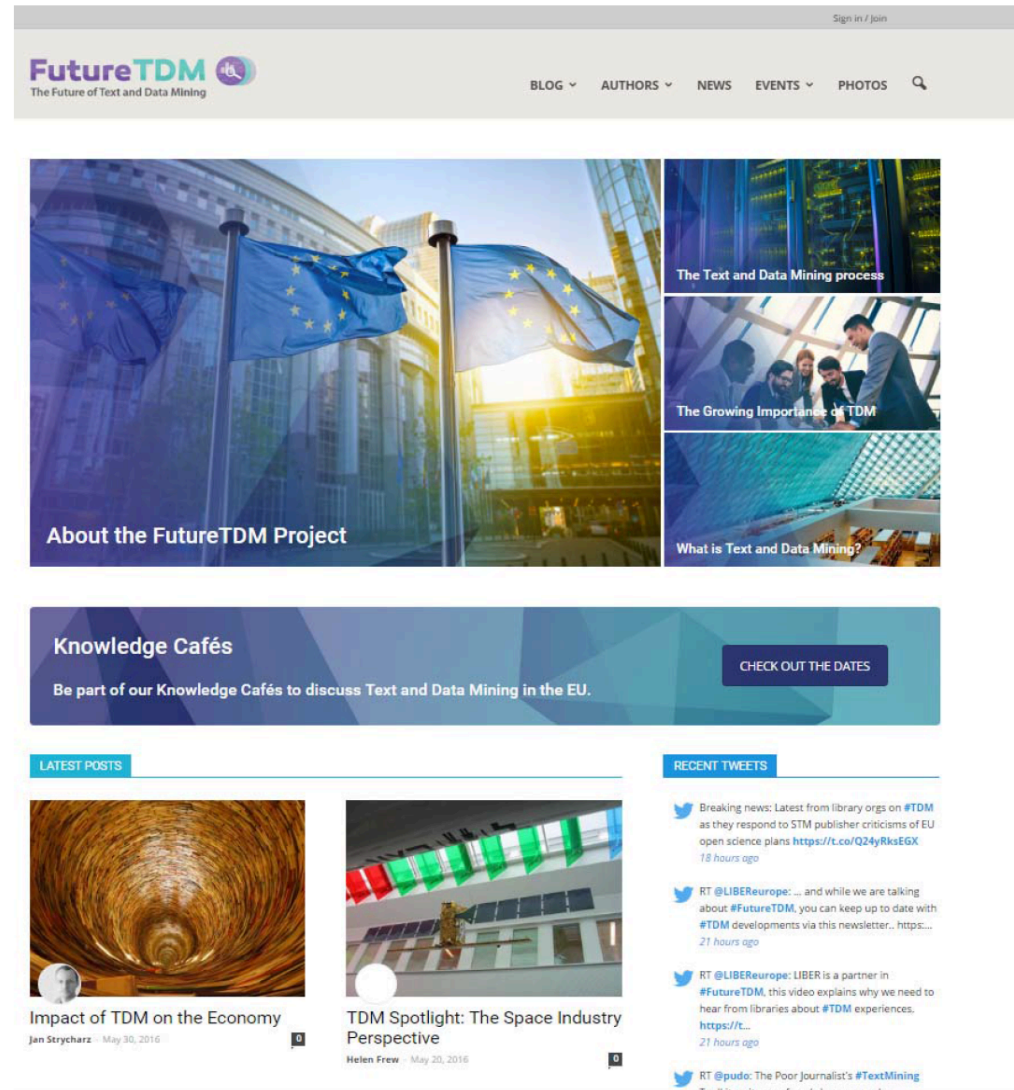
Main Posts

Knowledge Cafés

Latest Posts and Tweets

Authors according to Organisations, Project, People

Blog posts divided into six categories



# Dissemination: Blogging

All

Community & Events

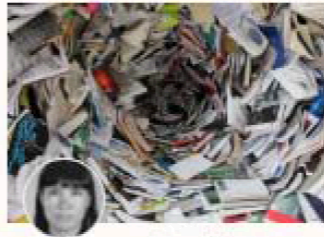
Education & Skills

Legal & Policies

Research & Insights

Stories & Cases

Technologies & Tools



TDM for Scientific Publications:  
Venues and Ideas



European Data Forum 2016



Future TDM is All About  
Stakeholder Feedback



The FutureTDM Open Science  
Cafe at the Dutch Presidency  
Open Science...

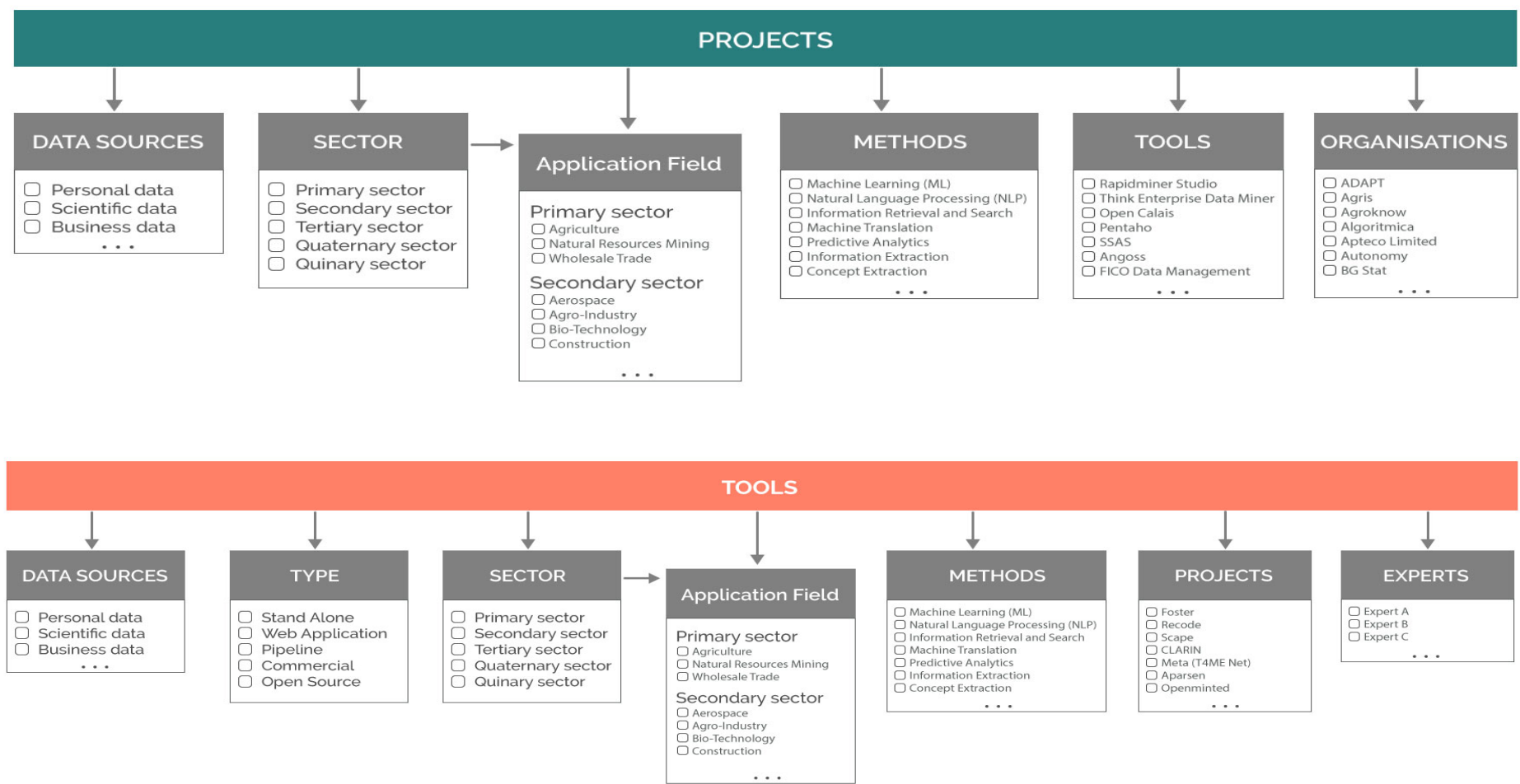


## Categories

- ☐ Community & Events
- ☐ Education & Skills
- ☐ Legal & Policy
- ☐ Research & Insights
- ☐ Stories & Cases
- ☐ Tools & Techniques



## Elaborated structure for TDM representation in the platform



# Dissemination: Awareness sheets



## TDM value chain:

- Content creation
- Content dissemination
- Text and data mining
- Knowledge utilisation

## Directions for recommendations

- Legal rules and policies (IP law and data protection law)
- Education, skills, and awareness (knowledge access)
- Economy and incentives (investment issues, full scale use of TDM potential)
- Technical and infrastructure (data&tools, language aspect, standards)



## COMING SOON

- FutureTDM Workshops in 2017
- We are present at:
  - Open Government Partnership Summit (December 2016, Paris, France)
  - RDA Plenary Meeting (April 2017, Barcelona, Spain)
  - LIBER Annual Conference (July 2017, Patras, Greece)
- Direct interviews with stakeholders & Online survey



WEB PLATFORM: [www.futuretdm.eu](http://www.futuretdm.eu)



TWITTER CHANNEL: [www.twitter.com/futuretdm](http://www.twitter.com/futuretdm)