European Commission

Tracking of Research Results

Tracking of Research Results (TRR) to support R&I Policy Making

Semantic Data Linking from Multiple Heterogeneous Sources

Intelcomp

A COMPETITIVE INTELLIGENCE CLOUD/HPC PLATFORM FOR AI-BASED STI POLICY MAKING (GRANT AGREEMENT NUMBER 101004870)

Opening event

27 April 2021

Dr. Sergio DI VIRGILIO, DSc, MEng
Data & Knowledge Management
DG RTD – GH6

European Commission

Context


Projects
Publications
Clinical Trials
Web sites
Awards
Proposals
Deliverables
Patents
Guidelines
News
Policy Documents
1-4 years
5-10 years
11-15 years
>15 years
Ontologies
National Data bases
Proprietary Data bases
Professional Data bases
Altmetrics
LOD
Tracking of Research Results
European Commission

Tracking of data beyond the duration of the projects for policy making

Output data:
need to be accurate even when monitoring data are not available (ind. 1-4 + 14)

Performance of researchers and organizations:
need to be tracked post-implementation (ind. 5-10)

Insights:
can/could be derived about the impact (ind. 11-13)

Issue

Because Policy priorities changes

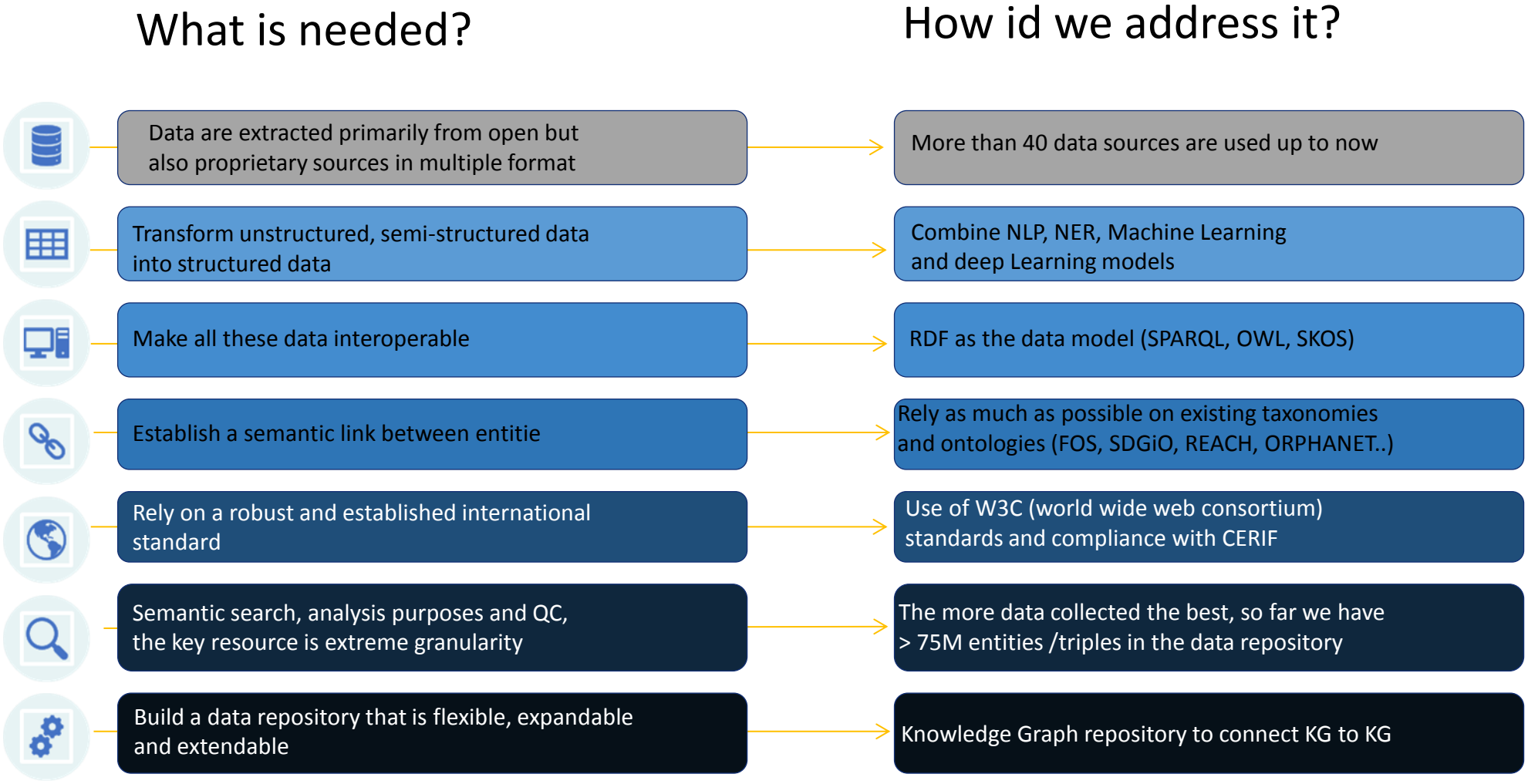
Future data needs are NOT known

New type of evidence need to be produced

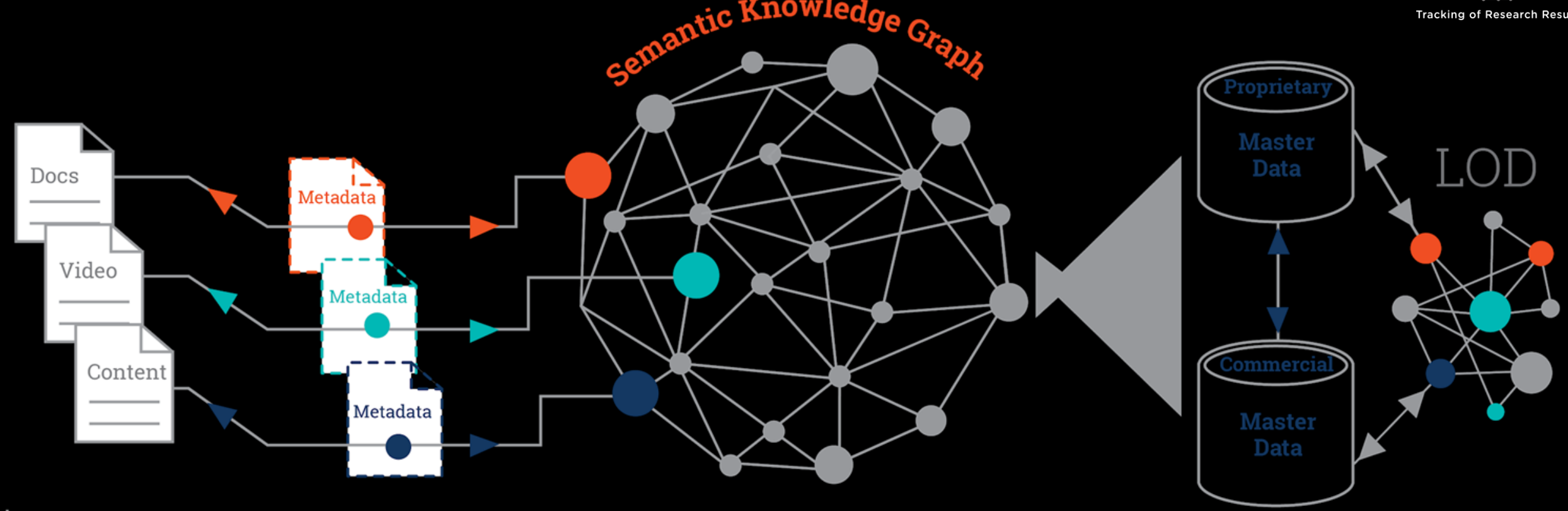
Solution

TRR use a bottom up approach on highly disaggregated (to the lowest possible denominator) data that can be later aggregated for the preferred indicator

Methodology



Tracking of Research Results



Build a Knowledge Graph, Integrate Data

Visualization is not considered as an output but an input for output we have charts and tables

The problems of scale are compounded by other challenges such as the breadth of topics covered, their jargon specific to each field and the changes in meanings of phrases over time.

European Commission

Ask less, ask anything (bar) TRR philosophy: a few key points

Unique coverage of data sources, with an aim to link them through specific entities
All key entities are tracked: researchers, organisations, projects

Coverage of all key stages of the R&I lifecycle
Tracking FP entities across time, incl. beyond the end of funding -> all output is relevant

Capture quasi-FP innovation sphere
Not just FP data -> better understand & benchmark

First projects to track data to medium- and long-term economic and societal/health impacts
Link previous project activities to events that happened recently

Tracking event/news data -> who, what, where in my vicinity doing what?
New indicators and line of thinking

1. Outputs, products and interventions

2. Collaborations

3. Scientific publications

4. Intellectual property rights

14. Scientific prizes

5. Innovation

6. Dissemination activities

7. Further funding/investment

8. Next destinations

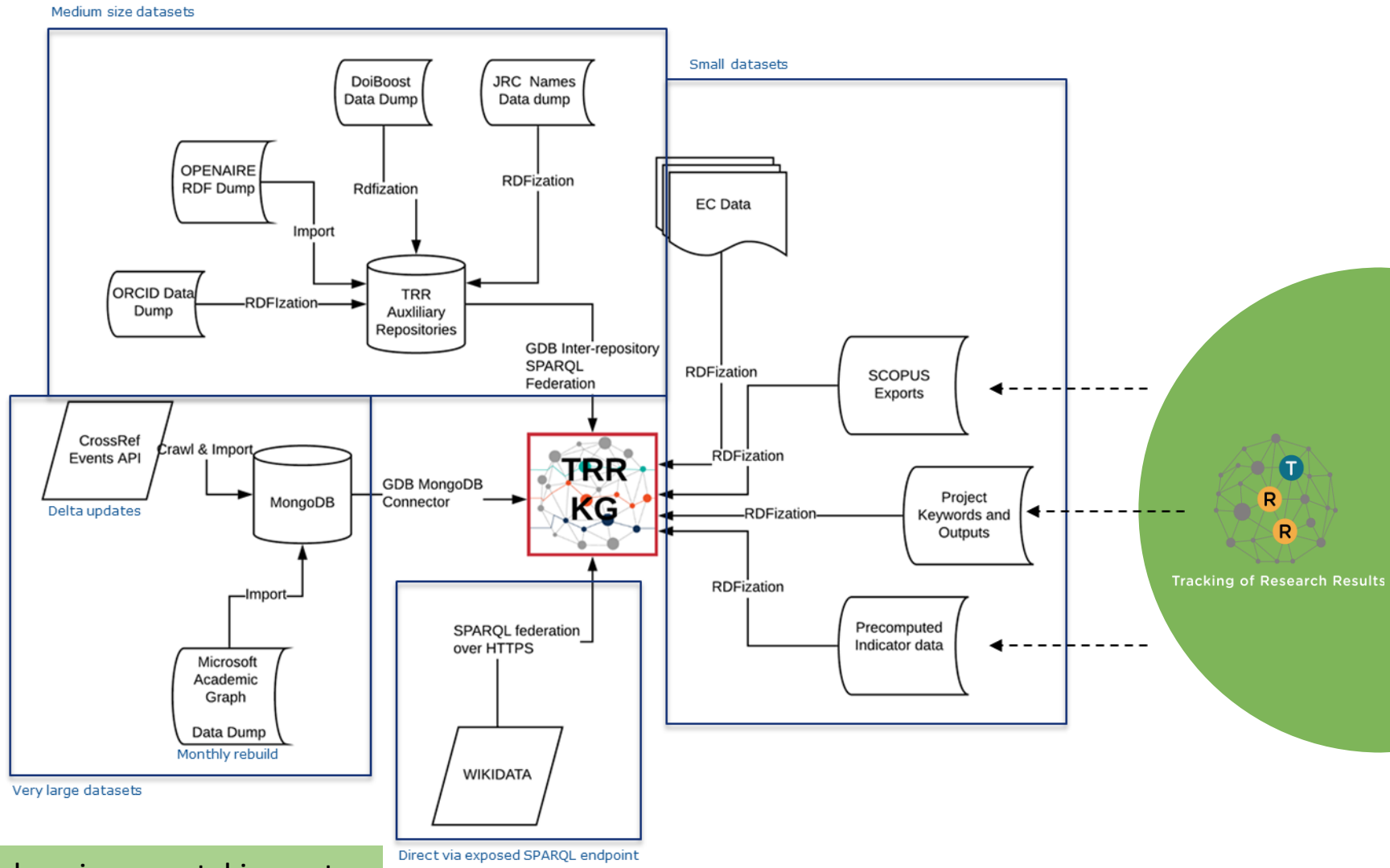
9. Effects on the company /private sector

10. New companies created

11. Impact on health and welfare/health and environmental impacts

12. Impacts on creativity, culture & society/social, economic, capability and cultural impact

13. Influence on policy making/political impact



European Commission

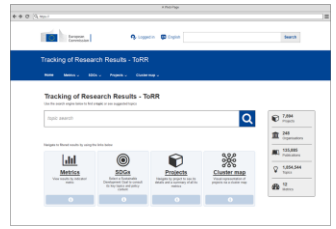
One multi-faceted, interlinked data infrastructure for all use cases

Tracking of Research Results

Because the underlying data is harmonised and curated in a Knowledge Graph repository , it is also *flexible and expandable* in how it can be used, and what it can be used for.
R&I: bottom-up and multidisciplinary
Missions oriented policies will be top-down and multidisciplinary

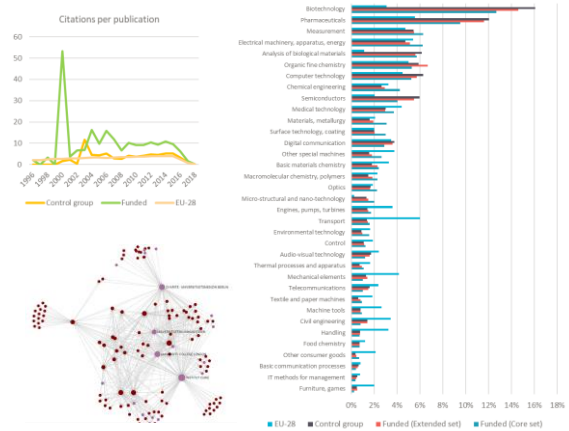
Discover

« Ask anything » semantic queries



Analyse and Report

Metrics and Indicators



Direct Access to Data

Bulk downloads

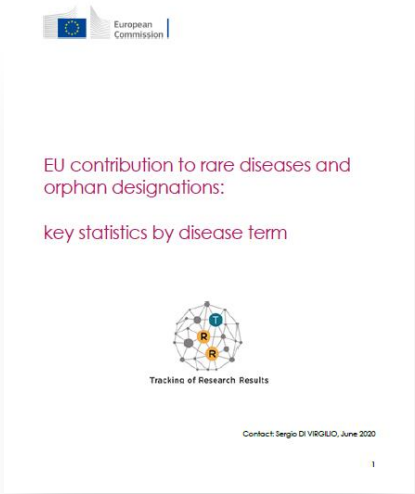
GraphDB

ArangoDB

jupyter

META DATA			
Project ID	200234	**Acronym**	CRUMBS IN SIGHT
Funding Scheme	FP7 HEALTH	**EU Contribution**	€ 3.0 mil.
Number of Participants	6	**Participant Organizations**	KNAW, AMT, RUMC, MPG, CNRS, USFD, EKUT
Insights extracted from project reports			
Project Outputs	**200234_mouse** : Crb1 knockout mice , Crb1 Crb2 double knock out mice , conditional knockdown mice , Mpp3 conditional knockout cKO mice , Mpp3 cKO mice , Crb1 / Crb2F/ Chx10Cre/ mice, conditional Crb2 knockout mice , double knockout mice **200234_vector** : pharmaceutical CRB gene therapy vector, Gene therapy vector, CRB1 Gene Therapy Vector, CRB gene therapy vector, AAV hCRB1 gene therapy vectors, AAV2/6 hCRB1 clinical gene therapy vector, gene therapy vectors and Müller glia progenitor cell therapy, clinical AAV hCRB1 gene therapy vector, clinical AAV2/6 hCRB1 gene therapy vectors **200234_mutant** : Crb1 mutants **200234_serotype** : AAV serotype **200234_platform** : baculovirus production platform for the AAV1 serotype		
Thematic Key-words / Phrases	Müller-glia-cells, CRB, photoreceptor, cell, retina, protein, gene-therapy, adheren-junction, eye-disease, retinal-degeneration, knock-out-mouse, eye, membrane		
Named Entities / IPR	AAV, AAV1, AAV2/6, CRB, CRB1, CRB2, CRB2F, hCRB, hCRB1, Mpp3, Crb1 / Crb2F/ Chx10Cre		
Additional Attributes			
Fields of Study / Themes	Biology, Cell-Biology, Anatomy, Genetics, Retina, Molecular-Biology, Retinal-Degeneration		
Relevant SDGs	SDG 3 – Good Health and Wellbeing		

European Commission



Overview of drug legislation for rare diseases in international jurisdictions

Region/ country	Criteria for defining rare disease, population (%)[8,9]	No. of designated orphan drugs (effective date)	No. of drugs approved* (effective date)	Legislation (supervising body)[10]	Elements of drug legislation for rare diseases[9,10]				
					Tax incentives	Fast-tracking of drug evaluation	Marketing exclusivity (yr)	Technical assistance to obtain approval	Other
Australia	≤2000 (0.01)	180 (2010)[11]	62 (2010)[11]	Australian Orphan Drugs Program (1997) (Therapeutic Goods Administration)	No	Yes	Yes (5)	Yes	Applications reconsidered every 12 months
European Union	5/10 000† (0.05)	664 (2010)[12]	51 (2010)[12]	Regulation No. 141/2000 (2000) (European Medicines Evaluation Agency)	Yes	Yes	Yes (10)	Yes	Not applicable
Japan	≤50 000 (0.04)	167 (2004)[10]	95 (2004)[10]	Orphan Drug Regulation (1993) (Ministry of Health, Labour and Welfare)	Yes	Yes	Yes (10)	Yes	Partial reimbursement of development costs Extended registration validity period
United States	≤200 000 (0.07)	2194 (2010)[13]	350 (2010)[13]	Orphan Drug Act (1983) (Food and Drug Administration)	Yes	Yes	Yes (7)	Yes	Not applicable

*This may also be defined as the number of orphan drug marketing authorizations.
†Measure of incidence.

Rare diseases

- Any disease affecting fewer than 5 people in 10,000 in the EU is considered rare. Although this might appear small, it translates into approximately 246,000 people. Most patients suffer from even rarer diseases affecting 1 person in 100,000 or more. Approximately 5,000-8,000 distinct rare diseases affect 6-8% of the EU population i.e. between 27 and 36 million people.

Maybe not so rare

Data Sources

Structured Thematic databases (Health) :

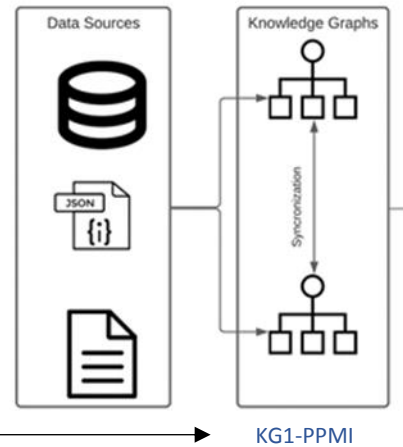
- Clinical Trials;
- Medicines (orphan drugs and human-medicinal products);
- Health Technology Assessments;
- Clinical Guidelines;
- Cochrane Reviews;
- Database of Abstracts of Reviews of Effects (DARE);
- NHS Economic Evaluation database (NHS EED);
- Chemicals (ECHA)

KG2-ONTOTEXT

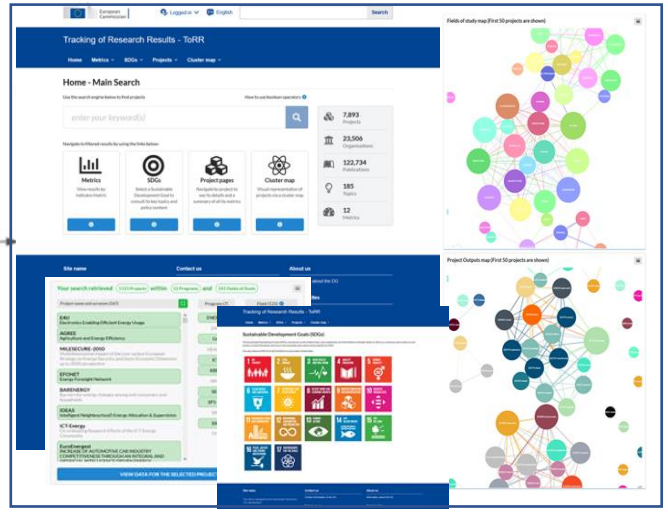
- CORDIS data;
- OpenAire;
- Microsoft Academic;
- Scopus;
- CrossRef Events;
- Wikidata;
- European Media MonitorThrough JRCNames links;
- Lens.org In the process of integration;
- Consortium output;

KG1-PPMI

- CORDA/SESSAM – documents and monitoring data;
- Company website data – additional data on companies and their products;
- MAG – publications, researchers, FOS;
- PATSTAT – patents ;
- EUIPO & TM-Link – Trademarks ;
- Web scraped EC data – policy documents (EU Publications Office; EC; EP repositories);
- Web scraped Media data (based on EMM sources) – media mentions;



ToRR



European Commission

Health Case study

Rare Diseases in EU research programmes

Tracking of Research Results

Topic clusters and FET/KET technology	Interdisciplinarity (field of study network)	Cross-cutting programmes analysis (FP7/H2020/3rd EU Health 2014-2020)	Contributin to diseases and EMA Orphan Designations
EU support to the development of EMA Orphan Designation in FP7 projects	EU support to clinical trials addressing Rare Diseases	Overview of outputs produced in the FP projects portfolio	Overview of innovations produced by participating organisations in the FP7 projects portfolio
Overview of citations in policy documents of FP7 projects portfolio	Tracking of research results after/beyond EU funding	Today Scientific breakthroughs linked to FP7 projects	What's happening on the market?

Contribution to diseases & EMA Orphan designations

Top-25 most frequently addressed rare diseases in the analysed set of FP7 rare diseases projects, by number of projects

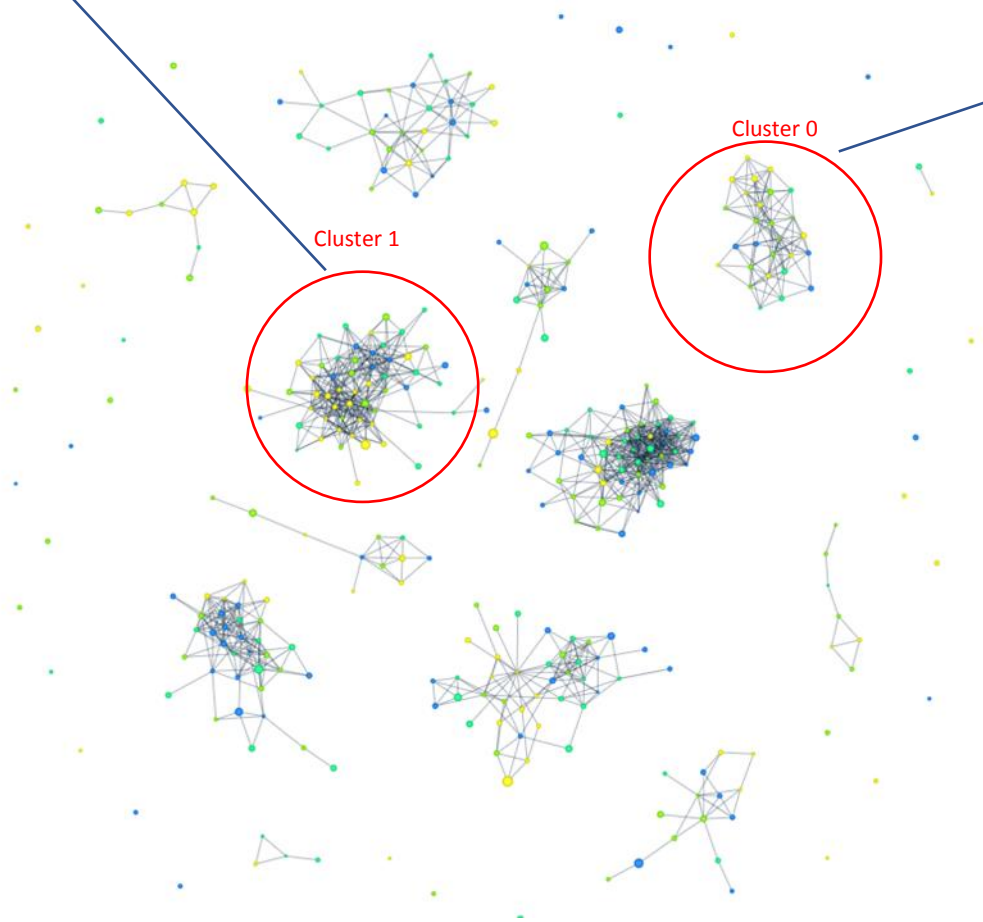
DISEASE	TOTAL EC CONTRIBUTION, EUR MILLION	TOTAL PROJECT COSTS, EUR MILLION	NUMBER OF PROJECTS
malaria	422,7	542,5	139
tuberculosis	436,6	582,5	98
pancreatic cancer	361,2	483,0	57
cystic fibrosis	220,0	290,9	48
glioma	304,8	419,9	37
Pseudomonas aeruginosa	235,0	323,0	33
sepsis	238,7	328,0	32
hepatocellular carcinoma	156,7	209,1	32
leishmaniasis	142,1	191,2	31
avian influenza	97,5	129,1	28
Duchenne muscular dystrophy	216,1	317,2	28
amyotrophic lateral sclerosis	124,6	180,8	25
multiple myeloma	132,2	175,8	25
Becker muscular dystrophy	131,0	176,3	23
severe combined immunodeficiency, adenosine deaminase deficiency, SCID, ADA	153,4	205,8	20
ovarian cancer	109,4	146,0	20
dengue	97,5	134,2	18
hepatocellular carcinoma, pexastimogene devacirepvec	128,6	173,2	18
Hearing loss, acute acoustic trauma, sudden deafness, surgery induced acoustic trauma	67,8	91,4	16
retinitis pigmentosa	36,2	46,5	16
non-small cell lung cancer anaplastic lymphoma kinase (ALK)-positive, NSCLC	103,0	162,2	16
small cell lung cancer	75,5	108,8	14
graft-versus-host disease	125,3	168,1	14
pre-eclampsia	57,7	75,2	14

Top-25 most frequently addressed EMA orphan designations in FP7 projects, by number of projects

ORPHAN DESIGNATION	RELATED DISEASES	NUMBER OF PROJECTS
Ciprofloxacin	cystic fibrosis	18
Artesunate	malaria	16
Dexamethasone	multiple myeloma	14
everolimus	gastric cancer, renal-cell carcinoma, tuberous sclerosis	9
Nitric oxide	cystic fibrosis	8
Sirolimus	beta thalassaemia intermedia, chronic non-infectious uveitis, sickle cell disease	8
Ribavirin	Crimean Congo haemorrhagic fever, Lassa fever	7
Complement factor H	haemolytic uraemic syndrome	7
Doxorubicin	hepatoblastoma	6
Mercaptopurine	acute lymphoblastic leukaemia	5
Itraconazole	invasive aspergillosis	5
6-mercaptopurine monohydrate	acute lymphoblastic leukaemia, B-cell acute lymphoblastic leukaemia	5
Glucagon	congenital hyperinsulinism	4
mifamurtide	hepatocellular carcinoma	4
Vorinostat	multiple myeloma	3
blinatumomab	acute lymphoblastic leukaemia, B-cell acute lymphoblastic leukaemia	3
Gastrin 17C diphtheria toxoid conjugate	pancreatic cancer	3
Givinostat	Becker muscular dystrophy, Duchenne muscular dystrophy	3
Chelidonii radix special liquid extract	pancreatic cancer	3
Deferiprone	sickle cell disease	2
Recombinant human minibody against complement component C5 fused with RGD-motif	ischaemia injury associated with solid organ transplantation, reperfusion injury associated with solid organ transplantation	2
Panobinostat	Hodgkin lymphoma, multiple myeloma	2
Avian polyclonal IgY antibody against Pseudomonas aeruginosa	cystic fibrosis	2
Sapacitabine	acute myeloid leukaemia, myelodysplastic syndromes	2
Adeno-associated viral vector	acute intermittent porphyria	2

Bottom-up methodology that not only cuts across programmes,
But clusters them to reveal new dimensions

cluster	fos_top30 per cluster	projects
1	University hospital, Peripheral blood, Disease progression, After treatment, Mrna expression, Positive control, Expression data, Mechanism of action, In patient, Expression analysis, Sample collection, Operating procedures, Protein expression, Clinical study, Cell transplantation, Immune system, Animal studies, Plasma samples, Inflammatory response, Signalling pathways, Exome sequencing, Cell based, Cell activation, Protein level, Genetic variants, Tumor cells, Clinical grade, Serum samples, Data management, Functional studies	BIOSHARE-EU, INTREALL, EUROSARC, BESTCILIA, RD-CONNECT, InSPiRe, METFIZZ, EEC, ASTERIX, AHEAD III, PERFORM, CHRONIOUS, ESI-TBVI, TBSUSGENT, ENCE-CF-LAM-LTX, EUCO-NET, LOULLA&PHILLA, SAGHE, TB PAN-NET, ENRIECO, EPIWORK, CHICOS, TRANSEURO, DALI, CHANCES, NIDIAG, EUROMEDICAT, THE HIP TRIAL, IMPACTT, EUROGENTEST2, COHEMI, PHARMAS, COACH, TIRCON, E-PREDICE, TAIN, INTERPREGGEN, CONTRAST, NANOMAL, DEVELOPAKURE, MEUSIX, IMPROvED, AfriCoLeish, SKIP-NMD, DSD-LIFE, DESSCIPHER, CHILD-EU, ODAK, COSMIC, ASPRE, PHAGOBURN, PANCARELIFE, NEOVANC, EUGENMED, CULPRIT-SHOCK, LENA, FEMNAT-CD, DESIRE, SYMPATH



cluster	fos_top30 per cluster	projects
0	World health, Health organization, Health sector, Health services, Health professionals, Health related, Climate change, Disease control, Human health, Primary care, Policy development, Country level, Public policy, Socio economic impact, Health problems, Resource allocation, Disaster risk reduction, Science policy, Millennium Development Goals, Priority setting, Pharmaceutical industry, Early detection, Community level, Research evidence, Public awareness, Research policy, Emergency situations, Emergency planning, Sub saharan, Burden of disease	CUREHLH, EUNEFRON, LEUKOTREAT, FIGHT-MG, PEMPHIGUS, HOMITB, ADAMANT, EURIPFNET, STAR-T REK, OPTISTEM, NIMBL, EPC-TM-NET, CURELUNG, CANCERALIA, LUNGTARGET, INTRICATE, CELL-PID, IMMOMEC, OPTATIO, OVER-MYR, OCTIPS, REGENER-AR, TRIAD, PathCO, TUMADOR, IACT, CAM-PAC

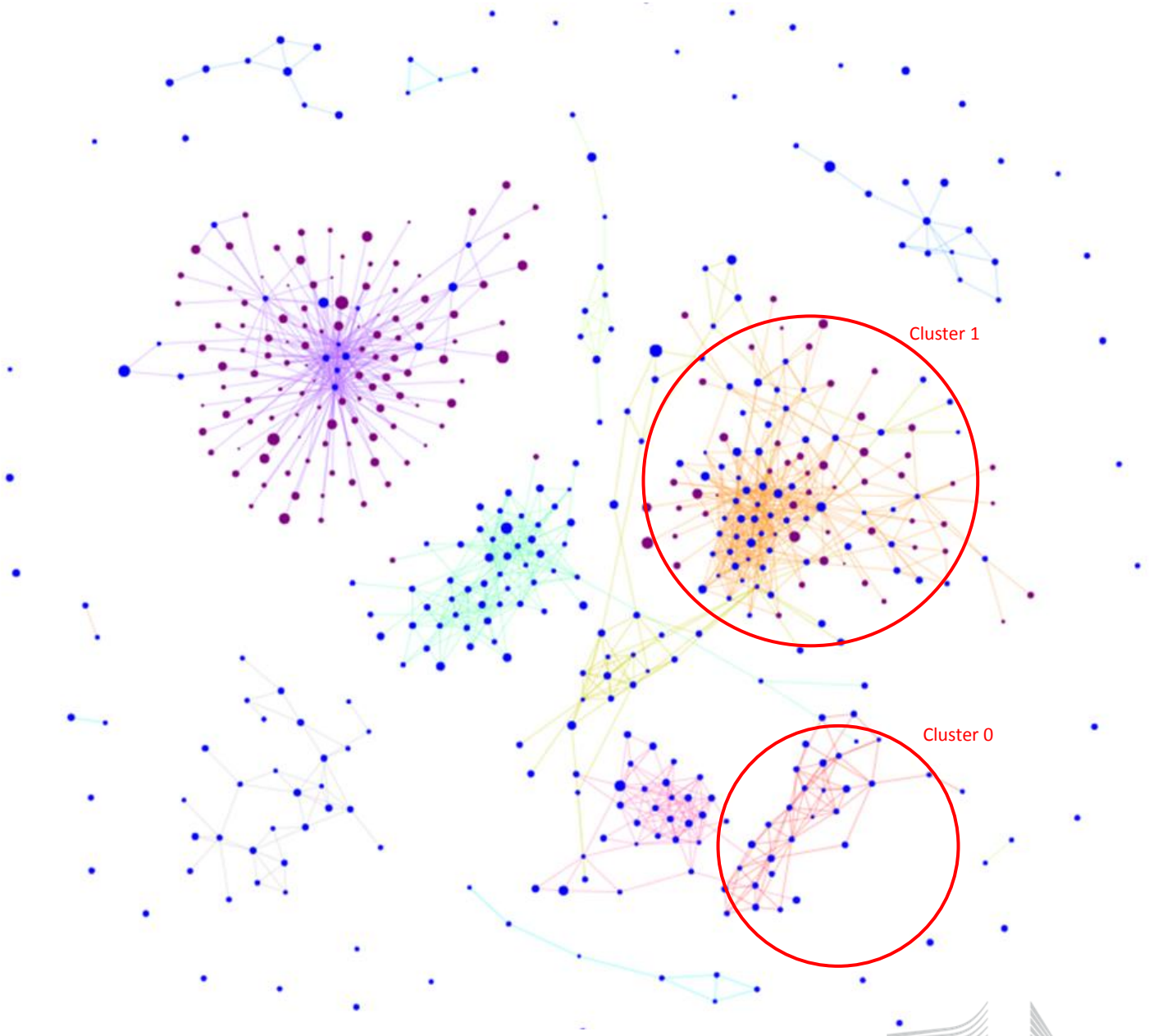
348 FP7 Rare diseases related projects in 10 clusters with prevalence scoring for FET and KET topics

European Commission

Comparison to other programmes

FP7 Cooperation Programme RTD

EU Health 2014-2020 Programme CHAFEA

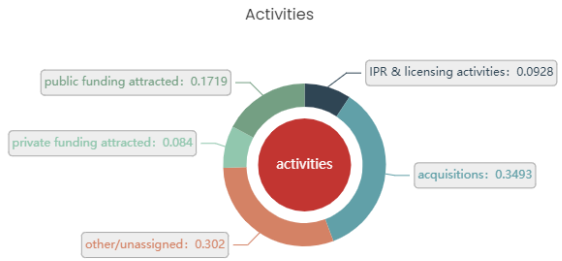


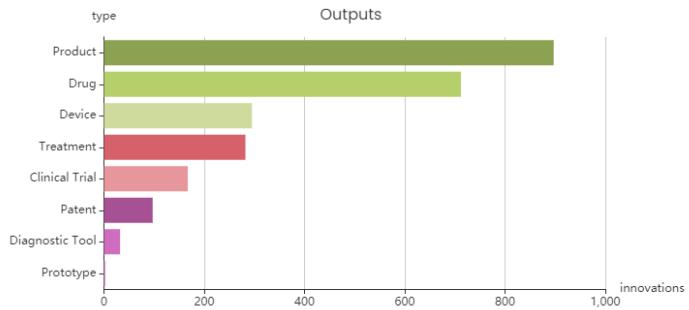
European Commission

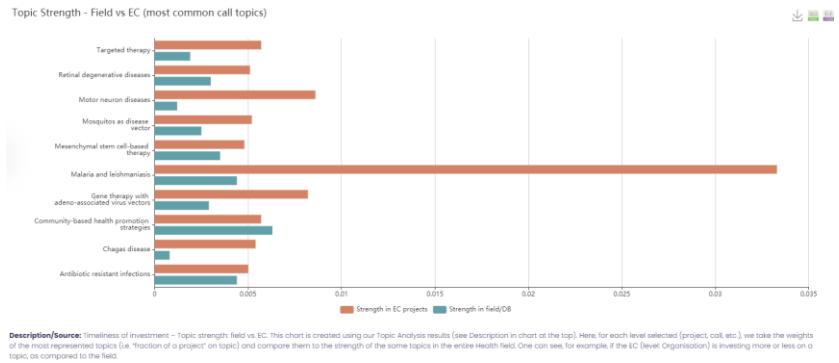
Indicators for portfolio_rare_diseases_RTD_FP7 (Call)

Input | Throughput/Output | Academic Impact | Economic Impact | Societal Impact

Innovations by Participant Companies

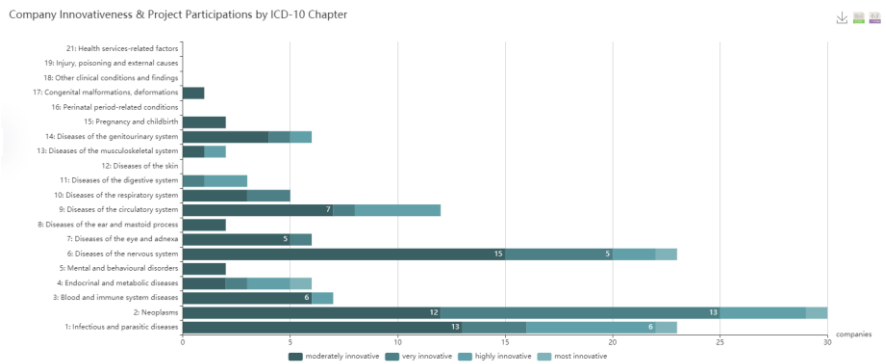






Description/Source: Timeliness of investment – Topic strength: field vs. EC. This chart is created using our Topic Analysis results (see Description in chart at the top). Here, for each level selected (project, call, etc.), we take the weights of the most represented topics (i.e. "fraction of a project" on topic) and compare them to the strength of the same topics in the entire Health field. One can see, for example, if the EC (level: Organisation) is investing more or less on a topic, as compared to the field.

Description/Source: Innovations by participant companies. For each project, we take the participants that are private-for-profit organizations (companies) and extract their current innovation activities and outputs from their websites. One can examine, how innovative prior project participants are today. For calls, programmes, organisations and the entire Health field, we aggregate the data from all corresponding projects. For topics and categories, we aggregate the data from all projects that are at least 30% on that topic or category. Each company is counted exactly once.



Description/Source: Company innovativeness and project participations by ICD class. We have classified each project to one or more ICD-10 chapters (https://icd.who.int/browse10/2019/en). Each chapter matched, receives all the private-for-profit participants (companies) of a project. The total length of the bars represents the number of companies involved in projects in that ICD chapter. Each bar is split into segments of companies of a particular innovativeness level. A company's innovativeness rank (among other participants) is determined by analyzing the text on its website. For calls, programmes, organisations and the entire Health field, we aggregate the data from all corresponding projects. In this chart, companies can be double-counted.

Description/Source: Clinical trials linked to projects per ICD-10 chapter. We have classified each project to one or more ICD-10 chapters (https://icd.who.int/browse10/2019/en). Each chapter matched, receives all the clinical trials linked to a project. For calls, programmes, organisations and the entire Health field, we aggregate the data from all corresponding projects. Clinical trials are linked to a project using a set of criteria (match in substance, participants, timeline, etc.). On the left, we list our findings for a general match ("looser" set of criteria) and on the right, for a narrow (tight) match to projects.

SOME RESULTS

Looking at outcome and impact of Rare Diseases Research Projects

EMA rare diseases and orphan designations analysis

Total number of EMA rare diseases and orphan designations analysed: 671 diseases

Total number of FP7 projects contributing to EMA orphan designations: 765 projects addressing 209 diseases (i.e. 31% of the 671 analysed rare diseases were found to be addressed by FP7 projects)

Total EU contribution allocated to the identified projects: EUR 3.36 billion

EU support to the development of EMA orphan designations

Estimated number of EMA orphan designations addressed in FP7 projects: 61 designations

EU support to clinical trials addressing rare diseases

Identified number of clinical trials linked to the analysed FP7 rare diseases projects :

104 clinical trials using 50 medicines/active substances

Overview of outputs produced in the analyzed portfolio of FP7 projects

Total number of researchers linked to the projects analysed: 34,225 researchers

Total number of publications produced: 25,838 publications

Total number of patents produced: 236 patents

Total number of innovation outputs produced: 8,139 outputs

Overview of citations of FP research in policy documents for the analysed portfolio of FP7 projects

Total number of FP7 publications cited in EC and EP documents:

107 publications cited 206 times in 84 EP and EC policy documents

Overview of innovations produced by the participating firms in the analysed portfolio of FP7 projects

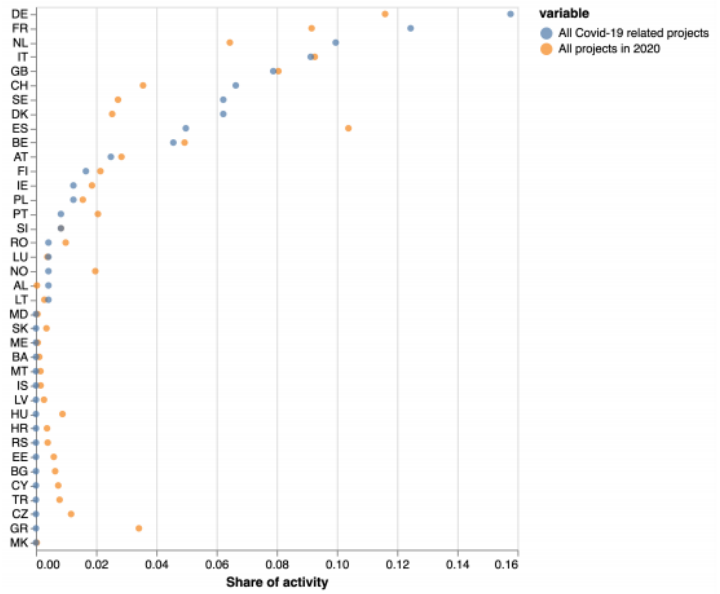
Total number of firms that participated in the projects analysed: 1060 firms

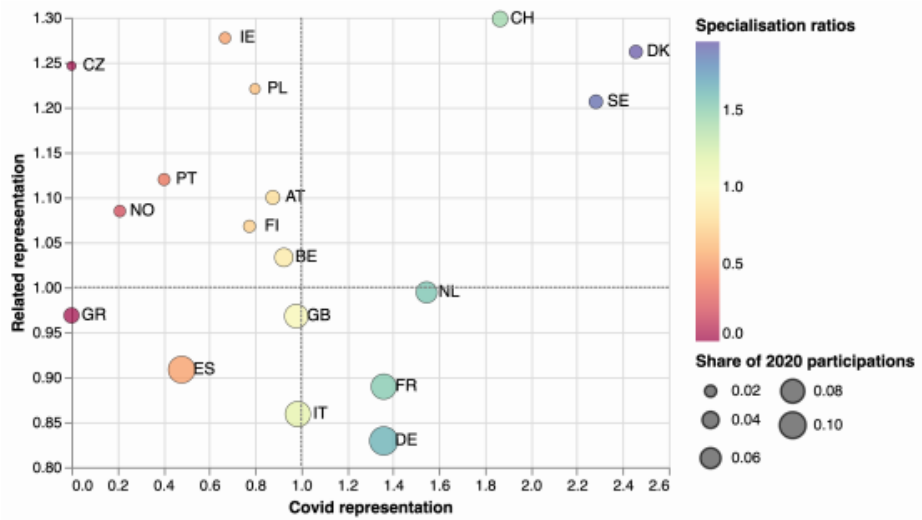
Total number of innovation announcements identified: 9605 innovation announcements, including 1427 announcements linked to new products/innovations on the market

Total number of announcements of additional funding attracted: 107 announcements

Of which: 55 announcements identified as public funding and 21 announcements identified as private funding

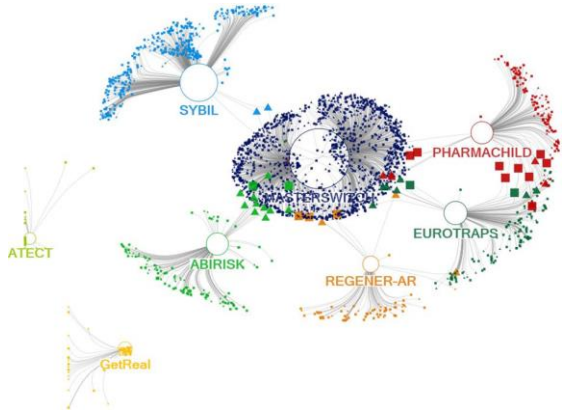
The COVID-19 Story





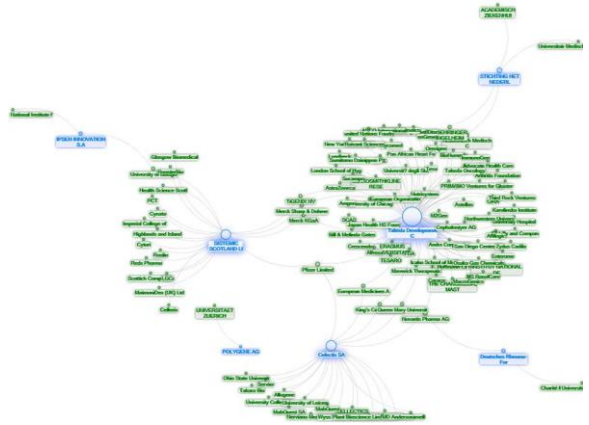
Tocilizumab

An immunosuppresive drug, mainly for the treatment of rheumatoid arthritis but today evaluated in patients admitted to hospital with COVID-19 (RECOVERY)



Gatekeepers

COLLABORATION NETWORKS OF RESEARCHERS LINKED TO PROJECTS RELATED TO TOCILIZUMAB



COLLABORATION NETWORKS OF COMPANIES INVOLVED IN PROJECTS ADRESSING TOCILIZUMAB

The SDGs Story

1. FP74SDG

Sustainable Development Goal(s) assigned to the projects by FP7-4-SD project (https://www.fp7-4-sd.eu/);

manually assigned by experts. reference data for the validation of the different predictions

2. ML Monitoring Data Main SDGs

Main Sustainable Development Goal(s) assigned by TRR using Natural Language Processing and Machine Learning;

Models were trained using the Deliverables and Monitoring Data for FP7 Projects;

3. ML Monitoring Data SDGs Full Dist.

Full probability distribution for all 17 SDGs from the Natural Language Processing /Machine Learning approach;

4. ML Monitoring Data & OECD Main SDGs

Main Sustainable Development Goal(s) assigned by TRR using Natural Language Processing and Machine Learning;

Models were trained using the Deliverables and Monitoring Data for FP7 Projects and OECD documents from SDG Path Finder (https://sdg-pathfinder.org/#/

5. ML Monitoring Data & OECD SDGs Full Dist.

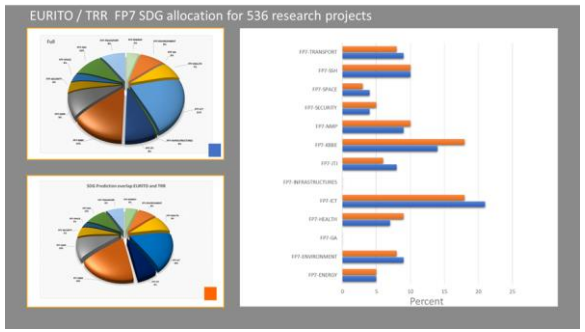
Full probability distribution for all 17 SDGs from the Natural Language Processing /Machine Learning approach;

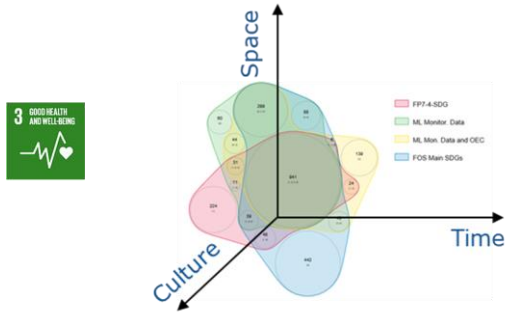
6. FOS Main SDGs

Main Sustainable Development Goal(s) assigned by TRR using the keyword mapping between Microsoft Academic Fields of Study (FOS) and SDGs;

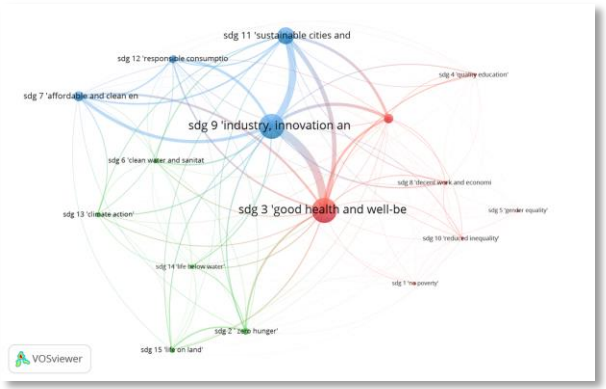
7. FOS SDGs Full Dist

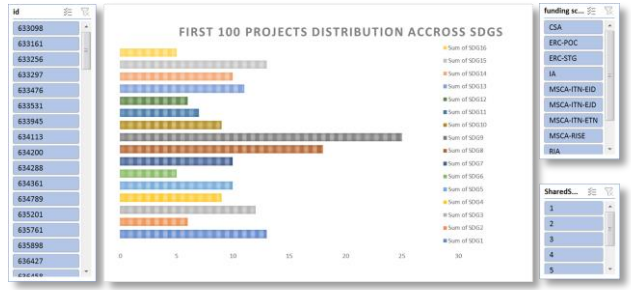
Full probability distribution for all 17 SDGs from the keyword mapping between Microsoft Academic Fields of Study (FOS) and SDGs;

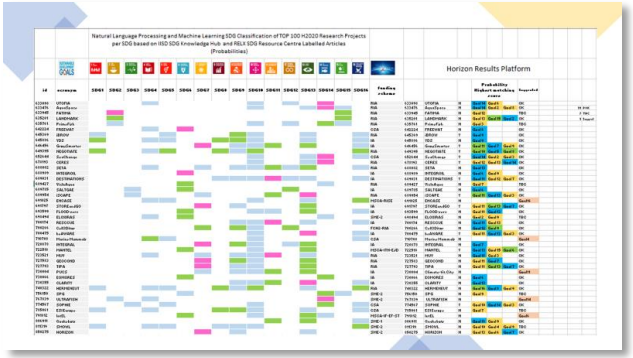


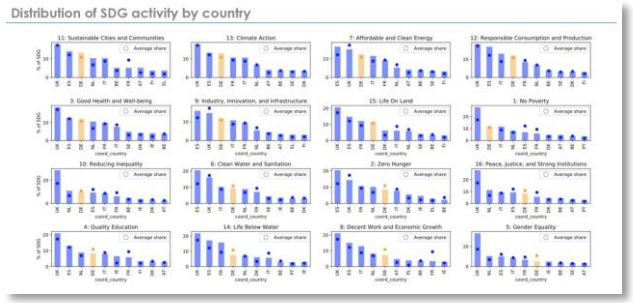


Mapping of links between SDGs based on FP7 project data



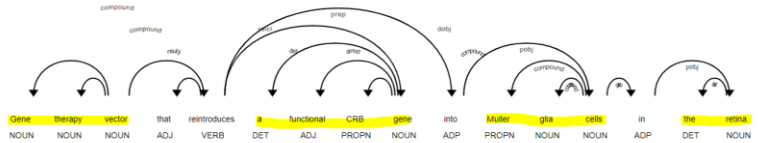






Recognition of innovation outputs in project reports

NOUN CHUNKS



EXTRACTED INSIGHTS

PROJECT ID	OUTPUT ID	OUTPUT NAMES FOUND IN THE REPORT	PROPER NOUNS/IPR
200234 (Crumbs in Sight)	200234_vector	pharmaceutical CRB gene therapy vector Gene therapy vector -> object mentioned in the sentence above CRB1 Gene Therapy Vector -> most frequently mentioned term CRB gene therapy vector most suitable CRB gene therapy vector AAV hCRB1 gene therapy vectors AAV2/6 hCRB1 clinical gene therapy vector production gene therapy vectors and Müller glia progenitor cell therapy clinical AAV hCRB1 gene therapy vector suitable AAV hCRB1 gene therapy vector AAV hCRB1 gene therapy vector AAV hCRB1 gene therapy vectors clinical AAV2/6 hCRB1 gene therapy vectors	CRB CRB1 AAV hCRB1 AAV2/6 AAV6 Crb1/Crb2 uniQure AAV-based

AUGMENTED Project Portfolio

META DATA			
Project ID	200234	**Acronym**	CRUMBS IN SIGHT
Funding Scheme	FP7 HEALTH	**EU Contribution**	€ 3.0 mil.
Number of Participants	6	**Participant Organizations**	KNAW, AMT, RUMC, MPG, CNRS, USFD, EKUT
Insights extracted from project reports			
Project Outputs	**200234_mouse** : Crb2 knockout mice , Crb1 Crb2 double knock out mice , conditional knockdown mice , Mpp3 conditional knockout cKO mice , Mpp3 cKO mice , Crb1 / Crb2F/ Chx10Cre/ mice, conditional Crb2 knockout mice , double knockout mice **200234_vector** : pharmaceutical CRB gene therapy vector, Gene therapy vector, CRB1 Gene Therapy Vector, CRB gene therapy vector, AAV hCRB1 gene therapy vectors, AAV2/6 hCRB1 clinical gene therapy vector, gene therapy vectors and Müller glia progenitor cell therapy, clinical AAV hCRB1 gene therapy vector, clinical AAV2/6 hCRB1 gene therapy vectors **200234_mutant** : Crb1 mutants **200234_serotype** : AAV serotype **200234_platform** : baculovirus production platform for the AAV1 serotype		
Thematic Key-words / Phrases	Müller-glia-cells, CRB, photoreceptor, cell, retina, protein, gene-therapy, adheren-junction, eye-disease, retinal-degeneration, knock-out-mouse, eye, membrane		
Named Entities / IPR	AAV, AAV1, AAV2/6, CRB, CRB1, CRB2, CRB2F, hCRB, hCRB1, Mpp3, Crb1 / Crb2F/ Chx10Cre		
Additional Attributes			
Fields of Study / Themes	Biology, Cell-Biology, Anatomy, Genetics, Retina, Molecular-Biology, Retinal-Degeneration		
Relevant SDGs	SDG 3 – Good Health and Wellbeing		

Researcher Portfolio

	Researcher
Name, Surname:	Jan Wijnholds
Affiliation: Name and PIC	KNAW (999518362)
Link to profile:	https://academic.microsoft.com/author/2008286538
Project:	CRUMBS IN SIGHT (200234);
Affiliated to beneficiary PIC?	Yes
Number of publications	15
Mentions in deliverables	236 mentions, of which 55 mentions in text Mentioned in the DoW
Author and project FOS	Author FOS: biology; retina; retinal; retinitis pigmentosa; CRB1; anatomy; molecular biology; cell biology; retinal degeneration Project FOS: Photoreceptor cell , Retinal degeneration , CRB1 , Gene therapy of the human retina , Outer nuclear layer

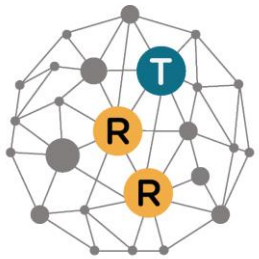
project_id	output_id	output_forms	output_fos_matched_w/names
200234	200234_mouse	['Crb1-Crb2 double knock-out mice', 'Mpp3 conditional knockout cKO mice']	{'crb1': 2, 'progressive retinal degeneration': 1, 'optical coherence tomography': 1, 'leber congenital amaurosis': 1, 'toluidine blue staining': 1, 'cross-sectional imaging': 1, 'vascular alterations': 1, 'retinitis pigmentosa': 1, 'functional analysis': 1, 'retinal lamination': 1, 'adherens junction': 1, 'retinal function': 1, 'double knockout': 1, 'spectral domain': 1, 'knockout mouse': 1, 'cross breeding': 1, 'cell division': 1, 'mouse strain': 1, 'early onset': 1}
200234	200234_serotype	['AAV serotype']	{'molecular evolution': 9, 'work package': 3, 'multidisciplinary approach': 1, 'degenerative disease': 1, 'production strategy': 1, 'recombinant virus': 1, 'lpl deficiency': 1, 'clinical grade': 1, 'cell specific': 1, 'start codon': 1, 'viral gene': 1, 'high titer': 1, 'phase i ii': 1, 'ex vivo': 1, 'crb1': 1}
200234	200234_vector	['AAV2/6 hCRB1 gene therapy vectors', 'gene therapy vectors and Müller glia progenitor cell therapy', 'AAV2/6 hCRB1 clinical gene therapy vector production', 'AAV hCRB1 gene therapy vectors', 'CRB1 gene therapy vectors generated in WPS', 'final CRB gene therapy vector', 'AAV hCRB1 gene therapy vector']	{'muller glia': 21, 'work package': 11, 'crb1': 7, 'retinal degeneration': 5, 'adeno-associated virus': 4, 'molecular evolution': 4, 'retinal pigment epithelium': 3, 'retinitis pigmentosa': 3, 'photoreceptor cell': 3, 'progenitor cell': 3, 'leber congenital amaurosis': 2, 'transmembrane protein': 2, 'cell transplantation': 2, 'therapeutic strategy': 2, 'single application': 2, 'double knockout': 2, 'polarized cell': 2, 'clinical grade': 2, 'high homology': 2, 'stem cell': 2, 'membrane-associated guanylate kinase': 1, 'ophthalmological equipment': 1, 'technology implementation': 1, 'drosophila melanogaster': 1, 'pharmaceutical industry': 1, 'manufacturing services': 1, 'functional importance': 1, 'intracellular domain': 1, 'membrane associated': 1, 'heterogeneous group': 1, 'future application': 1, 'post translational': 1, 'rod photoreceptors': 1, 'gene silencing': 1, 'gene transfer': 1, 'cell polarity': 1, 'visual system': 1, 'cell adhesion': 1, 'model system': 1, 'visual field': 1, 'cell therapy': 1, 'viral vector': 1, 'fund raising': 1, 'viral gene': 1, 'phase i ii': 1}

REITER

DATA for impact

EURITO
RELEVANT, INCLUSIVE, TRUSTED AND OPEN INDICATORS FOR RESEARCH AND INNOVATION POLICY

STIPCOMPASS
INTERNATIONAL DATABASE ON STI POLICIES



Tracking of Research Results

KNOWMAK
KNOWledge in the MAKing in European Society

Broader context

THANKS

Wider context: DG RTD has launched several big data projects lately

- Tracking of research results in FP7
- H2020 grants (Data4Impact, EURITO, KNOWMAK, REITER, RISE, etc.)
- Big data pilots contributing to the EIS

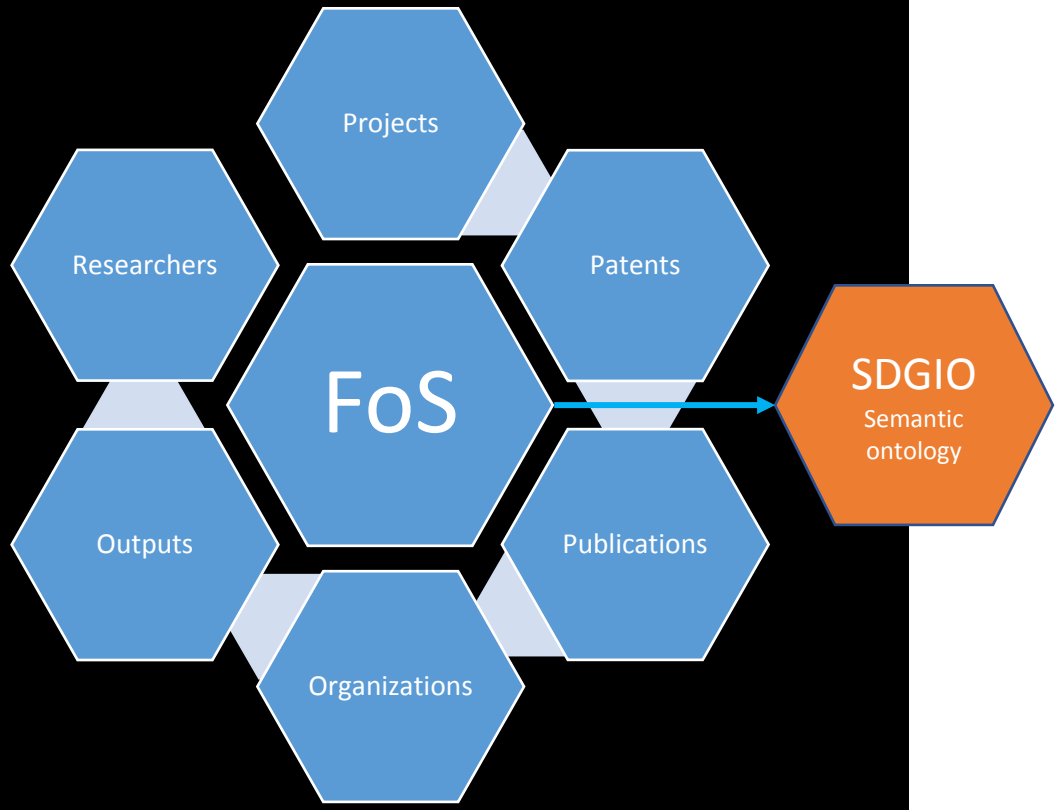
Key properties of Big Data:

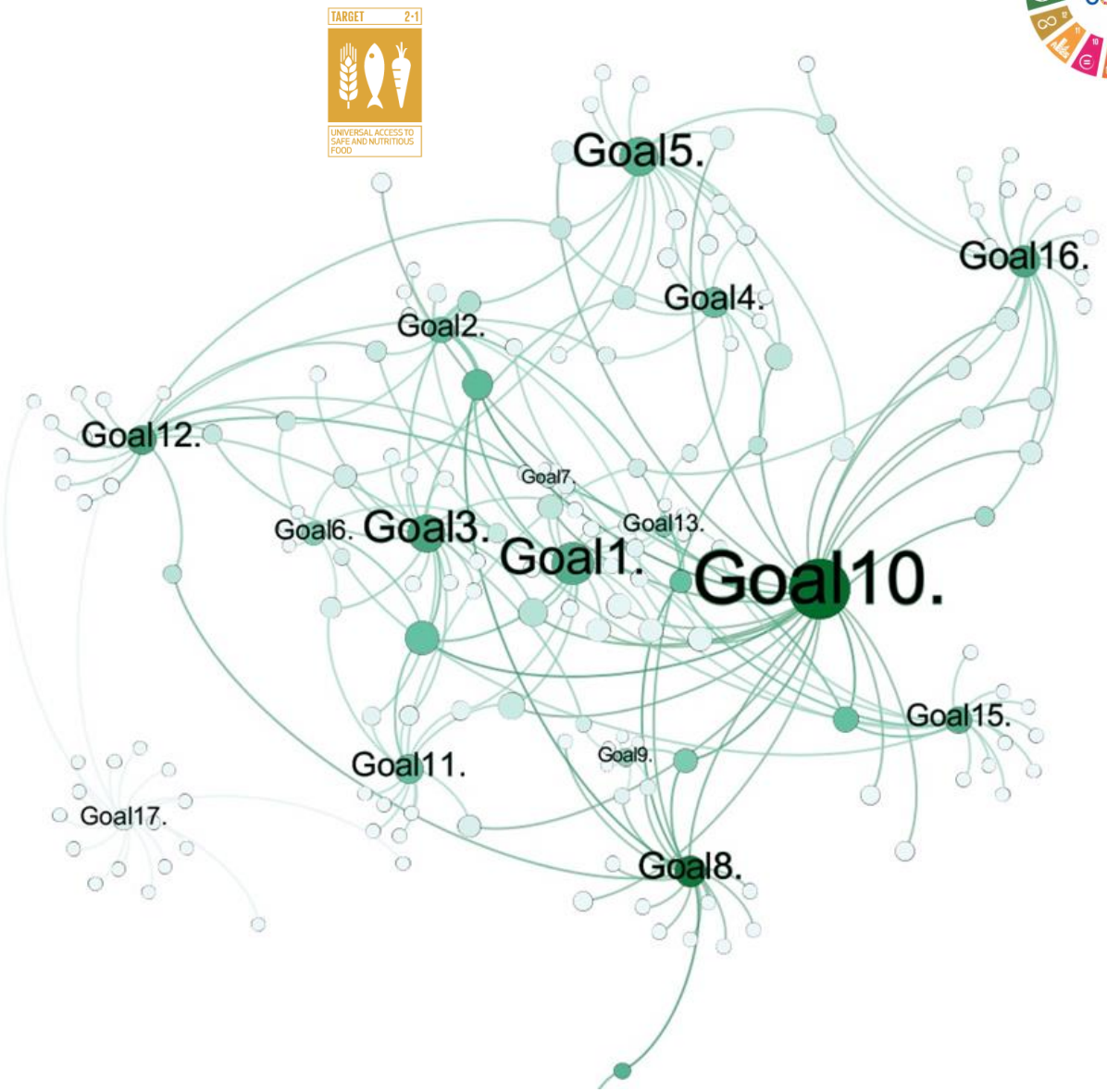
- Volume, i.e., no sampling is generally applied
- Variety, i.e., structured and unstructured data from various sources, in different formats
- Velocity, i.e., real-time/rapid data
- Veracity, i.e., variations in data quality, cleaning, processing, etc.

Non-intrusiveness → Big Data is a byproduct of digital interaction and communication

Key objective: make Big Data small!

Data linking through TRR Ontology-based text classification



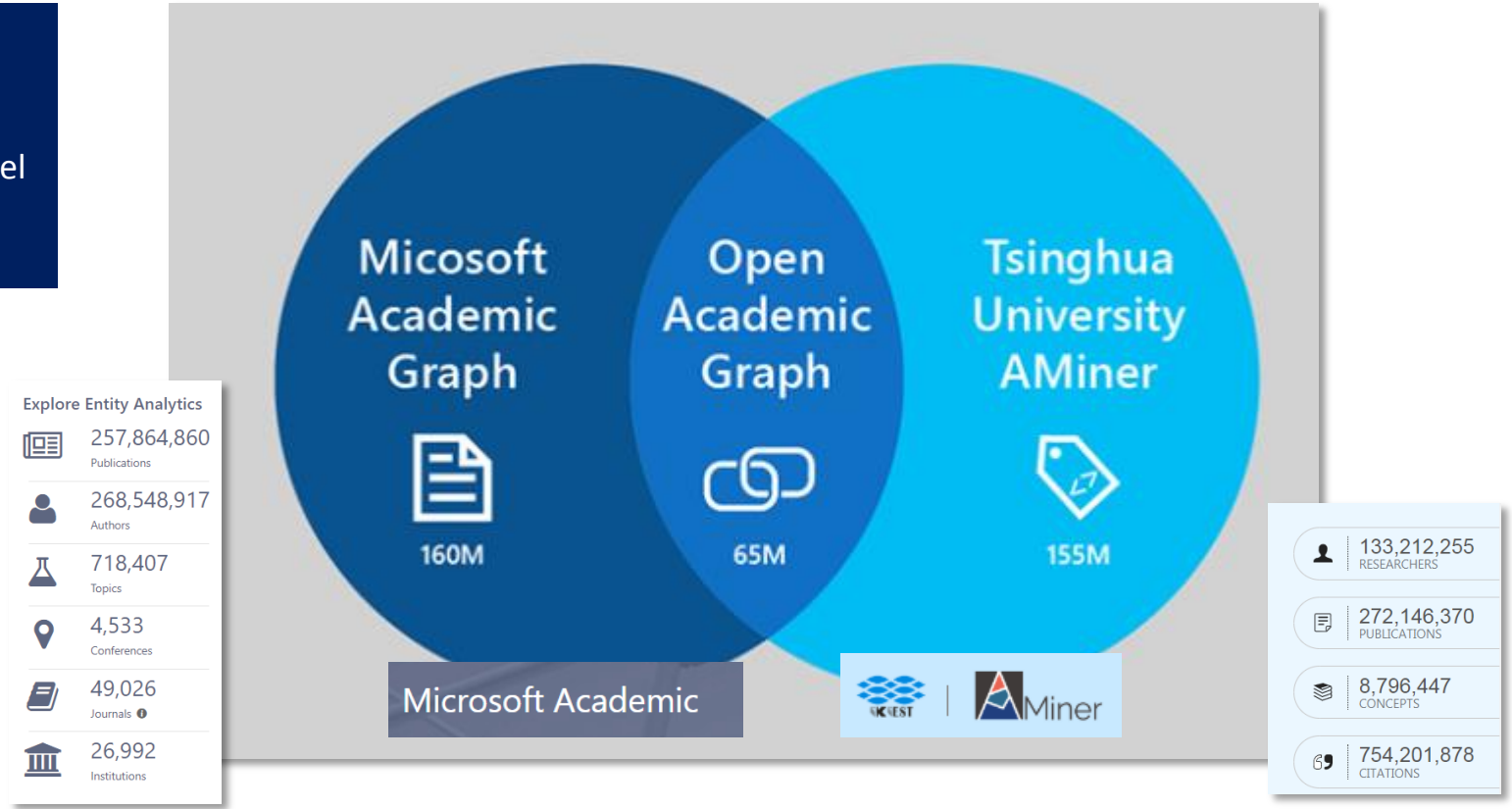


TRR Ontology-based text classification framework

Linking publications to specific field of study (FOS)

Open Academic Graph (billion-scale OAG)

- Open
- 6-level hierarchy
- Extreme granularity
- Lower complexity model to assign text to one or several themes



>700 million entities and > 2 billion relationships

Tracking of Research Results

European Commission

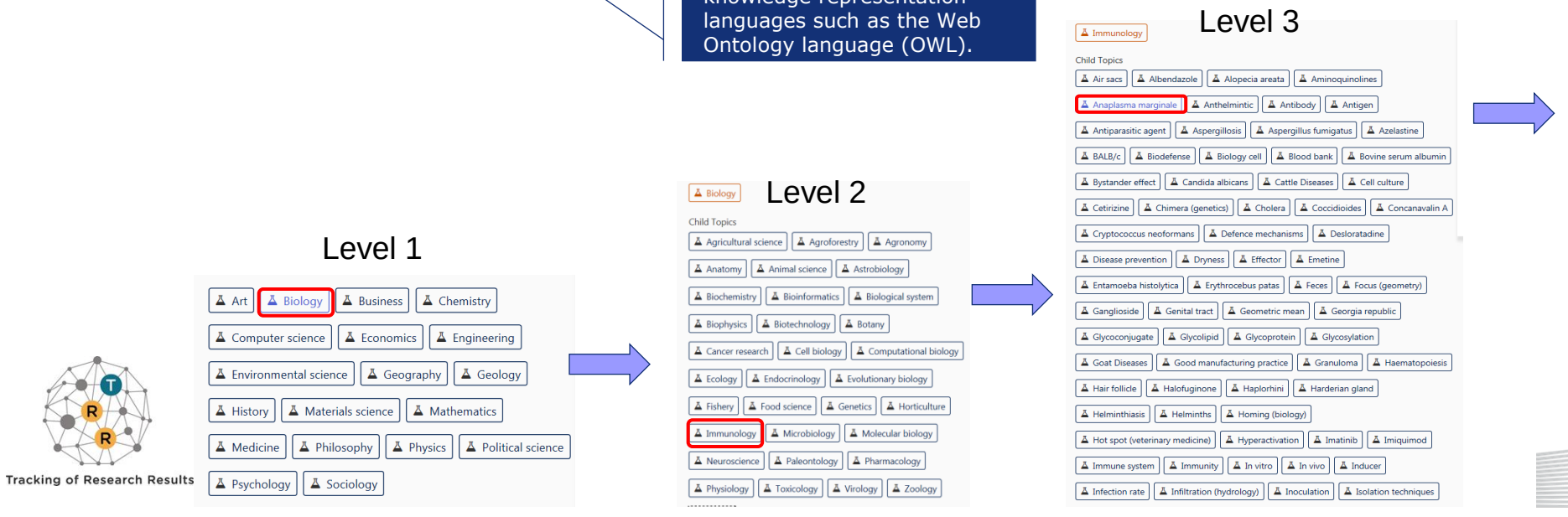
Microsoft Academic (MAG)Field Of Study (FOS)

- Microsoft Academic Graph 709K Fields of Study concepts semi-automatically constructed with ID and hierarchy
- 6 levels
  - Level 1 entirely manually constructed
  - Level 2 curated manually
  - Levels 3-6 : 100% automatically using hierarchical clustering

is a free and open knowledge base that can be read and edited by both humans and machines.

- Constructed from Wikipedia so quasi total mapping to Wikidata
  - In the process of importing MAG FOS in Wikidata (wdt:P6366)
  - Soon possibility to use any other thesaurus mapped to wikidata, through the FOS to retrieve/categorize publications/projects.
- Simple Hierarchy modeled by TRR in SKOS

SKOS Simple Knowledge Organization System Reference W3C Recommendation.
It may be used on its own, or in combination with formal knowledge representation languages such as the Web Ontology language (OWL).



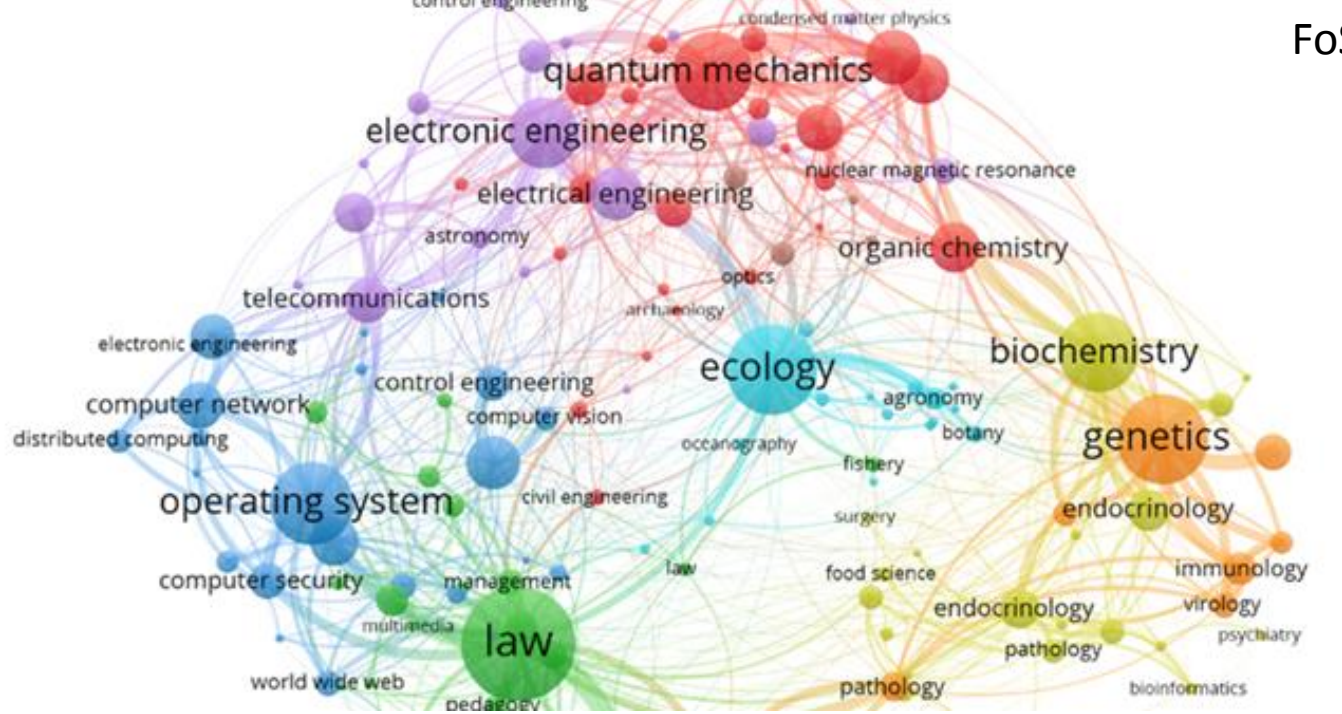
Tracking of Research Results

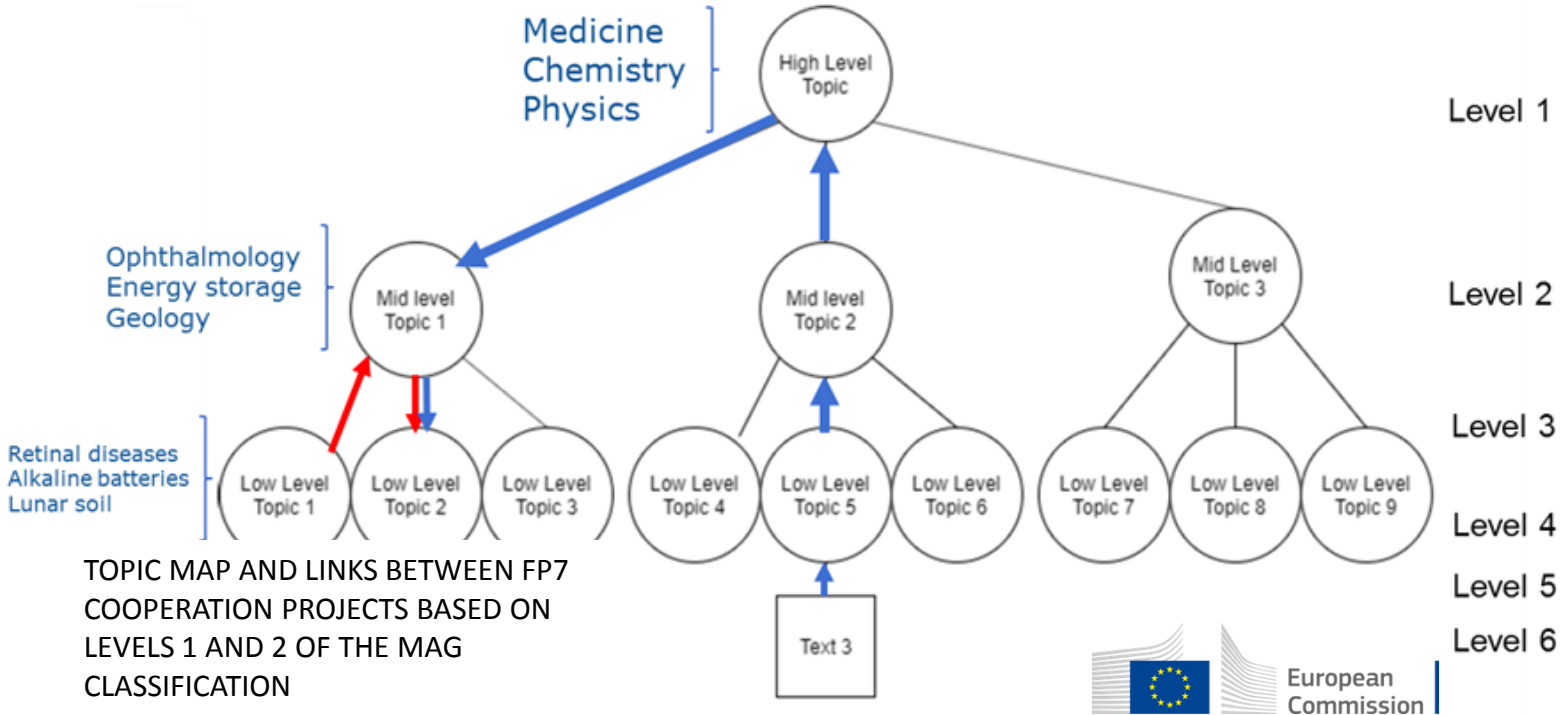
European Commission

Estimating distance between texts using topic ontology

- >200M publications = 709K FoS in the hierarchical MAG ontolgy

- > 200K topics are linked to Wikipedia articles and WikiData





TOPIC MAP AND LINKS BETWEEN FP7 COOPERATION PROJECTS BASED ON LEVELS 1 AND 2 OF THE MAG CLASSIFICATION