fteval JOURNAL
for Research and
Technology Policy Evaluation

July 2025, Vol. 57, pp. e5, 1-19
DOI: 10.22163/fteval.2025.697

ADVANCING RESEARCH IMPACT EVALUATION IN THE DIGITAL ERA:

INSIGHTS FROM EU-FUNDED RARE DISEASE PROJECTS

IOANNA GRYPARI, SERGIO DI VIRGILIO, HARIS PAPAGEORGIOU, ARIS FERGADIS AND DIMITRIS PAPPAS
DOI: 10.22163/FTEVAL.2025.697

ABSTRACT

This study presents a data-driven methodology for evaluating the impact of publicly funded research, addressing the growing complexity of research and innovation landscapes. By integrating diverse data sources (including publications, clinical trials, and company websites) and leveraging advanced analytics such as natural language processing (NLP) and deep learning workflows, this approach overcomes traditional limitations in research impact evaluation. A case study on rare diseases demonstrates how the methodology uncovers pathways linking research outputs to societal benefits while balancing automation with expert validation to ensure accuracy and relevance. These findings underscore the strategic importance of robust, data-driven insights for aligning research priorities with evolving societal imperatives.

Keywords: Research Impact Evaluation, Rare Diseases, Natural Language Processing, Publicly Funded Research, Horizon 2020

1. INTRODUCTION

Evaluating the societal impact of publicly funded research is a significant challenge, often constrained by extended timelines and the complex, interconnected nature of research landscapes. Research outcomes frequently require years or even decades to translate into societal or economic benefits, involving diverse actors, disciplines, and outputs. This misalignment between the extended timelines of research impact and the shorter cycles of policy evaluation underscores the need for innovative approaches. Traditional evaluation methods, reliant on structured data and statistical indicators, provide a baseline understanding but often fail to capture the intricate pathways through which research drives societal change. For instance, foundational knowledge from a project may indirectly influence innovations years later, connections that are difficult to trace without advanced tools.

To address these challenges, this study presents a methodology that integrates diverse data sources with advanced artificial intelligence (AI) approaches, including text mining, natural language processing (NLP), and machine learning (ML). By linking datasets such as projects, publications, and corporate activities, the methodology uncovers connections between research outputs and societal impacts, a key for evidence-based policymaking. Combining scalable automation with expert human oversight, it ensures both accuracy and contextual relevance, adapting to the complexity and diversity of modern research. Expert-informed interpretation ensures that subtle or long-tail pathways of influence are recognised.

The methodology is demonstrated through a case study on rare diseases, a domain of significant societal importance that exemplifies the need for collaborative and long-term research efforts. Rare diseases, while individually uncommon, collectively affect from 27 to 36 million people in the European Union.[1] Between 2014 and 2020, the EU invested more than €2.9 billion in over 600 rare disease research and innovation (R&I) projects under FP7 and Horizon 2020.[2] Despite this substantial investment, understanding how research outputs translate into tangible societal benefits remains a critical challenge.

1 https://research-and-innovation.ec.europa.eu/research-area/health/rare-diseases_en

2 European Commission, 2021

A central component of the methodology is the use of big data analytics to process vast amounts of structured and unstructured information. Recent advancements[3] in tools such as knowledge graphs and ensemble algorithms enable the extraction of meaningful insights from diverse datasets, mapping the lifecycle of research activities and providing policymakers with actionable intelligence for targeted interventions. Moreover, the inclusion of innovative indicators, which extend beyond traditional metrics, facilitates a richer understanding of research impact. These indicators can capture contextual dimensions, such as alignment with global priorities like the United Nations' Sustainable Development Goals (SDGs)[4] or relevance to specific health challenges.

While this approach represents a step forward in research impact evaluation, it is not without limitations. Indirect and nuanced pathways often require qualitative insights that cannot be fully automated, underscoring the importance of expert validation to ensure analytical robustness. Moreover, the interconnected nature of research landscapes introduces additional complexity: multiple developments frequently occur simultaneously, influenced by diverse actors, external events, and evolving societal needs. Even with advanced tools and methodologies, it is often impossible to definitively attribute specific outcomes to individual projects or interventions. This highlights the need for cautious interpretation and an appreciation of the dynamic, multifaceted nature of research impact. Even if definitive attribution is elusive, partial, or probabilistic, insights are invaluable for shaping policy decisions.

This paper is organised as follows: Section 2 details the methodology underpinning the evaluation framework, Section 3 presents the rare diseases case study, Section 4 presents the results, and Section 5 concludes with a discussion of the broader implications of this methodology for research assessment and evidence-based policymaking.

3 European Commission, 2023a

4 https://sdgs.un.org/goals

2. METHODOLOGY IN BRIEF

This study employs a multifaceted methodology for evaluating R&I activities, bringing together AI techniques and domain expertise. While this study applies the framework to rare diseases, the methodology is designed to be research-theme agnostic and can be adapted to various fields, including energy, climate, and digital technologies. The approach builds on prior work conducted in IntelComp[5] and Data4Impact,[6] which explored AI-driven frameworks for assessing research impact through large-scale data integration and advanced analytics.[7] While this section provides an overview, a forthcoming technical paper will elaborate on the specific workflows and computational models in greater detail.

The framework is designed to accommodate large-scale and heterogeneous data sources, producing policy-relevant indicators while maintaining robust oversight through expert validation. By blending automated and human-driven processes, the approach aims to strike a balance between scalability and interpretative accuracy.

GUIDING PRINCIPLES

Several guiding principles shape this methodology. First, it adopts a 360° view of data, integrating a broad range of R&I information, including publications, patents, industry records, and policy documents, among others. This holistic perspective is enriched by standardised frameworks, such as the SDGs and the International Classification of Diseases (ICD),[8] which situate research outputs within broader societal and policy contexts. Second, the workflow is modular and end-to-end, covering data cleaning, information extraction, integration, and final analysis. Third, it embodies the expert-in-the-loop paradigm, recognising that AI-generated outputs require human review and domain contextualisation to ensure validity and alignment with policy objectives. Finally, openness and transparency guide all activities, from data handling (e.g. adherence to FAIR principles) to methodological documentation, fostering trust and replicability.

5 Horizon 2020 project, with grant ID 101004870, https://cordis.europa.eu/project/id/101004870

6 Horizon 2020 project with grand ID 770531, https://cordis.europa.eu/project/id/770531

7 Grypari et al., 2020; Stanciauskas et al., 2020

8 https://www.who.int/standards/classifications/classification-of-diseases

DATA SOURCES AND PREPARATION

A core strength of the methodology lies in its capacity to merge structured and unstructured data reflecting multiple stages of the research lifecycle. Project databases offer foundational information on objectives, consortium structures, and funding levels. Scientific outputs, particularly publications, serve as an initial measure of research activity and dissemination. Patents, clinical trials, and other innovation-related data provide indicators of technology transfer and translational progress, whereas industry data, such as company websites, illuminate commercialisation pathways. The framework incorporates broader societal elements, such ESG metrics, regulations, policies, and human resources (skills demanded vs supplied), to create a comprehensive view of how research may impact economic, environmental, and societal imperatives. Finally, ontologies and standards, including the ICD and SDGs, facilitate semantic enrichment and consistent categorisation.

DATA PREPARATION

Each dataset undergoes a comprehensive cleaning, disambiguation, and deduplication process, removing inconsistencies and redundancies. Structured metadata, such as project IDs and publication DOIs,[9] are reconciled with unstructured content (e.g. abstracts, company websites) to form a unified database. In many cases, semantic linking is applied, mapping disease mentions or similar references to standardised terminologies (e.g. ICD codes). This ensures that subsequent analyses operate on harmonised, context-rich data.

EXTRACTION

Building on this curated dataset, machine learning (ML) and natural language processing (NLP) techniques extract and categorize relevant entities. Named Entity Recognition (NER) models identify diseases, technologies, and other key entities, while topic modelling detects thematic structures and helps capture how research priorities evolve over time.

To better capture relationships between research outputs, we apply semantic similarity analysis to identify connections between different publications, projects, patents and so on. Additionally, co-occurrence analysis helps detect recurring associations between key terms, providing insight into emerging research directions. These extracted entities and relationships are structured

9 https://www.doi.org/the-identifier/what-is-a-doi/

into knowledge graphs, which link research topics to relevant stakeholders, funding programs, and translational applications such as clinical trials and industrial uptake.[10] By organizing research impact pathways in this structured manner, the methodology enables downstream analysis to assess how publicly funded research contributes to long-term innovation and societal benefits.

INTEGRATION

After extraction, the framework integrates these varied data streams to illuminate broader connections. ML classifiers categorise research outputs according to established taxonomies (e.g. SDGs) to ensure alignment with recognised global priorities. In parallel, impact pathway analysis uses ML models to trace how early-stage findings (e.g. publications) transition into tangible applications (such as clinical trials, patents, or commercial products). Through this process, the methodology generates novel metrics – for example, gauging how far an idea has progressed from fundamental research to real-world implementation. As the database grows in size and quality, the framework's analytical precision improves, allowing for more reliable impact assessments across disciplines.

SYNTHESIS

In the final stage, inference methods evaluate the R&I ecosystem from a holistic standpoint. Citation networks depict the longevity and influence of foundational work, highlighting how discoveries spread across disciplines. Industry uptake scores measure how publicly funded research permeates ongoing industrial R&D, indicating potential commercialisation pathways. Beyond standard indicators like citation counts, contextualised measures (e.g. thematic momentum) provide a more nuanced understanding of research impact, one that is directly relevant to policymakers responsible for guiding future funding and innovation strategies.

TECHNOLOGICAL INFRASTRUCTURE

This methodological design operates within a cloud-native, modular architecture capable of supporting computationally demanding NLP and ML workflows:

10 In this paper, 'knowledge' refers to structured information about research outputs and their interconnections, derived from multiple data sources (e.g., publications, patents, clinical trials, and company websites).

- High-Performance Computing (HPC) environments enable large-scale data ingestion and batch-processing tasks, ensuring the timely analysis of extensive, heterogeneous datasets.
- Containerisation (e.g. Docker) and continuous integration/continuous deployment (CI/CD) pipelines facilitate rapid, iterative model development, allowing the framework to evolve in tandem with emerging analytical tools.
- Microservices and distributed infrastructure provide scalability, adaptability, and efficient resource utilisation, making it feasible to integrate additional modules (e.g. new entity classes or ontologies) without disrupting the overall pipeline.

By integrating diverse data sources, applying advanced NLP and ML techniques, and embedding expert validation throughout, this methodology presents a transparent and adaptable framework for R&I evaluation.

3. RARE DISEASES AS A CASE STUDY

Rare diseases (RD) pose a pressing public health challenge in the European Union (EU). They are defined as affecting no more than one person in every 2,000 individuals. Collectively, however, these conditions impact approximately 36 million people across the EU, and encompass 6,000 to 8,000 distinct disorders.[11] They are frequently characterised by high unmet medical needs, significant variability in clinical presentations, and limited treatment options, necessitating substantial collaboration at both European and international levels. Such collaboration draws on diverse expertise, from clinical practice to biotechnology and health policy, underscoring the inherent complexity of rare diseases and the need for robust, cross-sectoral approaches. Recognising the societal and economic implications of rare diseases, the EU has made them a key focus of research and innovation activities. Approximately 70% of these disorders manifest in childhood, often leading to diagnostic delays, challenging care requirements, and profound long-term impacts on patients and families.

11 European Commission, 2021

TAILORING THE METHODOLOGY TO RARE DISEASES

While the methodology outlined in Section 2 is broad enough to evaluate large-scale R&I activities across disciplines, our use case on rare diseases serves as a focused illustration. The framework itself is research-theme agnostic and can be applied to other domains. Rather than attempting a fully comprehensive analysis, we have selectively applied certain data sources to highlight how the framework can be adapted to specific domains. This narrower scope underscores its flexibility and capacity to generate meaningful insights across varying scales. A central task for the case study is establishing a rare disease project portfolio that adequately represents both the breadth of EU investments and the depth of targeted research activities. To achieve this, a multi-layered approach was adopted:

1. **Extended Portfolio**: Natural Language Processing (NLP) and probabilistic models were used to scan a wide range of EU-funded projects, capturing direct and indirect references to rare diseases (about 10,000 projects).

2. **Core Portfolio**: From this broad set, additional filters were introduced to isolate projects explicitly addressing rare disease topics, ensuring higher specificity (about 1,700 projects).

3. **Curated Portfolio**: Finally, manual review by domain experts confirmed a subset of projects with a primary and direct focus on rare diseases (about 400 projects).

By combining automation with expert input, this three-tiered structure enables flexible analyses: one can examine thematic diversity in the extended dataset while zeroing in on more specialised findings in the curated list. While expert validation was applied here in the context of rare diseases, this approach is adaptable to other fields by incorporating domain-specific expertise at key validation stages, ensuring accuracy and relevance regardless of the research area.

To navigate the multifaceted nature of rare disease research, the study relies on an array of NLP and graph-based approaches. For instance, entity extraction and topic modelling identify critical diseases, and thematic clusters

within project documentation[12], publications,[13] clinical trials,[14] and company websites. Citation graphs and knowledge graphs trace how research outputs are interlinked and how these relationships evolve over time. These techniques collectively address two prominent challenges: the fragmentation of data across scattered sources and the difficulty of following a project's influence through multiple, often indirect, pathways.

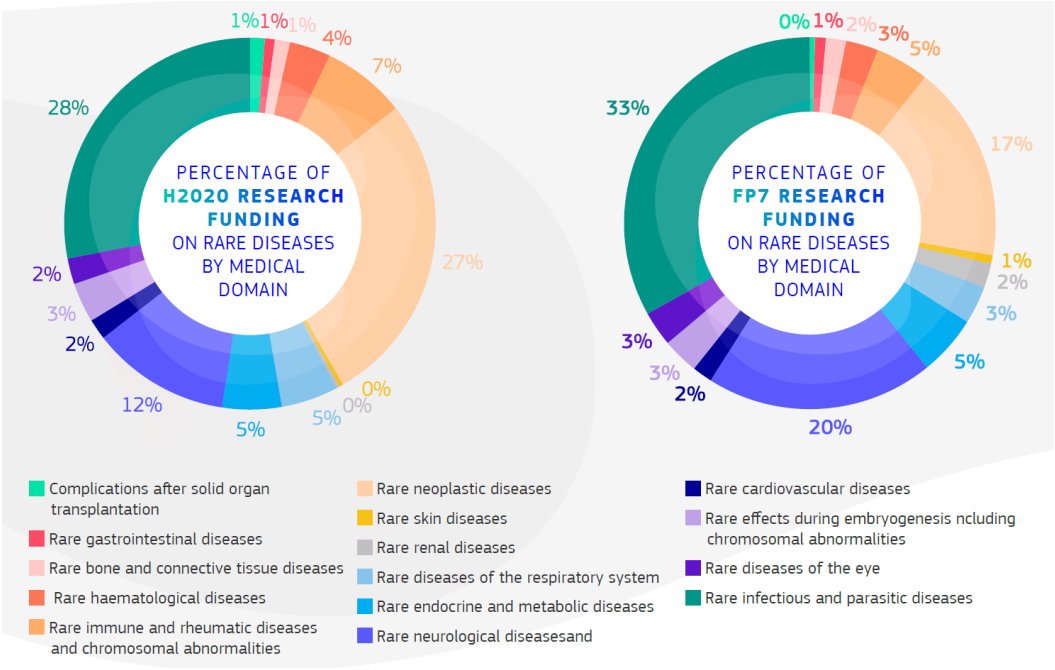

Figure 1: Distribution of Rare Disease Funding Across ICD-10. Source: European Commission, Directorate-General for Research and Innovation, Collaboration: a key to unlock the challenges of rare diseases research, Publications Office, 2021

A summary published by the European Commission (Figure 1) illustrates the percentage of rare disease funding allocated to different ICD-10[15] categories under FP7 and Horizon 2020. Compared to FP7, there appears to be a relative increase in funding for nervous system disorders and congenital anomalies under Horizon 2020. These shifts in emphasis offer a preliminary snapshot of how EU research priorities in rare diseases evolved between the two programmes.

12 Data from CORDIS https://cordis.europa.eu/

13 Data from the OpenAIRE Graph https://graph.openaire.eu/

14 Data from PubMed https://pubmed.ncbi.nlm.nih.gov/, and ClinicalTrials.gov https://clinicaltrials.gov/

15 https://icd.who.int/browse10/2019/en

Results and insights drawn from applying our adapted methodology to the curated project portfolio will explore patterns in greater depth, examining how the thematic distribution, collaboration networks, industrial uptake, and clinical links collectively shape the rare disease research ecosystem.

4. RESULTS & INSIGHTS

This section presents the principal findings obtained by applying the data-driven framework (described in Section 2) to our curated portfolio of rare disease projects under FP7 and Horizon 2020.

EMERGING THEMES IN RARE DISEASE RESEARCH

We begin by applying topic modelling to the full texts of project descriptions and scientific publications. This unsupervised approach clusters thematically related documents and labels them according to human expert input, enabling us to identify which disease areas or research themes gained (or lost) prominence between FP7 and Horizon 2020.

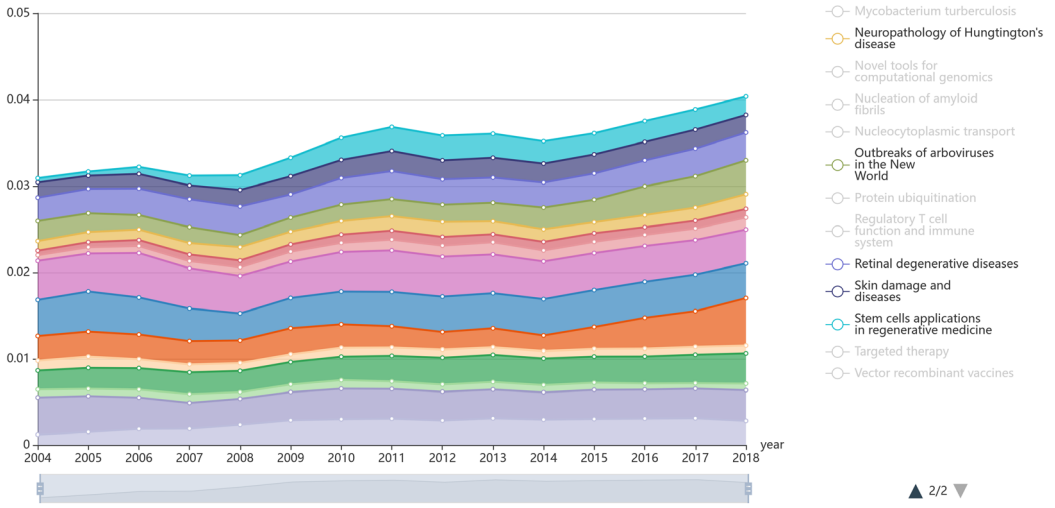

Figure 2: Evolution of Select Topics in Rare Disease Portfolio Over Time. Sources: OpenAIRE Graph, CORDIS.

Figure 2 displays examples of topic evolution across time. Notably, the 'Outbreaks of Arboviruses in the New World' topic rises markedly under Horizon 2020, coinciding with heightened global concerns over Zika and dengue, which have been particularly prominent in Latin America. In contrast, while malaria remains one of the most EU-funded research areas, its topic

momentum in H2020 is lower—despite its persistent burden, especially in Sub-Saharan Africa.

It is important to interpret these results in context. The increased focus on arboviruses can be seen as an illustration of the EC's capacity to address urgent crises. However, it does not necessarily imply diminished attention to other high-burden diseases, including malaria and leishmaniasis. In several calls, the natural draw of scientists and public health actors specialising in arboviruses, many located or collaborating in Latin America, led to a proportional rise in projects on these topics. As shown and explained in Figure 4 below, the EC's investments in both arboviral and malaria/leishmaniasis research still outpaces the broader health field's average proportion. Nonetheless, the data highlight how rapid shifts in global health needs can shape which research themes gain traction in any given funding cycle.

COLLABORATION PATTERNS AND REGIONAL DISPARITIES

To examine cross-organisational partnerships and co-publications, we constructed graphs linking projects, participating organisations, and the resulting publications. The visualisations below offer an "user friendly" birds' eye perspective on how project participants and coauthors collaborated over time in rare disease projects.

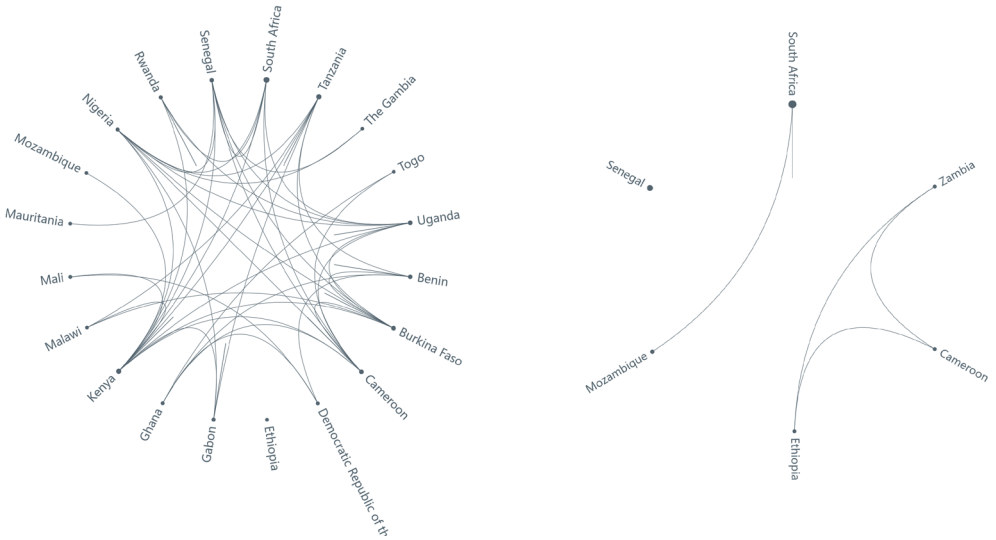

Figure 3: Project & Publications Collaborations in FP7 (left) and H2020 (right) in Sub-Saharan Africa. Sources: OpenAIRE Graph, CORDIS.

Figure 3 illustrates the collaboration network for sub-Saharan Africa organisations, indicating that African organisations formed stronger co-participation ties under FP7. During Horizon 2020, the data show a pronounced surge in collaborations between Latin American partners (not shown above), consistent with the rise of arbovirus-related topics. Crucially, this reorientation does not necessarily mean sub-Saharan Africa received fewer resources in absolute terms; rather, the number of joint publications and grants involving African partners decreased.

The figure below uses each topic's share of publications as a proxy measure of resource allocation and research focus. While this metric helps illustrate relative emphasis, it does not represent a precise accounting of budgets or project-level expenditure. Nonetheless, it supports the broader finding that the EC-funded topics depicted often exceed the global field's average proportion, indicating a deliberate policy to address areas requiring public-sector intervention. As shown, both "Outbreaks of Arboviruses in the New World" and "Malaria and Leishmaniasis" command disproportionate investment, reflecting the EC's strategic emphasis on these research challenges.

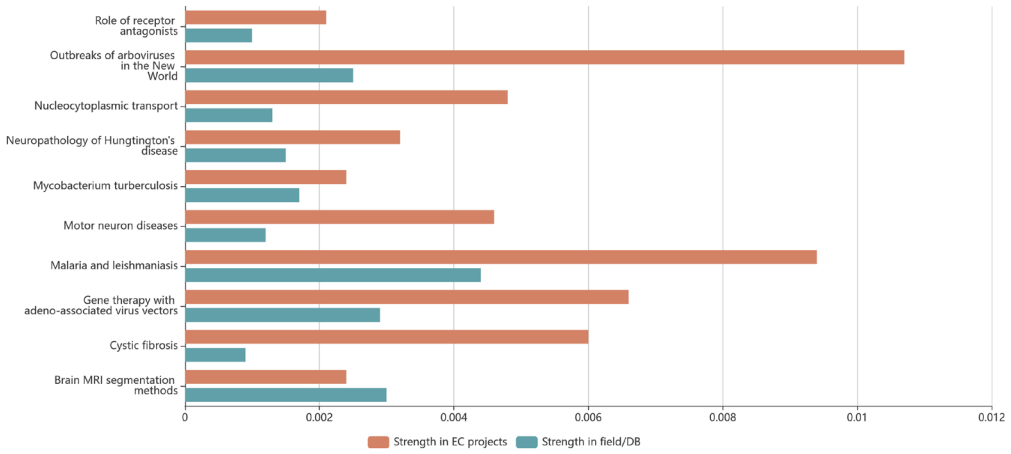

Figure 4: EC vs the Field, Investment in Topics as a Share of Total Publications.
Sources: OpenAIRE Graph, CORDIS

INDUSTRIAL UPTAKE AND R&D CONTINUITY

A major objective of EU-funded research is to stimulate industrial innovation. To assess whether companies continued working on project-related topics, we applied a deep learning method to analyse text from company websites, measuring semantic similarities with each company's past project deliverables. This yielded an R&D Uptake Score, which quantifies how closely a firm's current activities resemble its earlier, publicly funded research.

In Figure 5, companies closer to the outer edge of the circle (score ~1) exhibit strong continuity between their present-day research and the innovations they pursued under EC projects, whereas those nearer to the centre (score ~0) appear to have shifted focus. A low score does not imply low impact; firms may pivot strategically in response to market signals or integrate project methodologies into different domains. Conversely, a high score suggests thematic consistency but does not guarantee successful product development. Additional data, such as patent portfolios, licensing records, or clinical trial sponsorship, would further enrich assessments of how public investments translate into commercial outcomes.

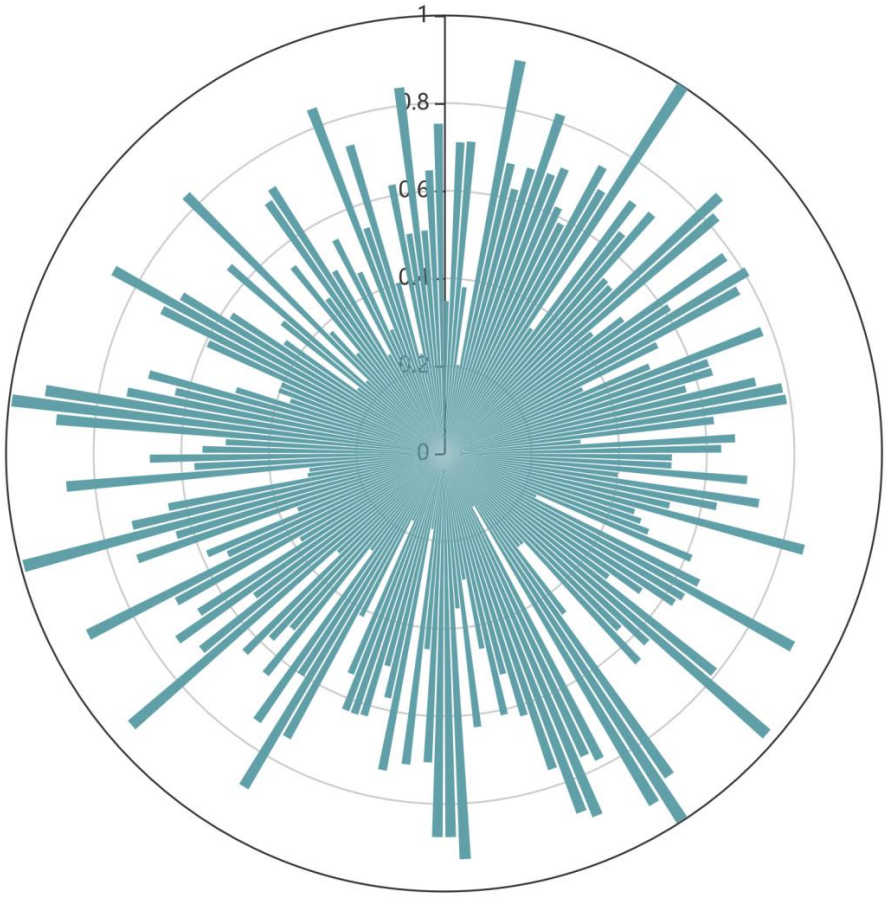

Figure 5: Company Uptake Score for Rare Disease Portfolio. Sources: OpenAIRE Graph, CORDIS, Company Websites. Each column represents a company, with column height reflecting the uptake score, a measure of how closely a firm's current research aligns with its past EC-funded projects. The score is calculated using AI-driven text similarity analysis, comparing company website content with project publications. Higher scores (taller columns) indicate continued work in the same field, while lower scores suggest a shift in focus.

LINKING TO HEALTH OUTCOMES VIA CLINICAL TRIALS AND GUIDELINES

A final cornerstone of our assessment tracks whether project outputs feed into clinical trials, a critical juncture between scientific discovery and patient outcomes. Our analysis revealed over 1,800 trials citing publications linked to the rare disease portfolio, including 843 trials in which at least one original project participant was directly involved. This engagement demonstrates the continuity from research funding to the clinical testing phase.

Nevertheless, progress from publication to improved patient care can be slow. 100 clinical guidelines were found to reference the curated projects' publications, and in six cases, the guidelines included a direct mention of the project in their metadata. Some guidelines emerged more than three years after the project's end date, illustrating the iterative and often protracted path from funded research to real-world application. Even "unsuccessful" trials can shape best practices or refine methodological approaches, serving as stepping stones toward future breakthroughs.

To capture these indirect but critical contributions, additional contextual indicators could offer a more complete understanding of how public research investments unfold in clinical practice. As the use case below demonstrates, interpreting the full chain of evidence demands careful triangulation among multiple data sources.

A CLINICAL TRIAL USE CASE: ANALYSING PATHWAYS TO SOCIETAL IMPACT

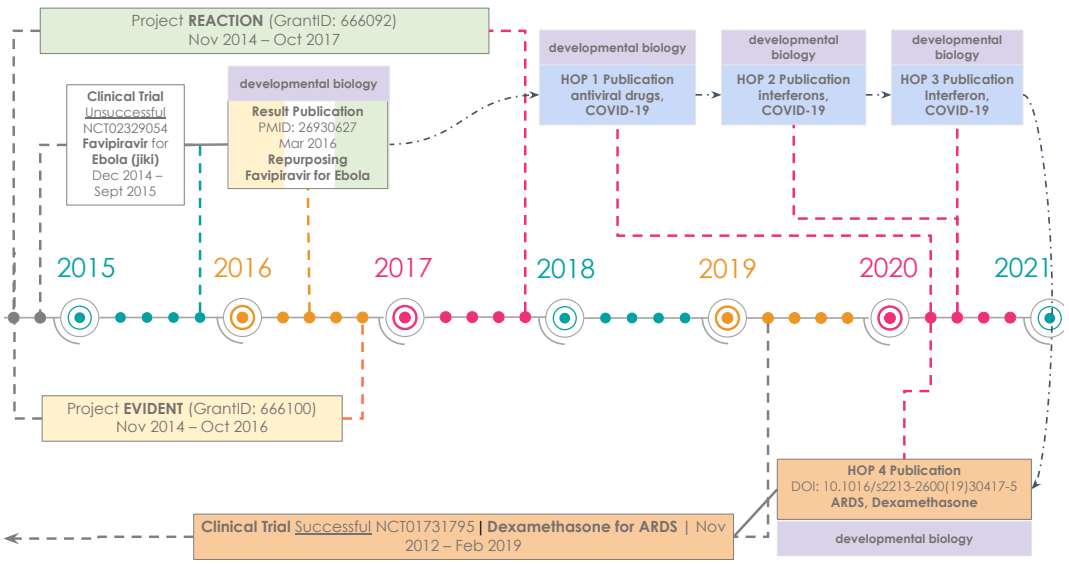

Figure 6: Flowchart of a Clinical Trial Use Case

In the domain of health research, clinical trials act as a pivotal bridge between laboratory innovations and tangible societal benefits. A compelling illustration of these research-to-impact pathways arises from two Horizon 2020–funded projects, REACTION and EVIDENT, which jointly produced a publication (PMID: 26930627) summarising the outcomes of the "jiki" trial (NCT02329054). This trial investigated the repurposing of Favipiravir for Ebola treatment and ended unsuccessfully at Phase 2. However, advanced citation analysis revealed a four-hop linkage connecting the "jiki" trial to a successful clinical trial (NCT01731795) on the use of Dexamethasone for Acute Respiratory Distress Syndrome (ARDS), as depicted in the flowchart above.

This multi-hop chain illustrates the cumulative nature of scientific discovery: even halted or "unsuccessful" trials can contribute knowledge that informs later breakthroughs. The "jiki" trial was cited in 190 subsequent publications, nine of which were associated with other trial efforts that themselves ended prior to commercial success. Techniques such as semantic similarity analysis, and topic evolution tracking helped uncover these indirect pathways. Yet, human expertise remains essential for validating weak or ambiguous connections, ensuring contextual accuracy, and mitigating the risk of over-attribution.

Although the path from Ebola research to an ARDS breakthrough might appear tenuous, tracing through multiple layers of citations and knowledge diffusion, the example underscores the incremental, interwoven nature of health research. While direct attribution is difficult, the original Ebola trial contributed to a growing body of knowledge, influencing subsequent studies that may have played a role in shaping later breakthroughs.

5. CONCLUSION

This study demonstrates how a data-driven, AI-augmented methodology can illuminate the often intricate and indirect pathways that link publicly funded research to societal outcomes. By integrating large-scale datasets, advanced NLP and ML tools, and graph-based analyses, the approach uncovers patterns and relationships that traditional methods might overlook. However, our findings also highlight the inherent complexity of attributing research impacts, given the diversity of actors, the interwoven nature of developments, and the delayed emergence of tangible benefits.

Despite its strengths, the framework requires careful interpretation. AI-driven analyses may oversimplify complex relationships or introduce spurious

correlations, making expert validation essential to ensure meaningful insights. Additionally, not all types of impact, such as policy influence, are easily captured with structured data alone, highlighting the need for complementary qualitative assessments.

For researchers and policymakers looking to apply this framework in other disciplines, its scalability and adaptability are key advantages, but expert knowledge remains critical for ensuring results are contextually valid. In fields with fewer structured indicators, AI's role may shift from directly identifying patterns to helping generate hypotheses, requiring an iterative process between automated insights and expert interpretation. Ensuring reliability in AI-driven findings demands cross-validation across multiple data sources, transparency in methodological assumptions, and active monitoring for biases in both data and model design.

From a policymaking perspective, this synergy between AI-driven analytics and expert validation provides a powerful tool for evidence-based decision-making. This approach enables stakeholders to better allocate resources, support high-impact collaborations, and track emerging research priorities, while remaining aware of the limitations of purely algorithmic methods. Finally, new impact metrics generated through this methodology help address gaps in traditional assessment frameworks, but their value depends on continuous refinement, interdisciplinary validation, and engagement with the broader research community. By maintaining a balance between advanced analytics and expert oversight, research investments can better align with societal goals, maximizing their long-term impact.

REFERENCES

Adam, G. P., Pappas, D., Papageorgiou, H., Evangelou, E., & Trikalinos, T. A. (2022). A novel tool that allows interactive screening of PubMed citations showed promise for the semi-automation of identification of Biomedical Literature. Journal of Clinical Epidemiology, 150, 63–71. https://doi.org/10.1016/j.jclinepi.2022.06.007

European Commission: Directorate-General for Research and Innovation, Collaboration – A key to unlock the challenges of rare diseases research – February 2025, Publications Office of the European Union, 2025. https://data.europa.eu/doi/10.2777/8029727

European Commission: Directorate-General for Research and Innovation, Collaboration – A key to unlock the challenges of rare diseases research, Publications Office, 2021. https://data.europa.eu/doi/10.2777/249334

European Commission: Directorate-General for Research and Innovation and Denham, S., Tracking of research results – Specific contract No 2018/RTD/TRR003 – Final data and data sources report, Denham, S. (editor), Publications Office of the European Union, 2023. https://data.europa.eu/doi/10.2777/57692

European Commission: Directorate-General for Research and Innovation and Denham, S., Tracking of research results – Specific contract No 2018/RTD/TRR003 – TRR final report, Denham, S. (editor), Publications Office of the European Union, 2023. https://data.europa.eu/doi/10.2777/128148

Fergadis, A., Pappas, D., Karamolegkou, A., & Papageorgiou, H. (2021, November). Argumentation mining in scientific literature for sustainable development. In Proceedings of the 8th Workshop on Argument Mining (pp. 100-111). https://doi.org/10.18653/v1/2021.argmining-1.10

Grypari, I., Pappas, D., Manola, N., & Papageorgiou, H. (2020, May). Research & Innovation Activities' Impact Assessment: The Data4Impact System. In Proceedings of the 1st Workshop on Language Technologies for Government and Public Administration (LT4Gov) (pp. 22–27).

Kotitsas, S., Pappas, D., Manola, N., & Papageorgiou, H. (2023). SCINOBO: a novel system classifying scholarly communication in a dynamically constructed hierarchical Field-of-Science taxonomy. Frontiers in Research Metrics and Analytics, 8, 114983. https://doi.org/10.3389/frma.2023.1149834

Saeidnia, H. R., Hosseini, E., Abdoli, S., & Ausloos, M. (2024). Unleashing the power of AI: a systematic review of cutting-edge techniques in AI-enhanced scientometrics, webometrics and bibliometrics. Library Hi Tech. https://doi.org/10.1108/LHT-10-2023-0514

Stanciauskas, V., Grypari, I., Nelhans, G., Papageorgiou, G., & Demiros, I. (2020). Policy report on new indicators and approaches for assessing the societal impact of research and innovation activities: Big Data approaches for improved monitoring of re-search and innovation performance and assessment of the societal impact in the Health, Demographic Change and Wellbeing Societal Challenge.

Stavropoulos, P., Lyris, I., Manola, N., Grypari, I., & Papageorgiou, H. (2023, December). Empowering Knowledge Discovery from Scientific Literature: A novel approach to Research Artifact Analysis. In Proceedings of the 3rd Workshop for Natural Language Processing Open Source Software (NLP-OSS 2023) (pp. 37–53). https://doi.org/10.18653/v1/2023.nlposs-1.5

ACKNOWLEDGEMENT

This work was partially funded by the European Union through the Horizon 2020 projects Data4Impact (Grant Agreement ID: 77053) and IntelComp (Grant Agreement ID: 101004870). The views and opinions expressed herein are solely those of the authors and do not necessarily reflect those of the European Union or the relevant granting authorities. Neither the European Union nor the granting authorities can be held responsible for them. Special thanks to Christina Kyriakopoulou for her meticulous curation of the list of rare disease projects and other project outputs and to Tsveta Schyns-Liharska (EC Blue Book Trainee) for offering their valuable insights in the earlier stages of this work.

AUTHORS

IOANNA GRYPARI (Corresponding Author)
Email: ioanna.grypari@opix.ai
OPIX P.C., https://www.opix.ai
Athena Research Center, https://www.athenarc.gr/en
OpenAIRE AMKE, https://www.openaire.eu/
ORCID: 0000-0002-7485-1591

SERGIO DI VIRGILIO
DG R&I, European Commission, 1049 Brussels, Belgium
ORCID: 0000-0001-5751-9661

HARIS PAPAGEORGIOU
OPIX P.C., https://www.opix.ai
Athena Research Center, https://www.athenarc.gr/en
ORCID: 0000-0002-7352-2403

ARIS FERGADIS
OPIX P.C., https://www.opix.ai

DIMITRIS PAPPAS
OPIX P.C., https://www.opix.ai
ORCID: 0000-0001-5784-0658