European Commission

The European Commission's Knowledge Centre for Biodiversity

Mapping interlinkages between EU-funded research projects and EU policy priorities

Application of the TIM open access tool to case studies in the context of the EU Biodiversity Strategy for 2030

TIM

In 2020 the European Commission (EC) adopted the **EU Biodiversity Strategy for 2030**[1] "Bringing nature back into our lives", which includes a comprehensive Action Plan for protecting nature and reversing the degradation of ecosystems. This mapping exercise - an initiative of the EC Knowledge Centre for Biodiversity[2] (KCBD) – aims to support this Action Plan by contributing to knowledge management at the science-policy interface. Conducted in support of the EC Directorate General for Research and Innovation, it explores ongoing and past EU-funded research projects carried out over the past 15 years and maps their linkages with the Biodiversity Strategy, particularly its four pillars: 1 - Protect Nature, 2 - Restore Ecosystems, 3 - Transformative change and 4 - EU as a global leader. It also includes a pilot case study that aims to address linkages with cross-sectoral topics such as Biodiversity and Health.

[1] https://ec.europa.eu/environment/strategy/biodiversity-strategy-2030_en
[2] https://knowledge4policy.ec.europa.eu/biodiversity_en

Data source and text mining software

The primary source of information for EU-funded research projects is the EC CORDIS database, which can be filtered using simple keywords searches. This database includes the metadata files of all projects. This study proposes an advanced analysis based on the metadata using the *Tools for Innovation Monitoring* (TIM)[3] Open Access version developed by the Joint Research Centre of the European Commission (JRC).

Methodology and workflow

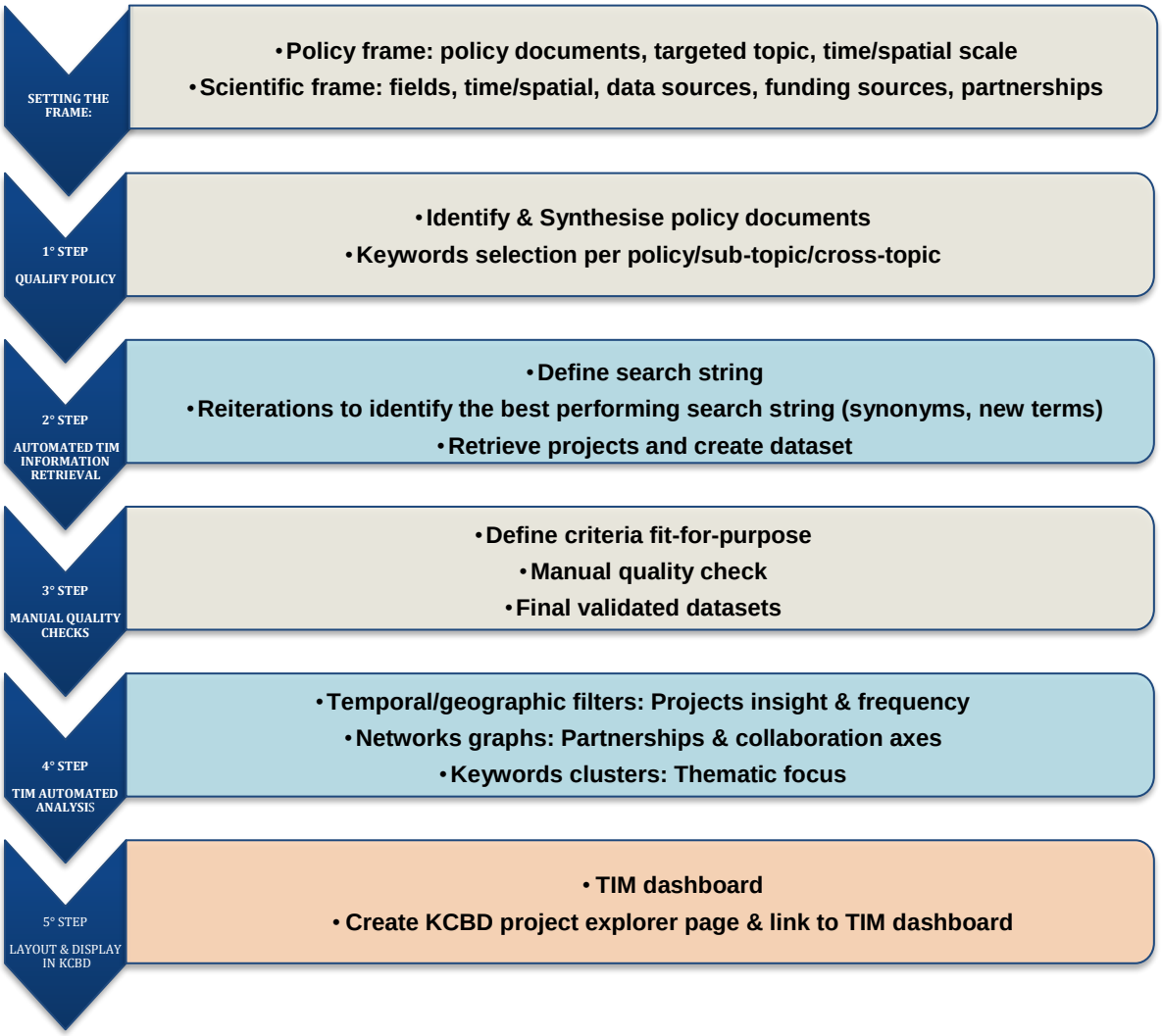
The main steps of the methodology used to scan EU-funded research projects for their linkages to policy priorities are described by the workflow in Figure 1.

After analyzing policy documents (step1), a conditional search string consisting of keywords and logical operators is defined by the user to best qualify the topic and/or policy of interest (step 2). A dataset is created based on the projects retrieved using the search string. In step 3 of the workflow, the dataset is finalised using expert judgement, particularly to exclude irrelevant projects. In steps 4 and 5, the dataset is explored through three main lenses in TIM and displayed as a TIM dashboard including network graphs. These show (1) general information on the dataset, (2) collaborative networks and partnerships at the level of countries and institutions and (3) thematic foci - keywords that are automatically retrieved by text mining and natural language processing algorithms.

Queries on a specific time frame, country, organisation or keyword can be carried out and analysed to understand and identify temporal trends, budget lines, main research institutions and their main collaborators, the most proactive countries, the main topics of research, and the frequency of keywords and their connections within the projects' data. Such information is relevant for policies and, especially in relation to strategies at the European level, helps inform the funding of new research areas, and tackle thematic gaps and under-represented countries, so as to ensure the achievement of policy goals and targets.

[3] EU, Joint Research Centre, TIM Analytics, http://www.timanalytics.eu/

Figure 1. Workflow to map interlinkages between EU-funded research projects and policies based on TIM Open Access software processing.



Source: JRC, 2022[4].

Application of the TIM open access tool to case studies

Mapping exercises were carried out according to the workflow in order to identify the relevance of EU research projects funded during the period 2007-2021 (using metadata from the CORDIS database for the Framework Programmes FP7 and Horizon 2020) in responding to biodiversity policy priorities. Two case studies were developed: one addressing the Biodiversity Strategy for 2030 (and each of its four pillars), and another addressing the Biodiversity and Health cross-policy domain. Expert judgement was used to define the search strings for the semantic translation of policy documents into keywords (see the search strings used in TIM Open Access in Table 1). Of the projects retrieved by the TIM tool, 19% for the Biodiversity Strategy and 14% for the cross-sector topic Biodiversity and Health were discarded after the manual check found them to be irrelevant (see Table 2 showing the TIM outcome).

[4] This brief is a summary of the published report EUR 31106 EN *Mapping interlinkages between EU-funded research projects and EU policy priorities*, 2022, doi:10.2760/37559, where all figures and tables are available.

Table 1. Search strings for policy needs under the Biodiversity Strategy for 2030 (best performing keywords for all pillars combined into a single string) and the cross-policy topic.

Biodiversity Strategy for 2030
topic: ("natura 2000" OR "natura2000" OR "protected area" OR "ecological corridor" OR "ecological network" OR "agro biodiversity" OR agroecology OR agrobiodiversity OR "aquatic ecosystem" OR "biodiversity conservation" OR "biodiversity loss" OR "biodiversity management" OR "biodiversity monitoring" OR "biological invasion" OR biomass OR "coastal ecosystem" OR "coastal marine" OR "conservation of biodiversity" OR "ecosystem restoration" OR "ecosystem services" OR "endangered species" OR fertiliser OR fertilizer OR "floodplain ecosystem" OR "forest conservation" OR "forest ecosystem" OR "freshwater biodiversity" OR "freshwater ecosystem" OR "habitat conservation" OR "invasive alien species" OR "invasive species" OR "loss of biodiversity" OR "marine biodiversity" OR "marine ecosystem" OR "marine litter" OR "management of biodiversity" OR "monitoring biodiversity" OR "nature based solutions" OR "nature-based solution" OR NBS OR "old growth forest" OR "organic farming" OR "pest management" OR pesticides OR pollinators OR "protected species" OR "riparian ecosystem" OR "river biodiversity" OR "river ecosystem" OR "soil biodiversity" OR "soil ecosystem" OR "species conservation" OR "species extinction" OR "sustainable agriculture" OR "sustainable fisheries" OR "terrestrial biodiversity" OR "terrestrial ecosystem" OR "threatened species" OR "urban greening" OR "biodiversity governance" OR business OR "ecological effect" OR "ecological footprint" OR "ecological impact" OR education OR "environmental compliance" OR "environmental effect" OR "environmental impact" OR "environmental law" OR financing OR "food chain" OR labelling OR "sustainable consumption" OR "sustainable food" OR "sustainable production" OR "value chain" OR "ocean governance" OR deforestation OR "forest degradation" OR "global biodiversity" OR "indigenous communities" OR "indigenous people" OR "local communities" OR "local people" OR "wildlife trade") AND **topic:** ("biodiversity" OR "bio diversity" OR "ecosystem services") AND **emm_programme:** (fp7 to h2020) AND **class:** euproject
----
---------- Biodiversity and Health
topic: ("health" OR "animal health" OR "animal welfare" OR "zoonotic") AND **topic:** ("biodiversity" OR "bio diversity") AND **emm_programme:** (fp7 to h2020) AND **class:** euproject

Source: JRC, 2022[4].

Table 2. EU-funded research projects (2007-2021) relevant to policy needs identified by TIM from CORDIS metadata files, those discarded after manual filtering (quality checks), those allocated per pillar and the final number of relevant projects to be analysed.

TIM	**EC Biodiversity Strategy 2030**					**Cross-policy topic**
	BDS 2030	**Allocation per pillar**				**Biodiversity & Health**
		1	**2**	**3**	**4**	
Automatically retrieved projects	**720**	/	/	/	/	**143**
Projects excluded after quality check	134	/	/	/	/	20
Projects after quality check	**586**	**30**	**431**	**170**	**258**	**123**
% of total number	81%	4%	60%	24%	36%	86%

Source: JRC, 2022[4].

Interlinkages between EU-funded research projects and the Biodiversity Strategy

The mapping of interlinkages between EU-funded research projects in the period 2007-2021 and the Biodiversity Strategy highlighted:

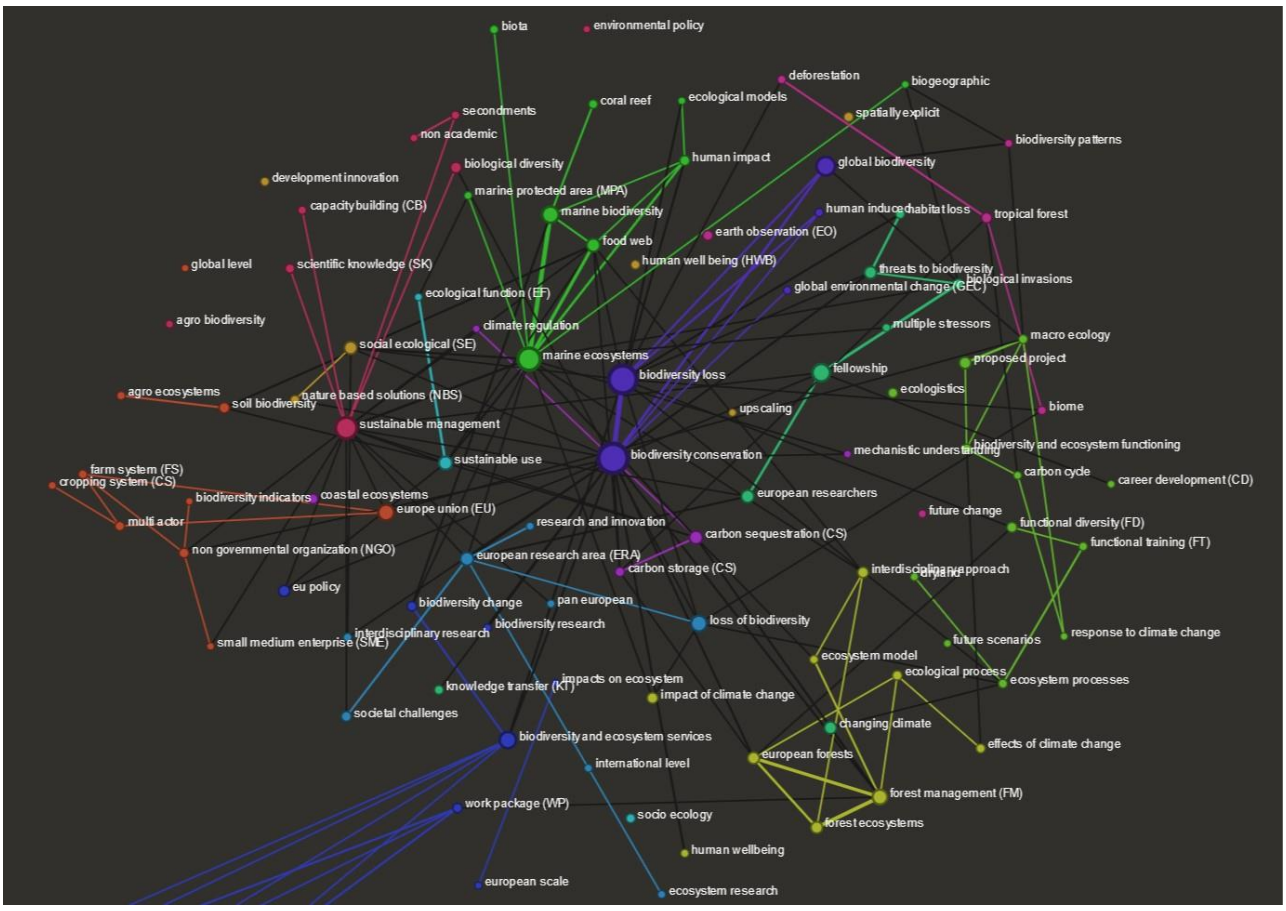
(1) an increase over time in the number of projects relevant to the Strategy and its pillars, except in the case of the first pillar on nature protection;
(2) a greater frequency in the number of projects that address cross-cutting concepts rather than topics with a single focus, such as in pillar 1;
(3) the most frequently represented countries in partnerships for biodiversity research across all four pillars were the UK followed by four EU countries (Spain, France, Germany and Italy);
(4) the same institutions from the EU featured in the top 10 of all four pillars, with one Spanish agency featuring most frequently;
(5) the main axes of frequent collaborations were UK-Germany and Spain-France, followed by the developing axis of Italy-the Netherlands; more extensive and inclusive collaborative networks were found for pillar 2 on ecosystems restoration;
(6) the distribution of projects seems to become more balanced across institutions over time.

We also analysed the specific case of the JRC. The JRC featured among the top 10 organisations in all pillars aside from pillar 1 on nature protection, but its ranking based on number of projects fell from 6th under FP7 to 18th under Horizon 2020. This is likely due to the fact that the JRC has limited its participation to mainly advisory roles in EU-funded research projects due to changes in administrative rules under H2020. The JRC was found to be very collaborative, with a total of 76 direct collaborators and an additional eight prominent partners, including the University of Wageningen.

At the global level, the mapping of projects confirmed a high concentration of projects both in Europe and coordinated by European organisations. Countries of the Middle East, Central Asia and the African continent were under-represented in the partnerships. There were few if any project partners from the Maghreb, Sahara and Congo regions.

The automated thematic analysis of projects relevant to the Biodiversity Strategy confirmed a few well-known foci (Figure 2). Biodiversity conservation and biodiversity loss appeared in relation to global biodiversity. Two prominent clusters emerged across all four pillars, aside from pillar 3: one regarding marine ecosystems (highlighted in green in Figure 2), and one for forest ecosystems (in yellow).

Figure 2. Thematic networks of keywords automatically retrieved from CORDIS metadata files of EU-funded research projects and relevant to the Biodiversity Strategy. Each network is assigned a colour and includes the keywords that most frequently occur together.



Source: JRC, 2022[4].

Exploring the results for each pillar, Earth observation appeared in relation to protected areas under pillar 1 on nature protection. Under pillar 2 on ecosystems restoration, soil biodiversity appeared as a keyword only when associated with agriculture and farming systems. Under pillar 3 on transformative change, keywords were poorly clustered and showed no specific foci. Tropical deforestation was the main issue that emerged from the forest cluster under pillar 4.

Interlinkages between EU-funded research projects and Biodiversity and Health

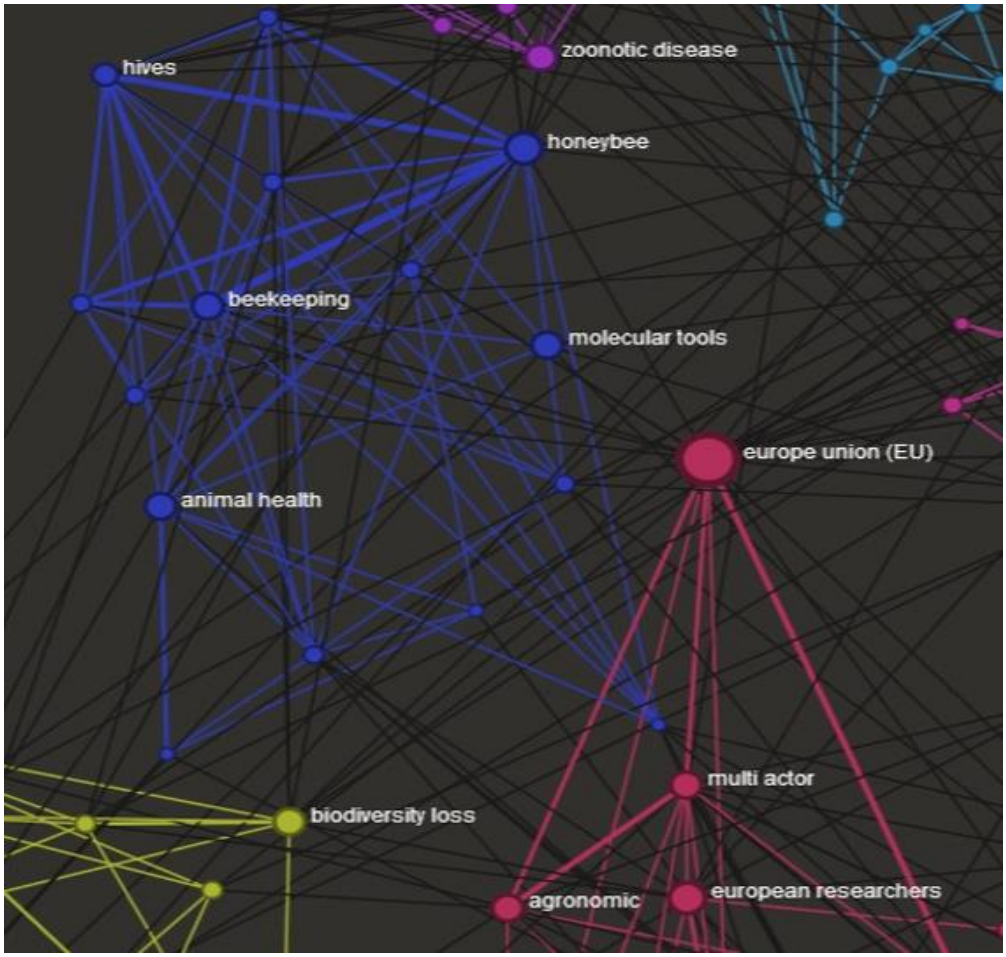
The mapping of collaborations on EU-funded research projects in the period 2007-2021 and related to the cross-policy domain of Biodiversity and Health revealed a doubling in the number of projects and the frequency of the UK's involvement over the past 15 years. The projects were evenly distributed across institutions, with none emerging as being particularly dominant and no frequent collaboration axes.

Over the period 2007-2021, the JRC did not show a clear research focus on studying the link between biodiversity and health. Interestingly, thematic clusters covered two topics of the One Health approach - animal health and environmental health in the context of sustainable development and biodiversity

conservation - but there were no keywords or explicit links to human health except in the case of the zoonosis cluster (Figure 3).

The mental health keyword did not appear in the automatic clustering process, indicating that its link with nature may not be explicitly developed in research on green infrastructure and biodiversity. The main clusters formed around honeybees and their population decline, the health of ecosystems linked to their integrity and ecological processes, and sustainable use, linked to natural ecosystems and plant health. There were no keywords that enabled the linking of impacts through the trophic chain, for example on the effects of chemical pollution and toxicity on the environment and the threat to human health. This has been developed at the JRC only recently.

Figure 3. Thematic networks of keywords automatically retrieved from CORDIS metadata files of EU-funded research projects and relevant to the Biodiversity and Health topic (only some keywords are shown and the most frequent keywords are labelled). Each network is assigned a colour and includes the keywords that most frequently occur together.



Source: JRC, 2022[4].

Critical aspects and looking ahead

From the methodological point of view, the workflow that was developed alternates automated and manual processes, and can be applied to any topic and policies of interest. The TIM Open Access software has the advantage of being freely accessible to non-specialist audiences; it does not require

any advanced technical skills or expertise in machine learning and text mining algorithms. It enables the listing of relevant projects, and break-down analysis per country and organisation (similar to other tools such as the recently developed Horizon Dashboard: (https://ec.europa.eu/info/funding-tenders/opportunities/portal/screen/opportunities/horizon-dashboard).

The **real value-added of TIM lies in the policy relevant insights it can provide for project results**, notably by enabling targeted and detailed policy relevant searches, by in depth text mining and the automated extraction of keywords within a project's metadata and network analysis for partnerships and thematic content.

One crucial aspect of the methodology is related to the potential bias introduced by expert judgement, notably in the formulation of the best-performing search string to translate policy documents into keywords and in the manual quality checks. Another critical point is the underperformance of the automated extraction process which discarded about 15% of projects. This is mainly due to the lack of precision, focus and/or reliability of the metadata associated with projects that are archived in the CORDIS database.

Our recommendations would first go to the scientists involved in EU-funded research projects, particularly to improve the preciseness and quality of their reporting with regard to the policy relevance of their projects (e.g. in the projects' metadata files). EC policymakers in research and innovation could provide criteria and guidelines for the upcoming EU research projects funded under the Horizon Europe Framework Programme (2021-2027). Guidelines could be provided for writing fit-for-purpose abstracts, ensuring real and precise objectives and actions, precisely categorising of the research project with carefully selected keywords and explicitly identifying the science-to-policy relevance. Improved metadata would also improve TIM's automated project extraction, thus minimising the need for human intervention in manual quality checks.

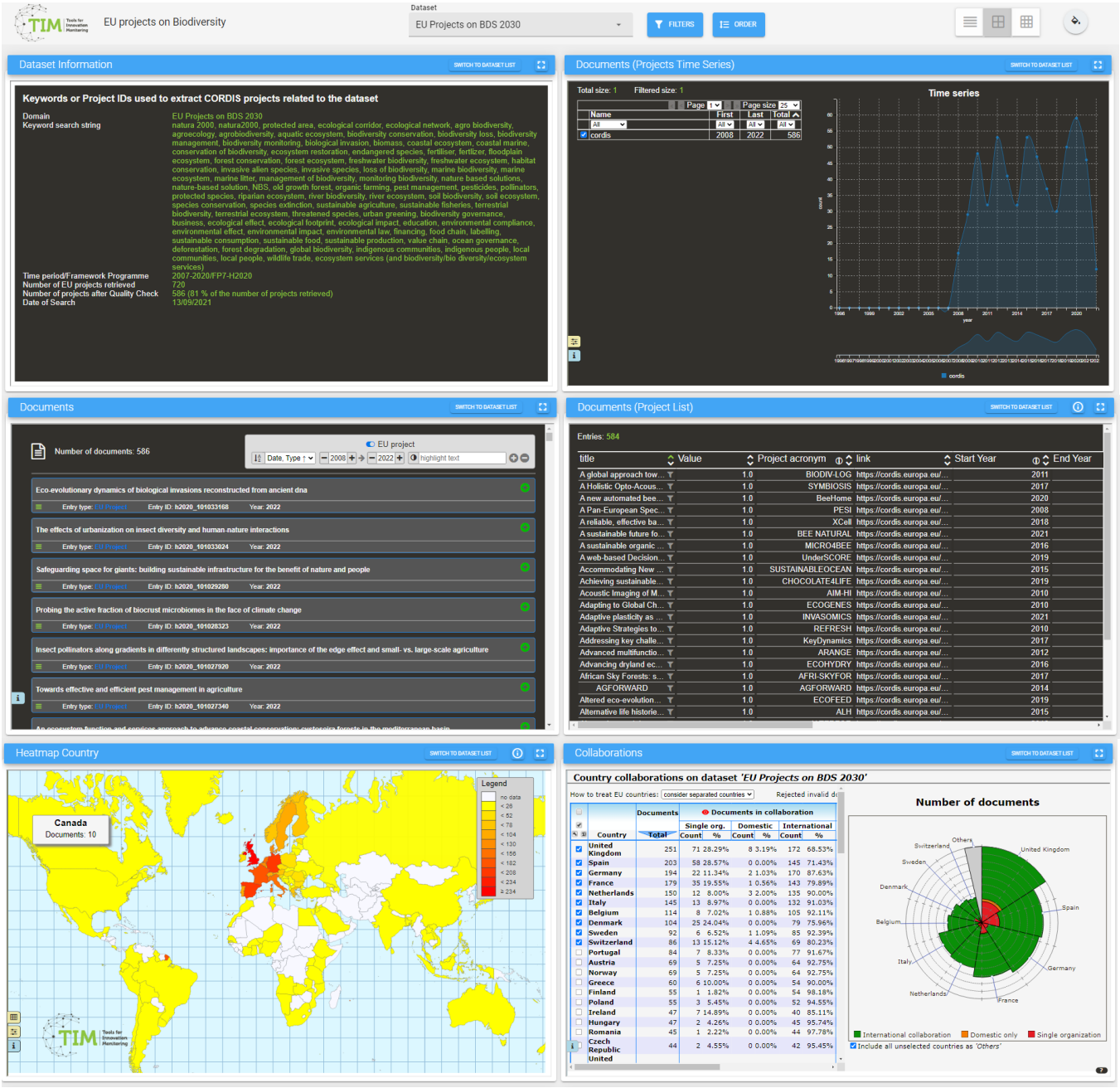
In the context of the Knowledge Centre for Biodiversity, analyses could be conducted on the relevance of EU-funded research projects to other cross-sectoral topics, e.g. Biodiversity and Agriculture, Biodiversity and Trade. In addition, to answer policy priorities of the Biodiversity Strategy for 2030, other conditional queries could be defined to retrieve EU-funded research projects that respond to specific biodiversity actions and commitments, such as the restoration of free-flowing rivers and the reduction of 50% of chemical pesticides.

The results of this study can be viewed as an interactive TIM dashboard on the *EU research project - policy link explorer*[5] page of the Knowledge Centre for Biodiversity website (Figure 4). The dashboard provides access to the separate datasets that help contextualise and identify links between EU-funded research projects and biodiversity policy priorities.

This work was conducted in close collaboration with the JRC's Competence Centre on Text Mining and Analysis. For further details, please download the report Parracciani, C., Ganisheva, K., Ventocilla Jaramillo, J.L., and Estreguil, C., *Mapping interlinkages between EU-funded research projects and EU policy priorities. Application of the TIM open access tool to case studies in the context of the EU Biodiversity Strategy for 2030*, EUR 31106 EN, Publications Office of the European Union, Luxembourg, 2022, ISBN 978-92-76-53299-6, doi:10.2760/37559, JRC127579.

[5] https://knowledge4policy.ec.europa.eu/biodiversity/topic/eu-research-project-policy-link-explorer_en

Figure 4. TIM dashboard showing the "EU Projects on BDS 2030" dataset to help contextualize the EU-funded research projects in relation to the Biodiversity Strategy for 2030 (accessible from the EU research project - policy link explorer section of the KCBD website).



Source: JRC, 2022[4].

JRC128604