EFSA ROADMAP TO AI ONGOING AND FUTURE PROJECTS

Dr. Yannick G. Spill, Senior Data Scientist, EFSA

15th scientific conference of the Bulgarian focal point of EFSA

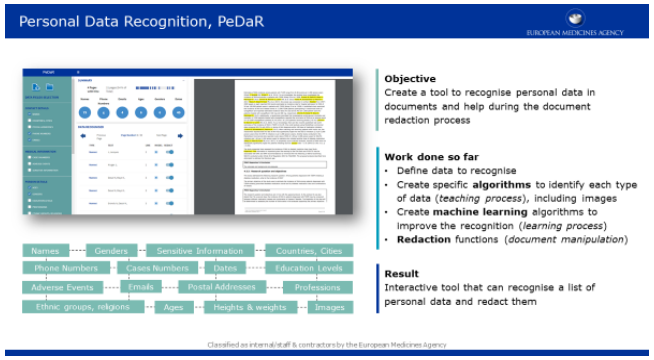
May 18th 2023

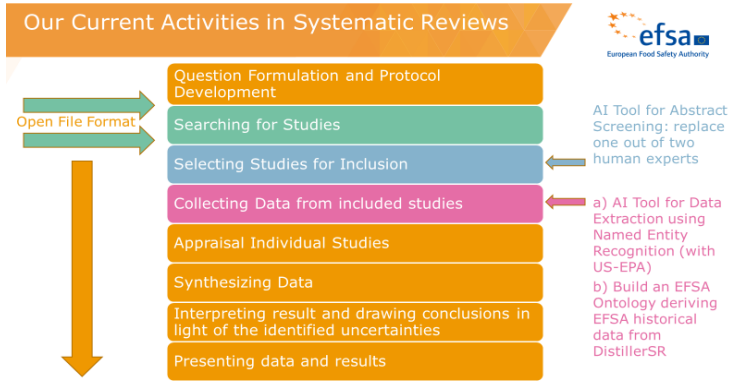
EUROPEAN FOOD SAFETY AUTHORITY

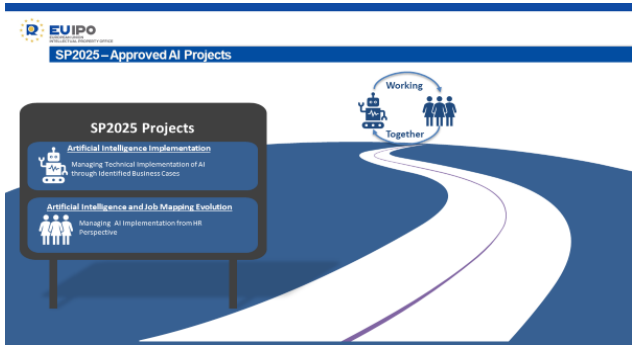
HOW IT STARTED (2018)

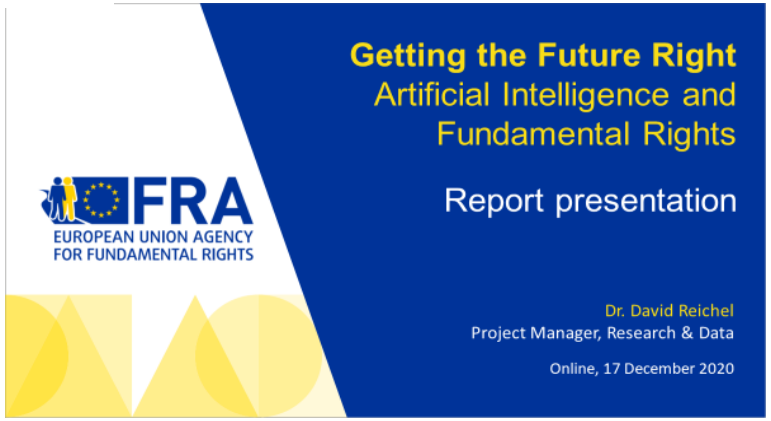
Use cases
#1 Signal Detection
#2 Evidence Management (EFSA)
#3 Forecasting
#4 Automated Reporting
#5 Image Processing (EUIPO)
#6 Content Sanitisation (EMA)
#7 Expert Identification

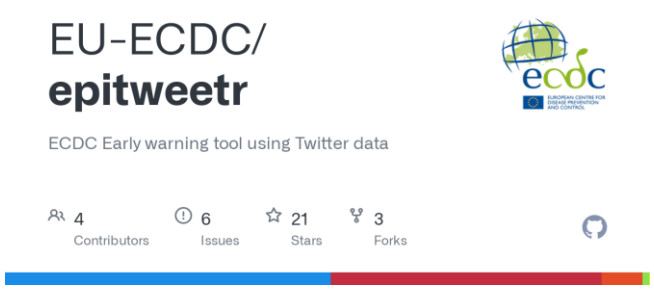
OUR JOURNEY SO FAR





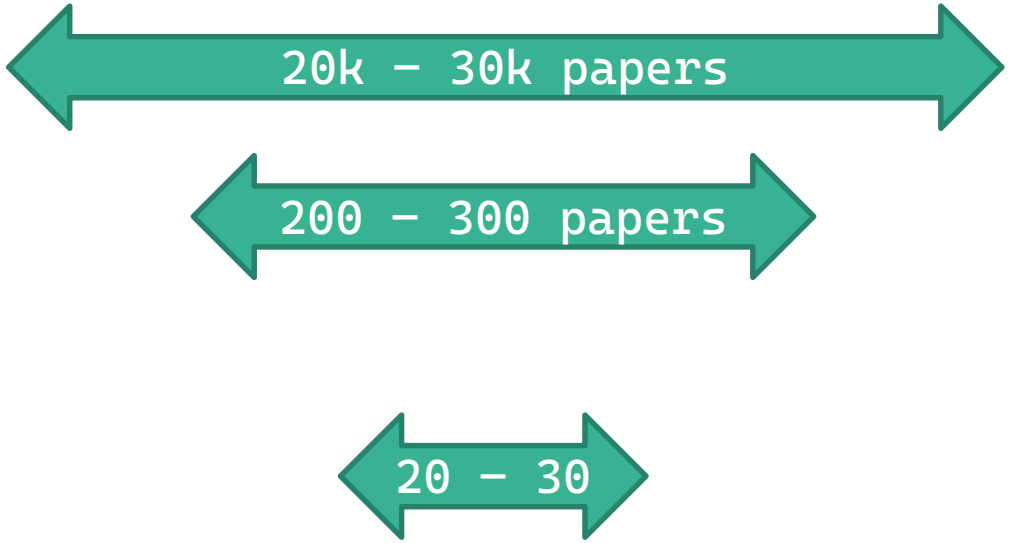






A TYPICAL EFSA SYSTEMATIC REVIEW

- Question Formulation and Protocol Development
- Searching for Studies
- Selecting Studies for Inclusion
- Collecting Data from included studies
- Appraisal Individual Studies
- Synthesizing Data
- Interpreting result and drawing conclusions in light of the identified uncertainties
- Presenting data and results



OUR CURRENT ACTIVITIES IN SYSTEMATIC REVIEWS

Question Formulation and Protocol Development

Searching for Studies

Selecting Studies for Inclusion

Collecting Data from included studies

Appraisal Individual Studies

Synthesizing Data

Interpreting result and drawing conclusions in light of the identified uncertainties

Presenting data and results

Topic Modelling

AI Tool for Abstract Screening: savings of 50%

AI Tool for Data Extraction using Named Entity Recognition

Automate Risk of Bias evaluation

TOPIC MODELLING PROJECT

How can NLP support EFSA to

- **Classify docs** according to pre-specified categories
- **Group comments** from a public consultation
- **Set up** the putative structure of an **adverse outcome pathway** through scientific papers
- **Detect emerging risks** in food & feed in news articles

Case studies

- CS1: **APRIO** classification
- CS2: **Sugar** public consultation
- CS3: **AOP** on developmental neurotoxicity
- CS4: **Beeswax** adulteration in news articles

Main finding

Classification and topic modelling are **not perfect** but provide **qualitative results** and can **accelerate** the work of subject-matter specialists via tailored algorithms and web applications

Explored NLP techniques

- **Classification**: XGBoost, FastText
- **Topic Modeling**: Top2Vec, LDA, ProdLDA, NeuralLDA, CTM, BERTopic
- **Embeddings**: TF-IDF, Doc2Vec, TSDAE, SPECTER, USE, MiniLM

Model optimization via hyperparameter tuning
Validation by involvement of subject-matter specialists

IT deliverables including NLP pipeline and web applications deployed via **infrastructure-as-code Templates** within EFSA cloud innovation zone

Proof-of-concept **web applications** for exploring NLP model results:

- Technology: Streamlit
- Built by data scientists
- Targeted at subject-matter specialists

EFSA'S 2020 RESEARCH AND INNOVATION THEME PAPER | AI

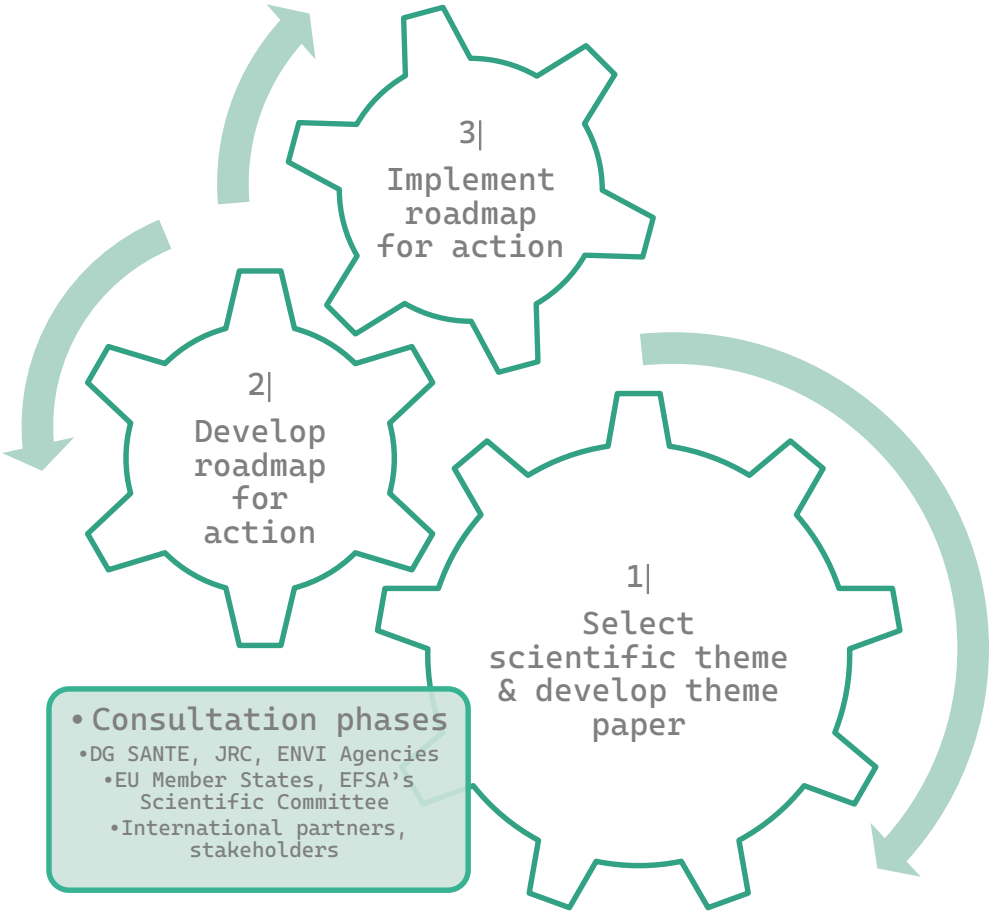
Vision: by 2027 EFSA to achieve

- i) an increase in the accessibility and the breadth of the body of evidence,
- ii) enhancing the trustworthiness in the risk assessment process, and
- iii) apply human centric artificial intelligence in close co-existence with the human expertise

Theme (concept) paper - Artificial Intelligence in risk assessment

Published: 31 Maggio 2022 **Approved:** 16 Ottobre 2020

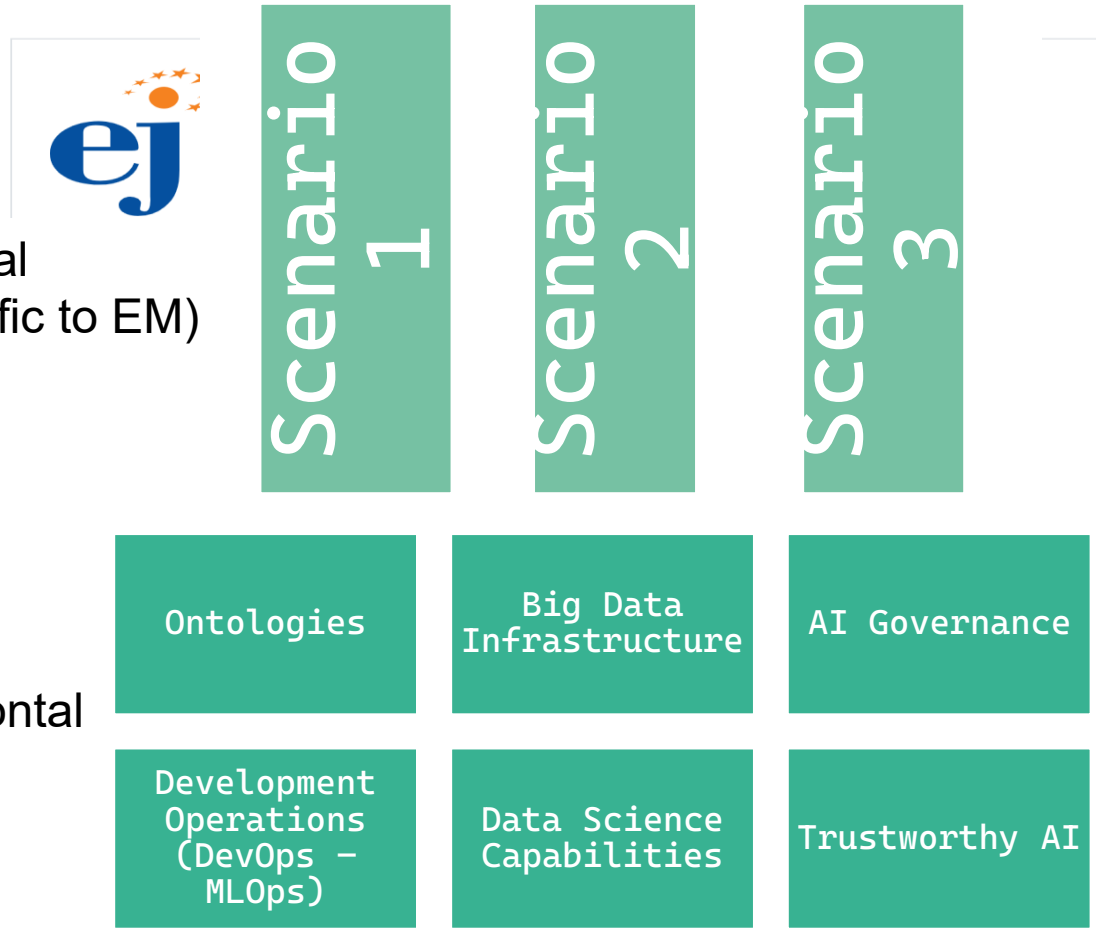
EFSA Journal on the Wiley Online Library



AI ROADMAPS: 4.6M€ UNTIL 2027

Roadmap for actions on artificial intelligence for evidence management in risk assessment

Published: 16 Maggio 2022 **Approved:** 1 Aprile 2022



Use cases	Vendor Scenarios
Data Collection Terminology Assessment SLR Keywords Identification SLR Relevant Abstract Screening SLR Literature De-duplication Text Summarization SLR Literature Findings Clustering	1: procurement
Quality Improvement of the Scientific Output	2: innovation procurement, including ProofOfConcept
Expert pool Identification Expert Selection Comparable Appraisals Detection	3: grant for research or innovation contest

PLAN TO 2027

Recommendation	H2 2022	H1 2023	H2 2023	H1 2024	H2 2024	H1 2025	H2 2025	H1 2026	H2 2026	H1 2027	H2 2027
Programme management	***Programme management***										
Horizontal 1		***Develop (or adopt existing) Ontologies relevant to EFSA***		Continuous maintenace required for this recommendation							
Horizontal 2				***Adopt a Data Infrastructure for Big Data and AI***		Training required for this recommendation					
Horizontal 3	***Design and Implement Data Governance***		Training required for this recommendation								
Horizontal 4					***Implement DevOps***	Training and adoption					
Horizontal 5		***Adopt a Trustworthy AI framework - CoC***		***Adopt a Trustworthy AI framework - Risk Mgmt Framework***							
			Training required for this recommendation								
Horizontal 6	***Assessment of Data Science Skills and recruitment/upskilling plan***		First round of recruitment			Second round of recruitment*					
		Upskilling of current staff									
Vertical scenario 1	*Scenario 1 UC 1*										
		Scenario 1 UC 2									
			Scenario 1 UC 3								
				Scenario 1 UC 4							
				Scenario 1 UC 5							
					Scenario 1 UC 6						
Vertical scenario 2						*Scenario 2 UC 7*					
Vertical scenario 3							*Scenario 3 UC 8*				
								Scenario 3 UC 9			
									Scenario 3 UC 10		

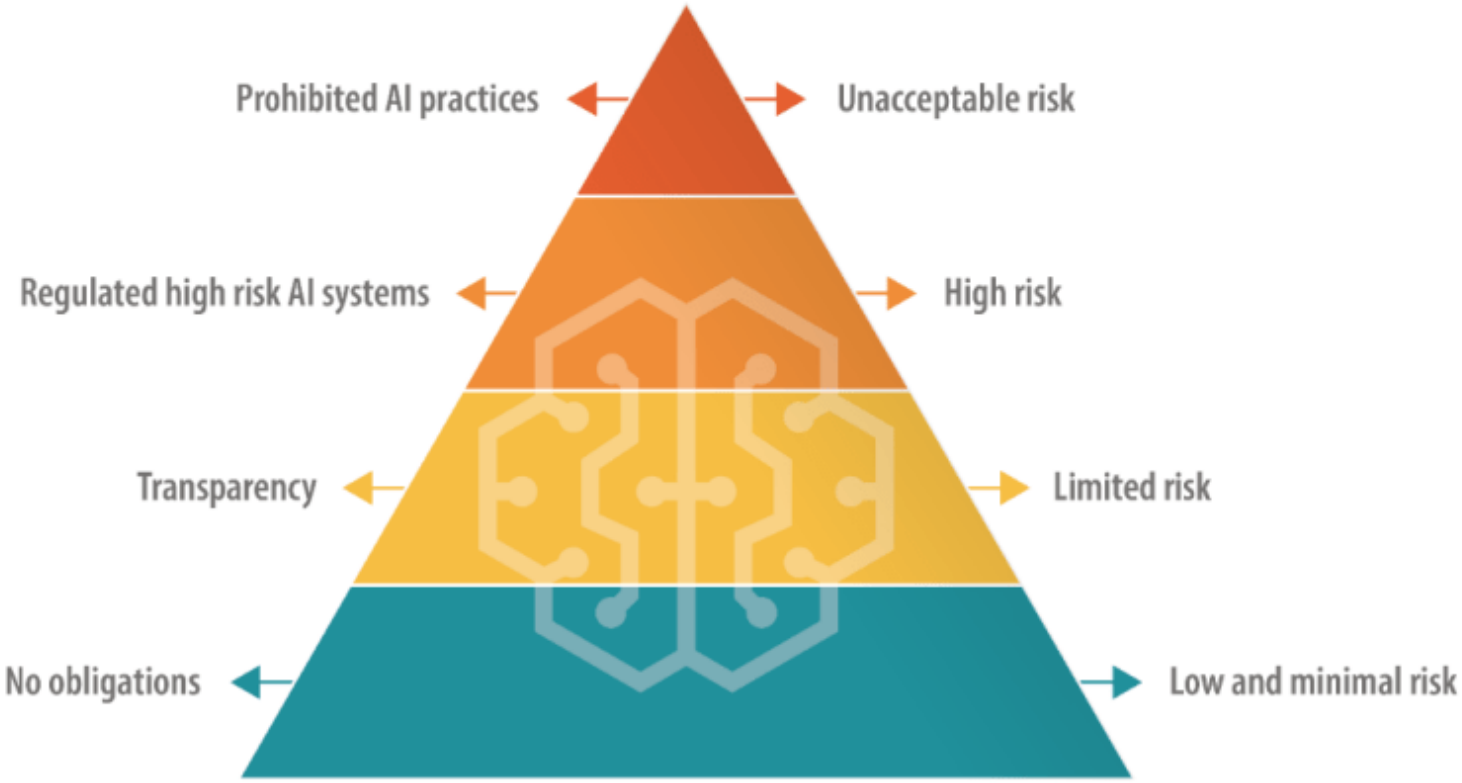
PLAN TO 2027

Recommendation	H2 2022	H1 2023	H2 2023	H1 2024	H2 2024	H1 2025	H2 2025	H1 2026	H2 2026	H1 2027	H2 2027
Programme management	***Programme management***										
Horizontal 1		***Develop (or adopt existing) Ontologies relevant to EFSA***		Continuous maintenace required for this recommendation							
Horizontal 2				***Adopt a Data Infrastructure for Big Data and AI***		Training required for this recommendation					
Horizontal 3	***Design and Implement Data Governance***		Training required for this recommendation								
Horizontal 4					***Implement DevOps***	Training and adoption					
Horizontal 5		***Adopt a Trustworthy AI framework - CoC***		***Adopt a Trustworthy AI framework - Risk Mgmt Framework***							
			Training required for this recommendation								
Horizontal 6	***Assessment of Data Science Skills and recruitment/upskilling plan***		First round of recruitment			Second round of recruitment*					
		Upskilling of current staff									
Vertical scenario 1	*Scenario 1 UC 1*										
		Scenario 1 UC 2									
			Scenario 1 UC 3								
				Scenario 1 UC 4							
				Scenario 1 UC 5							
					Scenario 1 UC 6						
Vertical scenario 2						*Scenario 2 UC 7*					
Vertical scenario 3							*Scenario 3 UC 8*				
								Scenario 3 UC 9			
									Scenario 3 UC 10		

ETHICS AND AI

4 Ethical Principles	- Respect for human autonomy - Prevention of harm - Fairness - Explicability
7 Key Requirements	- Human agency and oversight - Technical robustness and safety - Privacy and data governance - Transparency - Diversity, non-discrimination and fairness - Societal and environmental wellbeing - Accountability

AI ACT



Data source: European Commission.

UNACCEPTABLE RISK

"The proposal also prohibits AI-based **social scoring** for general purposes done by public authorities."

"Finally, the use of 'real time' **remote biometric identification systems** in publicly accessible spaces for the purpose of law enforcement is also prohibited unless certain limited exceptions apply."

HIGH RISK

"… AI systems identified as high-risk should be limited to those that have a significant harmful impact on the **health, safety and fundamental rights of persons** in the Union …"

"… Similarly, in the **health sector** where the stakes for life and health are particularly high, **increasingly sophisticated diagnostics systems and systems supporting human decisions** should be reliable and accurate…"

"… High-risk AI systems should only be placed on the Union market or put into service if they comply with certain **mandatory requirements** …":

- Risk management system
- **Data and Data Governance**
- Technical documentation
- Record keeping
- Transparency and provision of information to users
- Human oversight
- Accuracy, robustness and cybersecurity

PLAN TO 2027

Recommendation	H2 2022	H1 2023	H2 2023	H1 2024	H2 2024	H1 2025	H2 2025	H1 2026	H2 2026	H1 2027	H2 2027
Programme management	*Programme management*										
Horizontal 1		*Develop (or adopt existing) Ontologies relevant to EFSA*		Continuous maintenace required for this recommendation							
Horizontal 2				*Adopt a Data Infrastructure for Big Data and AI*		Training required for this recommendation					
Horizontal 3	*Design and Implement Data Governance*		Training required for this recommendation								
Horizontal 4					*Implement DevOps*	Training and adoption					
Horizontal 5		*Adopt a Trustworthy AI framework - CoC*		*Adopt a Trustworthy AI framework - Risk Mgmt Framework*							
			Training required for this recommendation								
Horizontal 6	*Assessment of Data Science Skills and recruitment/upskilling plan*		First round of recruitment			Second round of recruitment*					
		Upskilling of current staff									
Vertical scenario 1	*Scenario 1 UC 1*										
		Scenario 1 UC 2									
			Scenario 1 UC 3								
				Scenario 1 UC 4							
				Scenario 1 UC 5							
					Scenario 1 UC 6						
Vertical scenario 2						*Scenario 2 UC 7*					
Vertical scenario 3							*Scenario 3 UC 8*				
								Scenario 3 UC 9			
									Scenario 3 UC 10		

AI VERTICAL USE CASES – SCENARIO 1

- **UC1: Text summarization** (applied to attachments of public consultation's comments)
- **UC2: SR Keyword Identification** (selection of tools to automate keyword identification)
- **UC3: SR De-duplication** (analysis of DistillerSR automatic de-duplication feature)
- **UC4: SR AI Screening** (producing a manual to perform Title and Abstract screening using AI features in DistillerSR)
- **UC5: SR Finings Clustering** (producing a manual to use AI Classifiers features of DistillerSR)
- Additional deliverable: external scientific report summarizing UC1-5

PLAN TO 2027

Recommendation	H2 2022	H1 2023	H2 2023	H1 2024	H2 2024	H1 2025	H2 2025	H1 2026	H2 2026	H1 2027	H2 2027
Programme management	***Programme management***										
Horizontal 1		***Develop (or adopt existing) Ontologies relevant to EFSA***		Continuous maintenace required for this recommendation							
Horizontal 2				***Adopt a Data Infrastructure for Big Data and AI***		Training required for this recommendation					
Horizontal 3	***Design and Implement Data Governance***		Training required for this recommendation								
Horizontal 4					***Implement DevOps***	Training and adoption					
Horizontal 5		***Adopt a Trustworthy AI framework - CoC***		***Adopt a Trustworthy AI framework - Risk Mgmt Framework***							
			Training required for this recommendation								
Horizontal 6	***Assessment of Data Science Skills and recruitment/upskilling plan***		First round of recruitment			Second round of recruitment*					
		Upskilling of current staff									
Vertical scenario 1	*Scenario 1 UC 1*										
		Scenario 1 UC 2									
			Scenario 1 UC 3								
				Scenario 1 UC 4							
				Scenario 1 UC 5							
					Scenario 1 UC 6						
Vertical scenario 2						*Scenario 2 UC 7*					
Vertical scenario 3							*Scenario 3 UC 8*				
								Scenario 3 UC 9			
									Scenario 3 UC 10		

AI GOVERNANCE

The project aims at supporting EFSA in the definition of an AI governance framework, becoming an AI enabled organization

AI governance is a system of rules, practices, processes, and technological tools that are employed to ensure an organization's use of AI technologies aligns with the organization's strategies, objectives, and values; fulfils legal requirements; and meets principles of ethical AI followed by the organization.

How will the project impact EFSA?

Project's output

A *new operating model* among POTI dimensions:

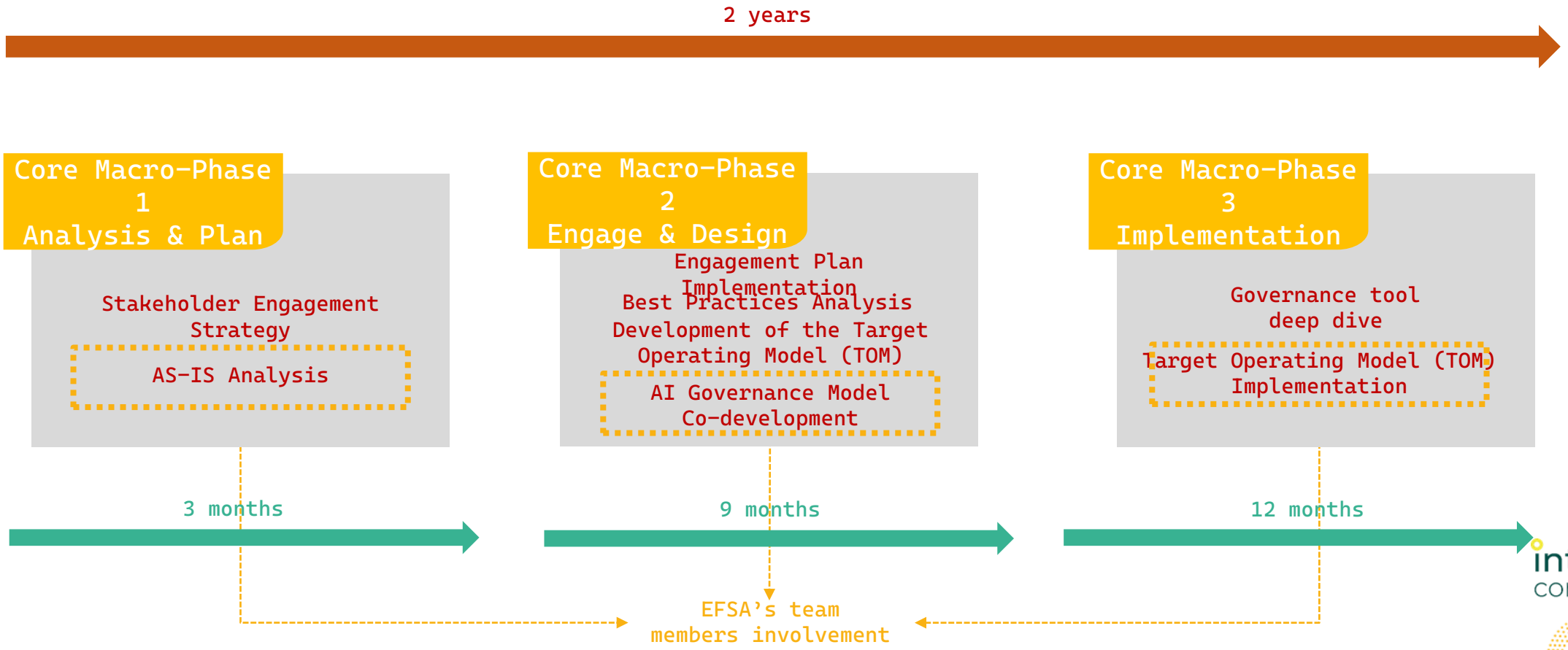
- Process: change of processes and introduction of new ones
- Organization: organizational change to adopt the new operating model
- Technology: new technological tools
- Information: change of data flows, data requirements and input

Project's outcome

EFSA as an AI-enabled organization with respect to data governance

intellera consulting

PROJECT OVERVIEW



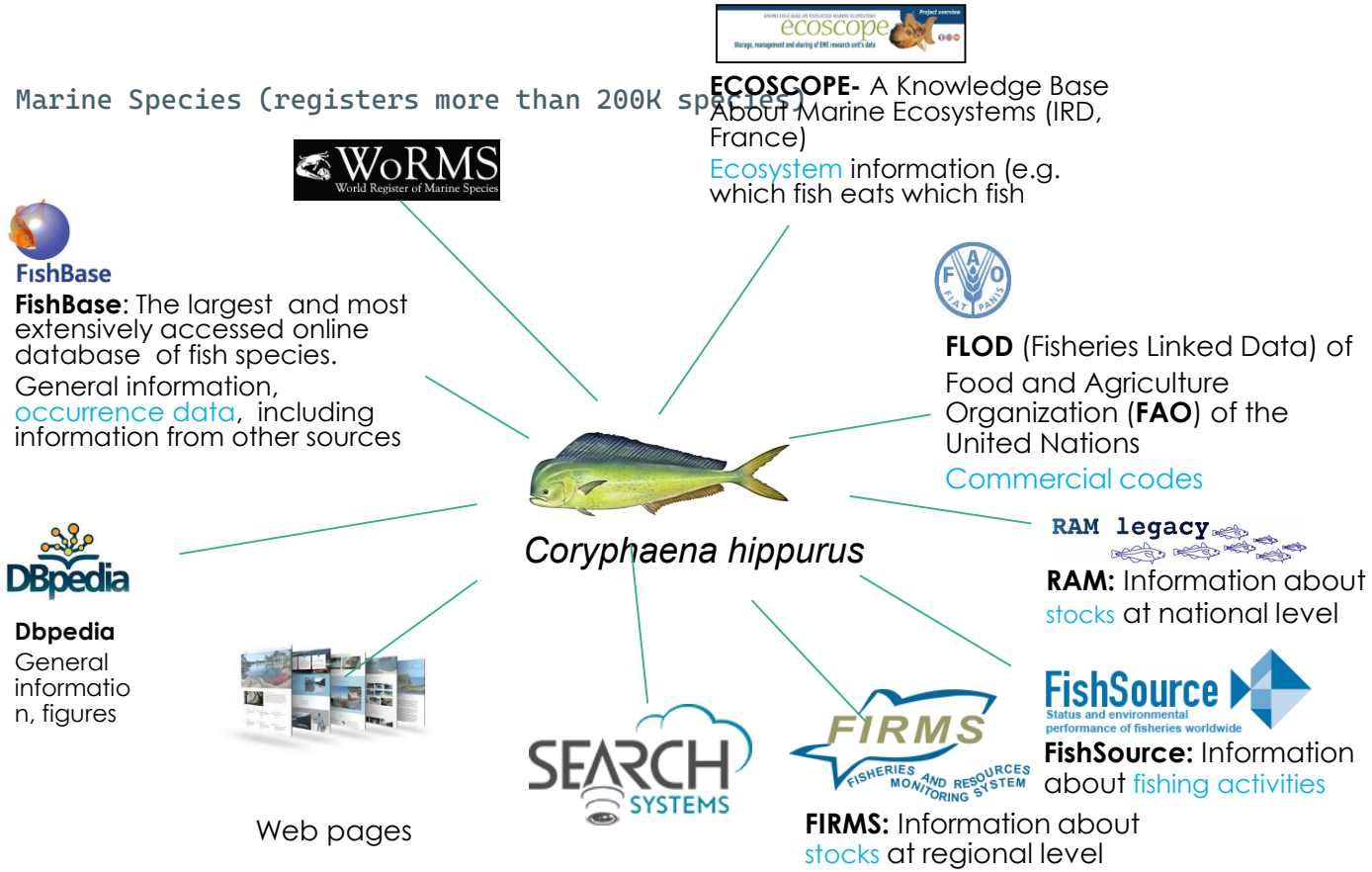
PLAN TO 2027

Recommendation	H2 2022	H1 2023	H2 2023	H1 2024	H2 2024	H1 2025	H2 2025	H1 2026	H2 2026	H1 2027	H2 2027
Programme management	***Programme management***										
Horizontal 1		***Develop (or adopt existing) Ontologies relevant to EFSA***		Continuous maintenace required for this recommendation							
Horizontal 2				***Adopt a Data Infrastructure for Big Data and AI***		Training required for this recommendation					
Horizontal 3	***Design and Implement Data Governance***		Training required for this recommendation								
Horizontal 4					***Implement DevOps***	Training and adoption					
Horizontal 5		***Adopt a Trustworthy AI framework - CoC***		***Adopt a Trustworthy AI framework - Risk Mgmt Framework***							
			Training required for this recommendation								
Horizontal 6	***Assessment of Data Science Skills and recruitment/upskilling plan***		First round of recruitment			Second round of recruitment*					
	Upskilling of current staff										
Vertical scenario 1	*Scenario 1 UC 1*										
		Scenario 1 UC 2									
			Scenario 1 UC 3								
				Scenario 1 UC 4							
				Scenario 1 UC 5							
					Scenario 1 UC 6						
Vertical scenario 2						*Scenario 2 UC 7*					
Vertical scenario 3							*Scenario 3 UC 8*				
								Scenario 3 UC 9			
									Scenario 3 UC 10		

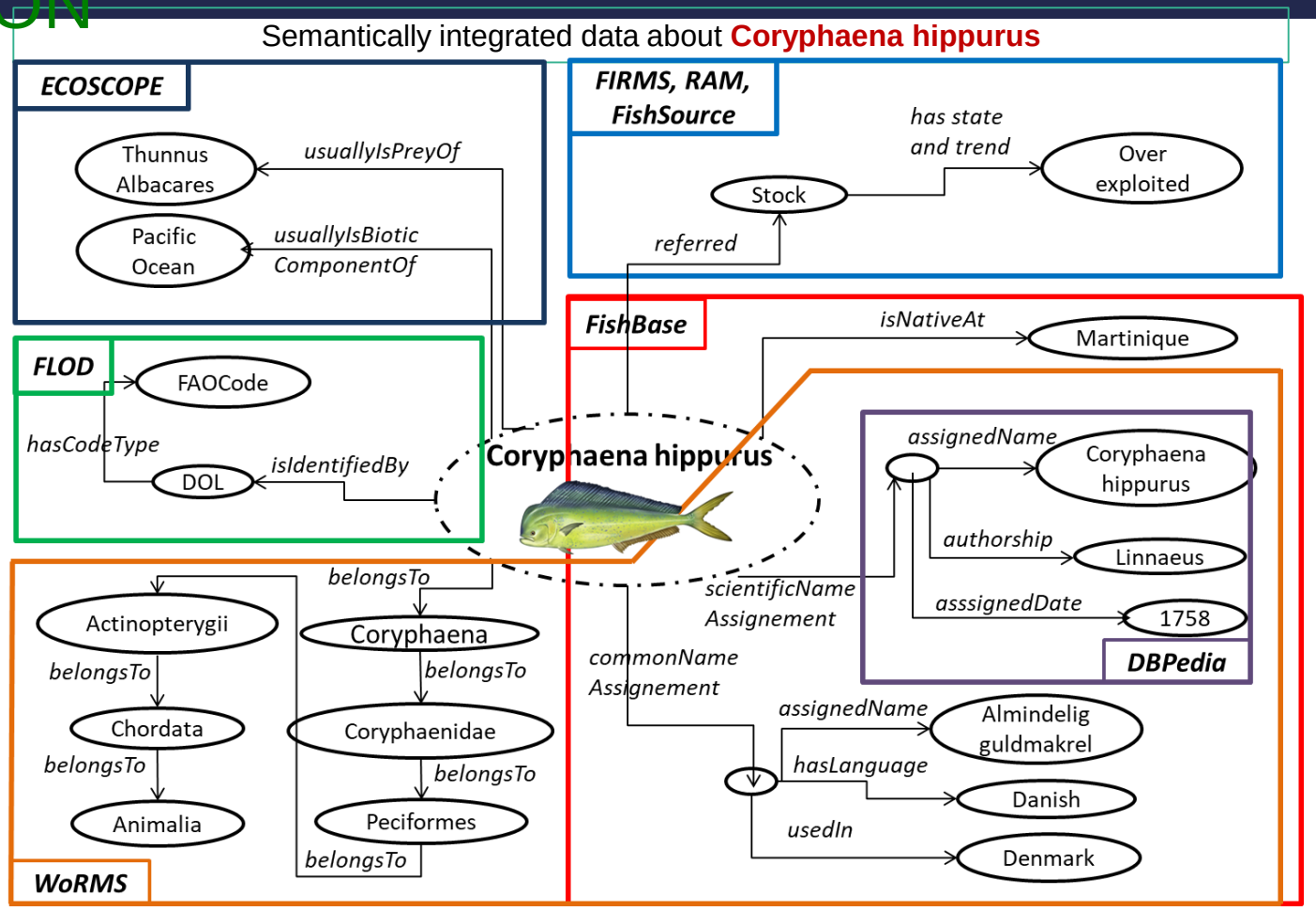
MARINE DOMAIN: BEFORE ONTOLOGY-BASED INTEGRATION

WoRMS: World Register of Marine Species (registers more than 200K species)

Taxonomic information

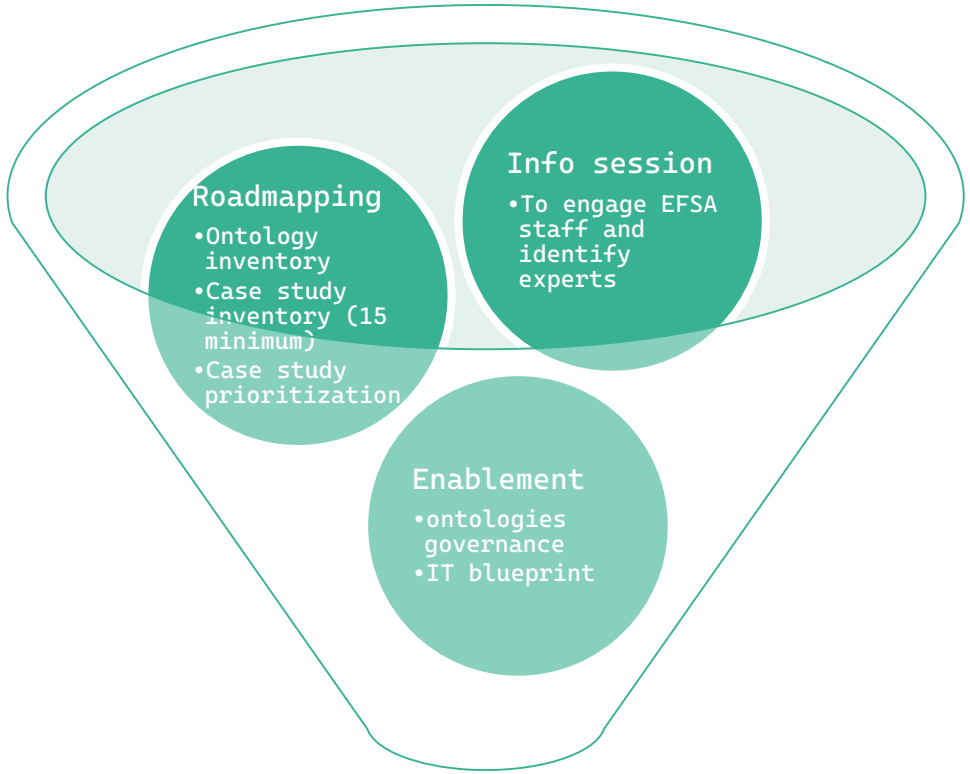


MARINE DOMAIN: AFTER ONTOLOGY-BASED INTEGRATION



Knowledge Graph obtained after defining a top-level ontology (**MarineTLO**) and transforming data from the existing sources

ONTOLOGY ROADMAPPING AND CASE STUDY IMPLEMENTATION



Implementation

- Minimum 3 case studies
- incl. AI component

EFSA-RELATED CASE STUDIES THAT INVOLVE ONTOLOGIES (ONGOING)

- 1/ Standardization of vocabulary in SLR data extraction step, aiming to make SLR extracted data searchable and retrievable
- 2/ Automated toxicological data extraction from scientific opinions to populate openfoodtox/IUCLID, building on Toxicological endpoint ontology and database. OECD Harmonised Templates (OHTs) should be considered when analysing current terminology used for chemical and hazard data reporting.
- 3/ Improved data sharing from/to external stakeholders through ontology-induced constraints in the data models
- 4/ EFSA Catalogue browser uses ontology-driven constraints and inference (instead of programmatic ones) for FoodEx2 encoding including facets
- 5/ FoodEx2-SCA made multilingual and/or more resilient to synonyms/typos using ontologies
- 6/ Use of ontologies to classify reported FoodEx2 terms within individual regulation classifications
- 7/ EFSA website glossary uses an ontology backend
- 8/ EFSA website's tags on scientific outputs added automatically using Named Entity Recognition and building on ontology (derived from website taxonomy)
- 9/ Scientific output keywords are set before publication (e.g., on Wiley or Appian) and can be linked to other similar keywords and associated scientific outputs through their relationship in an ontology
- 10/ Task 2 of the SPIDO AI vertical use-cases: Systematic Literature Review Keywords Identification
- 11/ Sci-ASK: Knowledge management ontology for the knowledge organization process in EFSA scientific output (opinions, protocols, model, etc)

The analysis will be completed end of May 2023

UPCOMING ACTIVITIES

Recommendation	H2 2022	H1 2023	H2 2023	H1 2024	H2 2024	H1 2025	H2 2025	H1 2026	H2 2026	H1 2027	H2 2027
Programme management	*Programme management*										
Horizontal 1		*Develop (or adopt existing) Ontologies relevant to EFSA*		Continuous maintenace required for this recommendation							
Horizontal 2				*Adopt a Data Infrastructure for Big Data and AI*		Training required for this recommendation					
Horizontal 3	*Design and Implement Data Governance*		Training required for this recommendation								
Horizontal 4					*Implement DevOps*	Training and adoption					
Horizontal 5		*Adopt a Trustworthy AI framework - CoC*		*Adopt a Trustworthy AI framework - Risk Mgmt Framework*							
			Training required for this recommendation								
Horizontal 6	*Assessment of Data Science Skills and recruitment/upskilling plan*		First round of recruitment			Second round of recruitment*					
		Upskilling of current staff									
Vertical scenario 1	*Scenario 1 UC 1*										
		Scenario 1 UC 2									
			Scenario 1 UC 3								
				Scenario 1 UC 4							
				Scenario 1 UC 5							
					Scenario 1 UC 6						
Vertical scenario 2						*Scenario 2 UC 7*					
Vertical scenario 3							*Scenario 3 UC 8*				
								Scenario 3 UC 9			
									Scenario 3 UC 10		

STAY CONNECTED

SUBSCRIBE TO
efsa.europa.eu/en/news/newsletters
efsa.europa.eu/en/rss
Careers.efsa.europa.eu – job alerts

LISTEN TO OUR PODCAST
Science on the Menu –Spotify, Apple Podcast and YouTube

FOLLOW US ON TWITTER
@efsa_eu @methods_efsa
@plants_efsa @animals_efsa

FOLLOW US ON LINKEDIN
Linkedin.com/company/efsa

FOLLOW US ON INSTAGRAM
@one_healthenv_eu

CONTACT US
efsa.europe.eu/en/contact/askefsa

EUROPEAN FOOD SAFETY AUTHORITY