parsuk
portuguese association of researchers and students in the uk

appendices

'PUMP' Mapping Scientific Collaborations Between Portugal and the United Kingdom

2023

contents

A. scoping document

background

PUMP (Portugal – United Kingdom Mapping Project) is a project led by PARSUK with the goal of mapping the recent evolution of scientific collaborations between Portugal and the United Kingdom. It takes into account the priorities of each partner organisation – the Portuguese Foundation for Science and Technology (FCT) and the British Embassy in Lisbon – in the context of the EU Framework Programme for Research and Innovation. Thisprogramme of work emerged from the signature of a protocol with FCT in 2019. The specific timeframe of this project is 6 months, from March 01 to August 31 2022.

The PUMP timeline is divided into four major phases:

» **Phase 1:** Pilot (March)

» **Phase 2:** Development (April - June)

» **Phase 3:** Validation (June - July)

» **Phase 4:** Closure (July - August)

This scoping document outlines our method for data collecting and database assembling in the context of **Phase 1**. However, the bulk of the work will be spent on Phase 2, with the key deliverable of **developing an Excel-based, comprehensive database of scientific collaborations between UK-PT**, which will build the foundation for a final project report.

contents

scope

research question

For the duration of Horizon 2020 (2014-2021), which scientific collaborations have been established between Portugal and the United Kingdom, in the 5 mission areas of Horizon Europe?

timeframe

01/01/2014 - 01/01/2021

This mapping exercise will cover a time period ranging from January 01 2014 to January 01 2021, the duration of Horizon 2020 programme. This time span aims to provide an overview of the evolution of the scientific collaboration landscape over a period of time that preserves coherence in terms of European research goal areas. Posterior timeframes more relevant to other programmes will be left to be explored in future iterations of this mapping exercise.

Our rationale lies in the hypothesis that scientific projects created during the Horizon 2020 mandate stand a greater chance of converging on their highlighted research areas in favour of greater European relevance (as a partner, which may also provide more fertile ground for international collaborations). Additionally, the Horizon 2020 programme had the United Kingdom as one of its top performing countries. The UK's 'university researchers participated in and led more Horizon 2020 projects than any other participant country, and the UK had one of the programme's highest funding bids success rates (15.0%)'[1].

areas of research

Given the multitude of thematic areas that could factor into our scope, our Phase 1 for this mapping exercise will be limited to those encompassed by Horizon Europe Programme's **5 missions**:

I. **Adaptation to climate change**

II. **Cancer**

III. **Restore our oceans and waters**

IV. **Climate neutral and smart cities**

V. **Soil deal for Europe**

This selection provides a thematic focus that fits neatly into the broader scope of its predecessor, ensuring a sense of continuity that is congruent to our chosen timeframe, while limiting the breadth of areas to explore. By analysing current European priorities as they were featured during the last European Framework Programme, this mapping exercise aims to gather insights about the past in order to inform present and future directions in research policy and fund management.

Each mission has been analysed and properly accommodated into the scope of our project by reviewing various official documents of the European Commission (see References). This allowed us to better understand the targets and scientific fields involved and, as a result, establish working definitions of the mission areas, list examples of areas of research for each mission and determine keywords aimed at filtering the text in our data in search for relevant results for each mission. Some keywords display a truncation with an apostrophe when they represent word stems that allow searching for several word variations at once.

I. ADAPTATION TO CLIMATE CHANGE[2]

Definition

Anticipation of adverse effects of climate change and action to prevent or minimise their possible damage.

Examples

- **Extreme weather effects**: hydrogeological instability, risk mitigation, flood relief investment programmes, disaster prevention.
- **Infrastructure**: resilient infrastructure, energy efficiency and resilient infrastructures, water and wastewater infrastructure, adaptation of health infrastructures.
- **Food**: resilient agricultural system, sustainable and resilient fishery, soil treatment, crop losses, disease vector, food security.
- **Mobility**: sustainable transport, transition to electric vehicles.
- **Ecosystems**: coastal areas and coastal ecosystems, climate services.
- **Energy transition**: energy source diversification, wind, wave, solar, biomass, biofuels, biogas, and hydrogen.
- **Economy**: sustainability and adaptation of tourism.
- **Societal transformation**: horizontal issues, behaviour changes, mental health.

Keywords

'climate change', 'global warming', 'sustainab', 'extreme weather', 'food security', 'climate service', 'coastal ecosystem', 'renewable energy'.

II. CANCER[3]

Definition

Allowing more people to live without cancer, more cancer patients to be diagnosed earlier, suffer less, and have a better quality of life after treatment.

Examples

- **Innovation in medicine**: gene editing, immune therapies, interplay of immune system and cancer, epigenetics, regenerative medicine, organoids, organ transplantation, 3D modelling in surgery, microbiota, functional (MR) imaging, molecular imaging, theranostics, minimal invasive and robotic surgery, smart medical devices, biohybrids, artificial intelligence, machine learning and augmented reality in medicine, personalised (precision) medicine.
- **Social Innovations**: agriculture and nutrition, general aspects, chemistry and pesticides, organic farming, general nutritional aspects, guidelines, innovations in food, food labelling, taxation of unhealthy foods and drinks (and tobacco).
- **Health Systems**: organisation of EU health systems, expenditure on cancer, cancer burden in the EU, cancer screening, access to cancer diagnosis and care, cross border cancer care, cancer care guidelines.
- **Social developments**: prevention, early detection, survival, cancer plans, patient-centric cancer, quality of life, social cohesion and support, inequity and prosperity, socioeconomic status and cancer control, European code against cancer.
- **Environment**: urban development, built environment and physical activity, built environment and depression, built environment of schools, attractive environments, health equity and built environment, healthy and sustainable cities.

Keywords

'cancer', 'tumour*', 'onco*', 'carcinogen*'.

III. RESTORE OUR OCEANS AND WATERS[4]

Definition	Full recovery and regeneration of European marine and freshwater ecosystems by 2030, including oceans, seas, coastal and inland waters.
Examples	» **Climate-resilient coastlines**: scope sea-level rise, coastal vulnerability, coastal protection. » **Clean water for the blue planet**: clean water cycle, industrial pollutants (nutrients, plastics, pharmaceuticals and pesticides) in European waters. » **Vital aquatic ecosystems**: aquatic biodiversity, sustainable aquatic ecosystem productivity, aquatic ecology. » **Human settlements and operations at sea**: developments in aquaculture, aquaponics, water desalination, energy generation. » **Open digital twin of oceans and waters**: observation and early warning, technological advances to manage data.
Keywords	'climate-resilient coastlines', 'clean water', 'aquatic ecosystems'.

IV. CLIMATE NEUTRAL AND SMART CITIES[5]

Definition	Mitigate cities' greenhouse gas emissions, offset their remaining unavoidable emissions, improve their adaptive capacities to the negative impacts of climate change, use digital and telecommunication technologies to make buildings, infrastructures, and services more efficient and less-resource-intensive.
Examples	» **Decarbonizing the energy grid**: centralised renewables and distributed renewables. » **Optimising energy use in buildings**: new building standards, building envelope retrofits, appliances HVAC, heating and hot water, smart lighting, refurbishment lighting, building automation and controls, smart meters. » **Enabling sustainable mobility & land use**: public transport-oriented mobility, walking and cycling, shared mobility, next generation vehicles (electric, autonomous), commercial freight, dense and mixed land use. » **Improving waste management**: re-use of waste materials. » **Improving urban green & food system**: carbon sequestration in urban green, urban water infrastructure, vertical farming, aquaponic/hydroponic farming, diet. » **Socio-cultural aspects**: changing consumption behaviours, using social engineering and marketing for raising awareness.
Keywords	'climate neutral cities', 'smart cities', 'urban food'.

V. SOIL DEAL FOR EUROPE[6]

Definition	Achieve 'healthy soils' across ecosystems for food, nature and climate.
Examples	» **Land degradation and desertification**: soil management practices. » **Soil organic carbon**: conservation of carbon soils, carbon sequestration, reverse carbon loss, vegetation cover, forests as carbon sinks, biomass production, biodiversity preservation. » **Soil sealing and net land take**: loss of agricultural, forest and other semi-natural and natural land to urban and other artificial land development. » **Soil pollution**: organic farming, integrated crop management system, conservation agriculture, remediated sites, soil pollutants and salts. » **Erosion** » **Soil structure** » **Global footprint**: carbon tax, trade regulations, reduction of carbon footprint.
Keywords	'land degradation', 'desertification', 'erosion', 'carbon soil sequestration', 'soil pollution'.

relevant definitions

Establishing a clear working set of definitions for the different concepts we aim to map and describe is critical to ensure proper understanding of this study.

For the purposes of this project, we define **scientific collaboration** as a joint partnership between researchers and/or research bodies that emerges in the service of an R&D project and manifests through any **publicly available evidence** of its research activity. The collaborations that fall within the scope of this study require at least one collaborator affiliated to a Portuguese institution, and another affiliated to an institution in the United Kingdom. According to the European standard reference[7], a **research & development (R&D) activity** comprises the sum of actions deliberately undertaken by R&D performers to fulfil these qualities: novel (aimed at new findings); creative (based on original concepts, interpretations and hypotheses); uncertain: (uncertain about its final outcome); systematic (planned for and budgeted); transferable and/or reproducible (lead to results that could be possibly reproduced). An **R&D project** consists of a set of R&D activities that is organised and managed for a specific purpose, and has its own objectives and expected outcomes, even at the lowest level of formal activity.

According to the Frascati Manual, research projects can engage in three kinds of research:

» **Basic research** is experimental or theoretical work undertaken primarily to acquire new knowledge of the underlying foundations of phenomena and observable facts, without any particular application or use in view.

» **Applied research** is an original investigation undertaken in order to acquire new knowledge. It is, however, directed primarily towards a specific, practical aim or objective.

» **Experimental development** is systematic work, drawing on knowledge gained from research and practical experience and producing additional knowledge, which is directed towards producing new products or processes, or improving existing products or processes.

The classification of fields of research and development[7], also known as **scientific disciplines**, are: Natural Sciences; Engineering and Technology; Medical and Health Sciences; Agricultural and Veterinary Sciences; Social Sciences; and Humanities and the Arts. This definition is relevant for secondary indicators and further analyses, as described in this document. This study considers R&D performers as being either researchers or research bodies.

Researchers are professionals engaged in the conception or creation of new knowledge. They conduct research to improve or develop concepts, theories, models, techniques, instrumentation, software, or operational methods.

This mapping exercise includes **researchers of all career levels**, categorising them according to the four career stages outlined and defined in the European Commission's communication 'Towards a European Framework for Research Careers'[8]:

- **R1:** First Stage Researcher (up to the point of PhD).
- **R2:** Recognised Researcher (PhD holders or equivalent who are not yet fully independent).
- **R3:** Established Researcher (researchers who have developed a level of independence).
- **R4:** Leading Researcher (researchers leading their research area or field).

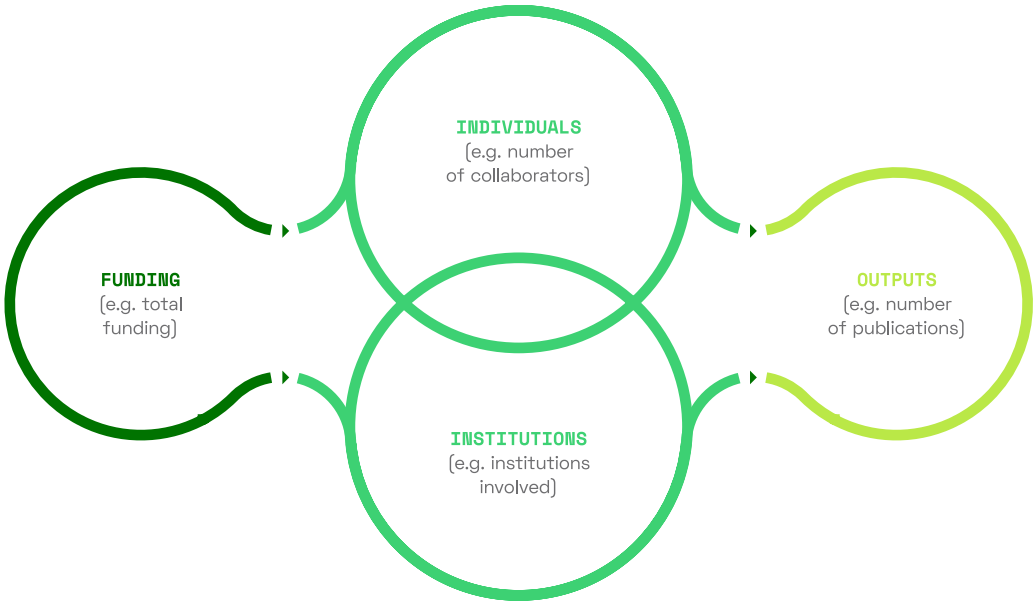
On the other hand, a **research body**, henceforth also designated as an **institution**, is an economic entity capable of decision-making in respect to the conduct of R&D, from the allocation of financial resources for internal or external use to the management of R&D projects. PUMP aims to include all **types of institutions**, categorising them according to the definitions already used by SciVal[9]:

- ACA (academic): university, college, medical school and research institutes.
- COR (corporate): all entities capable of generating a profit or other financial gain for their owners, and that are set up for purposes of engaging in market production.
- GOV (government): government and military organisations.
- MED (medical): hospitals.
- OTH (other): any type of institution not encompassed by the above categories.

R&D involves significant funding flows between units, sectors, and countries in order to cover its costs (which may include: direct and externally provided staff, subcontracted R&D, consumables, software, trials, prototyping, and independent research costs). R&D is conducted by the research performer with funds either from the institution's own sources, or from sources outside of the statistical unit. From the perspective of funding, R&D funds pay for the costs of R&D performance within or outside an institution. R&D funds can be assigned expecting certain performance requirements in return or not (as in R&D grants). **Funding sources** were classified, adapted from Frascati Manual[7], as:

- Business enterprise (private funding).
- Public (government and college grants).
- Private non-profit (philanthropic funding).
- Other (other funding).

This project conceptualises R&D activities as having the goal of translating their novel findings into some document or product, henceforth known as a ***scientific output***. This term encompasses all artefacts generated downstream of the scientific process, that allow the R&D activity to be discovered online, be it co-authored scientific publications, such as editorials, papers, preprints, peer-reviewed journal articles, book series, conference papers, reports, and theses published in peer-reviewed journals and grey literature, or intellectual property such as patents.



Conceptual framework and indicators

figure 1.
Conceptual Framework for PUMP

Data index

The following primary indicators represent data fields that (when available) will factor into our data collection in order to provide an adequate profile of characteristics about each collaboration item:

DOMAIN	INDICATOR	COLLABORATIVE	DESCRIPTIVE
Funding	Number of Grant Submissions	×	
	Total Funding	×	
	Funders	×	
	Type of Funding	×	
	Funding Scheme		×
Institutions	Number of Institutions	×	
	Type of Institutions*	×	
	Primary Institution	×	
	UK/PT Regions Involved		×
	Country of Primary Institution		×
	Institutions Involved		×
Individuals	Number of Collaborators	×	
	Career Level of Collaborators	×	
Output	Number of Scientific Publications	×	
	FWCI	×	
	Journal Quartile	×	
	Number of Patents	×	
	Project ID		×
	Scientific Discipline*		×
	Relevant HE Mission		×
	Acronym		×
	Title		×
	Year of Creation		×
	Duration		×

* This indicator may vary in designation according to the preferred nomenclature of each data source, which may require further translation of each category.

For each **collaboration indicator** (See Data Index), the team established its definition, type of data input (numerical, text), domain (institutions, individuals, funding and outputs) and the justification for its use (see References). The **descriptive indicators** can be found in the Database Architecture, and there is the possibility of creating more descriptive indicators, according to the findings in this first phase. Likewise, if any indicator is found to be inadequate for the purposes of the current exercise, after a first assessment of the data, it can be phased out of the study.

In a later stage, post the data collection phase, a set of secondary indicators will be derived by compounding the previous indicators, in order to obtain more insightful information. These secondary indicators, while prone to further elaboration after the first batch of data collection, should include metrics such as:

- Total funding per funder, institution, country and scientific discipline.
- Success rate of each funding scheme per scientific discipline.
- Total number of collaborations per funder, institution involved, country of primary institution, and scientific discipline.
- Average proportion of funds distributed to Portugal or the United Kingdom.
- Average number of collaborators per collaboration, institution, country, and scientific discipline.
- Average proportion of Portugal-United Kingdom collaborators per collaboration and scientific discipline.
- Average number of outputs per collaboration per total funding and duration.

Inclusion and exclusion criteria

CRITERIA	INCLUSION	EXCLUSION
Timeframe	01/01/2014 - 01/01/2021	Projects that were created before and after our established timeframe, yet containing scientific output within that timeframe.
Themes	Horizon Europe Programme's **5 missions**	Research collaboration whose primary outcomes are not directly related to one of the missions.
Types of research	Basic research and experimental developments	Applied research, such as general purpose data collection (such as recording weather statistics): » scientific and technical information services (collecting, coding, recording, classifying disseminating, translating, analysing and evaluating) except when integral to an R&D project. » testing and standardisation. » feasibility studies. » programmatic evaluations. » purely R&D financing activities. » indirect supporting activities. » routine testing and analysis of materials, components, products, processes, etc. » phase IV of clinical trials (unless they result in a further scientific or technological advance).

methodology

The standard approach outlined for the Phase 1 of PUMP will be to undertake a series of mutually reinforcing steps to build a robust set of data of documented collaborations, along the lines of a PRISMA-like pipeline. This will not only ensure and improve transparency, but also to enlighten all interested parties about the process's logical framework. The steps of this mapping exercise will be:

1. **Identify the collaborations**, based on the agreed inclusion and exclusion criteria, creating a comprehensive description of each collaboration according to previous selected indicators.

2. Collect additional information on each collaboration and fill gaps in primary data sources.

3. Explore the **context and possible ramifications** of these relations.

4. **Validate our findings** with the project team and advisory group.

Our Phase 1 methodology predicts the collection of two types of **evidence** to pinpoint publicly stated (or '**visible**') R&D collaborations: 1) **Reports and data on research funding**, and 2) '***Scientific outputs***', therefore excluding from this phase any other type of evidence. The data extracted from each source will be separately screened and properly filtered down in their respective Excel tab. The filters necessary to obtain the data within our scope are:

1. **Timeframe:** first select only projects with dates of creation within the timeframe criteria; in case of scientific outputs, the dates will refer to the date of publication.

2. **Affiliation:** seek out only projects referring to both Portugal and the United Kingdom; we expect to apply this criteria according to the locations of PI's, projects and/or institutions involved; in case of scientific outputs, the data will be filtered according to author affiliation.

3. **Themes:** apply necessary filters to only show data with themes that fit Horizon Europe's missions, by searching for the established keywords, and guarantee that the primary outcome is directly related to one of the missions, by its title and/or abstract. At this stage, the team will also assess the nature of the documents, to ensure that the types of research match the inclusion criteria.

The next stage of the PUMP exercise will be to tackle the myriad data gaps that we expect to find on our individual data extracts, by **converging them into a single database**. This process involves the cross-matching of different evidence types and sources, in order to tie separate indicators and traits into one data entry for each collaboration, in order to deliver more useful, insightful, and user-friendly data for all interested parties.

When a successful or partial convergence is possible, we will be able to cross-match the scientific output data with project funding reports data and characterise the collaborations with all or most of the desired indicators. This will also ensure exclusion of projects that were created before our established timeframe, yet still contain scientific output within that timeframe.

If it is not possible to obtain all the desired indicators from the defined data sources, the research team will resort to secondary data sources and collection in order to try to fulfil all indicators, in Phase 2. If this is not possible, they shall then proceed to interpret what those gaps in data might signal about those collaborations, and possibly deliberate on further measures. Therefore, a perfect convergence of data extracts for each project, in a way that satisfies all data points (i.e.: indicators), can not be expected for all projects, which implies, in more practical terms, that this database will include publications without funding data and grants without publications to this date, so long as they fit the inclusion criteria.

Collaborative instances where no funding reports and scientific outputs can be assigned will be excluded from our Phase 1 study. These '**invisible collaborations**', more easily traceable through news sources or anecdotal evidence, will only be adequately explored via qualitative methods in later PUMP stages.

types of sources

There are several possible sources from which the desired data might be collected. This study will utilise the most comprehensive and robust sources in Phase 1, and later attempt to harness less accessible data sources in order to either expand or complement the database. The key information sources (primary sources) include:

Databases of public funding

- CORDIS
- FCT
- UKRI
- ERC

Databases of *scientific outputs*

- SCOPUS
- WOS (Clarivate)
- SciVal

Regarding the data sources selected, it is pertinent to mention that:

- In Phase 1, all funding indicators are expected to be obtained via the mentioned databases of public funding (that may or may not include references to partially private or philanthropic funded projects).
- In Phase 2, it is expected that researchers will retrieve information regarding projects that have only private or philanthropic funding, by tracking the collaborations already established.
- The indicator FWCI can only be extracted from one data source, since it will vary between databases, hindering further comparison. Therefore, the single selected data source for this indicator is SciVal.
- For all the other indicators mentioned previously, the team expects to use all databases in order to obtain the desired data.

Not only is there the possibility of retrieving information from these sources, but also of contacting the authors or managers of these databases to obtain additional information. Besides the primary sources, the team is aware of additional, complementary sources. However, these data sources might not be included in Phase 1 of the project because they present greater obstacles to collecting their data and will require further steps than the primary sources. These may include (while not being limited to) the following:

- Ciencia Vitae
- GPS
- PARSUK
- BRF
- PURE
- OSF
- Science Exchange
- Altmetric
- Espacenet

The team will also incorporate extra sources about reports or scientific outputs in posterior phases, if it is determined that the present ones are not sufficient for the established needs.

Example Database
https://docs.google.com/spreadsheets/d/1XQZZrvzQ275ZX7IL8Qp_oB7EbUAfQLwDi8yri-yIn7N0/edit#gid=0

database architecture

quality assessment and validation

Similarly to a systematic review, after the screening process is complete, the team must assess its findings for quality and bias. This quality control can happen internally (1) or externally (2):

1. **Internal PUMP review**: all collected data will be reviewed by both research analysts and, with the feedback between them, there will be self-refinement in methods and in function of the data landscape as it reveals itself throughout the data collection.

    A. **In tandem with quality-assuring literature review**: the selection of indicators and methods is continuously scrutinised in light of new literature that was set out to be progressively gathered throughout this project, in order to assemble a robust literature review able to support and confirm procedural validity.

2. **External validation and review**: the quality of findings will be verified by other elements of governance, including the project team and advisory group. Formal input will be periodically provided, as well as informally via virtual meetings and electronic correspondence, as needed.

Based on a systematic review perspective, some problems and gaps that might arise during the research were identified, and corresponding methodologies or tools to tackle those same problems were suggested[10]:

PROBLEM	DEFINITION	METHODOLOGY/TOOL
Publication bias	Collaborations that do not have an outcome or publication, and therefore were not represented.	Finding 'invisible collaborations' will be a priority in Phase 2.
Time lag bias	Collaborations that occurred during the defined timeframe, but had delayed publications.	See inclusion and exclusion criteria.
Duplicate publications	Multiple publication of research findings, or publications regarding projects already screened in funding sources.	Excel tools to remove duplicates. Mendeley can also be utilised, which automatically removes duplicates.
Location bias	Databases with their own levels of indexing and regimes of accessibility, which may exclude some collaborations.	Several databases used and their information cross-matched in order to try to minimise this problem.
Outcome reporting bias	Selective reporting of some outcomes, but not others.	External validation.

references

1. UUK. (2021) *International facts and figures*. London: Universities UK.

2. European Commission, Directorate-General for Research and Innovation, (2021). Mission area: adaptation to climate change, including societal transformation: foresight on demand brief in support of the Horizon Europe mission board. https://data.europa.eu/doi/10.2777/53790

3. European Commission, Directorate-General for Research and Innovation, Giesecke, S., Jongh, T., Wasserbacher, D., et al. (2021). Mission area: cancer: foresight on demand brief in support of the Horizon Europe mission board. https://data.europa.eu/doi/10.2777/17186

4. European Commission, Directorate-General for Research and Innovation, Lacroix, D., Boero, F., Ligtvoet, A., et al. (2021). Mission area, healthy oceans, seas, and coastal and inland waters: foresight on demand brief in support of the Horizon Europe mission board. https://data.europa.eu/doi/10.2777/054595

5. European Commission, Directorate-General for Research and Innovation, Borsboom, J., Haindlmaier, G., Dinges, M., et al. (2021). Mission area : climate-neutral and smart cities: foresight on demand brief in support of the Horizon Europe mission board. https://data.europa.eu/doi/10.2777/123417

6. European Commission, Directorate-General for Research and Innovation, Voglhuber-Slavinsky, A., Bisoffi, S., Helming, K., et al. (2021). Mission area : soil health and food: foresight on demand brief in support of the Horizon Europe mission board. https://data.europa.eu/doi/10.2777/038626

7. OECD (2015). Frascati Manual 2015: Guidelines for Collecting and Reporting Data on Research and Experimental Development, *The Measurement of Scientific, Technological and Innovation Activities*, OECD Publishing, Paris. https://doi.org/10.1787/9789264239012-en.

8. European Commission (2011). *Towards a European Framework for Research Careers*. Available at: https://cdn5.euraxess.org/sites/default/files/policy_library/towards_a_european_framework_for_research_careers_final.pdf (Accessed: 11 March 2022).

9. Elsevier, Research Intelligence. (2019). *Research Metrics Guidebook*. Elsevier B.V.

10. University of North Carolina at Chapel Hill, LibGuides. (2022). *Types of bias* [Online]. Quality Assessment - Systematic Reviews. Available at: https://guides.lib.unc.edu/systematic-reviews/quality-assessment (Accessed: 11 March 2022).

11. García-Villar, C., & García-Santos, J.M. (2021). Indicadores bibliométricos para evaluar la actividad científica. *Radiología*, 63:228-235.

12. Tigges, B.B., Miller, D., Dudding, K.M., Balls-Berry, J.E., Borawski E.A., Dave, G., Hafer, N.S., Kimminau, K.S., Kost, R.G., Littlefield, K., Shannon J., & Menon, U. (2019). The Measures of Collaboration Workgroup of the Collaboration and Engagement Domain Task Force, National Center for Advancing Translational Sciences, National Institutes of Health. Measuring quality and outcomes of research collaborations: An integrative review. *Journal of Clinical and Translational Science*, 3: 261-289. doi: 10.1017/cts.2019.402

B. thematic screening keywords

thematic screening keywords

The keywords used to identify the projects and publications in scope were:

1. **Adaptation to climate change**: climate change, climate service*, extreme weather, food security, global warming, renewable energ*, sustainab*, carbon emission*, carbon neutral*, carbon reduction, carbon sequestration, decarbonisation, ecosystem service*, energy efficien*, energy transition, environmental solution, greenhouse gas*, hydropower, resilient agriculture, resilient fisher*, resilient infrastructure*, solar energy*, solar power, wind energy*, wind power.

2. **Cancer**:cancer,carcinogen*,onco*,tumour,chemotherapy,immunotherapy, mutagen*, radiotherapy.

3. **Restore our oceans and waters**: aquatic ecosystem*, clean water, climate-resilien*, aquaculture, aquaponic, coastal ecosystem*, coastal protection, coastal vulnerability, hydroponic, marine ecosystem*, marine environment, sea-level rise.

4. **Climate neutral and smart cities**: climate neutral cit*, smart cit*, urban food, autonomous vehicle, bio-based, bioeconomy, building automation, carbon-intensive, circular economy, electric vehicle, energy source diversification, envelope retrofit, environment-friendly, green procurement, recycling, refurbishment lighting, shared mobility, smart building, smart lighting, sustainable procurement, urban agriculture, vertical farming, waste management, water heating.

5. **Soil deal for Europe**: desertification, erosion, land degradation, agroforestry, biodiversity preservation, biological agriculture, biological farming, bioremediation, carbon sequestration, carbon sink, conservation agriculture, crop loss, crop management, loss of land, organic agriculture, organic farming, remediated site, reverse carbon loss, soil, vegetation cover, water desalination.

C. PRISMA diagrams

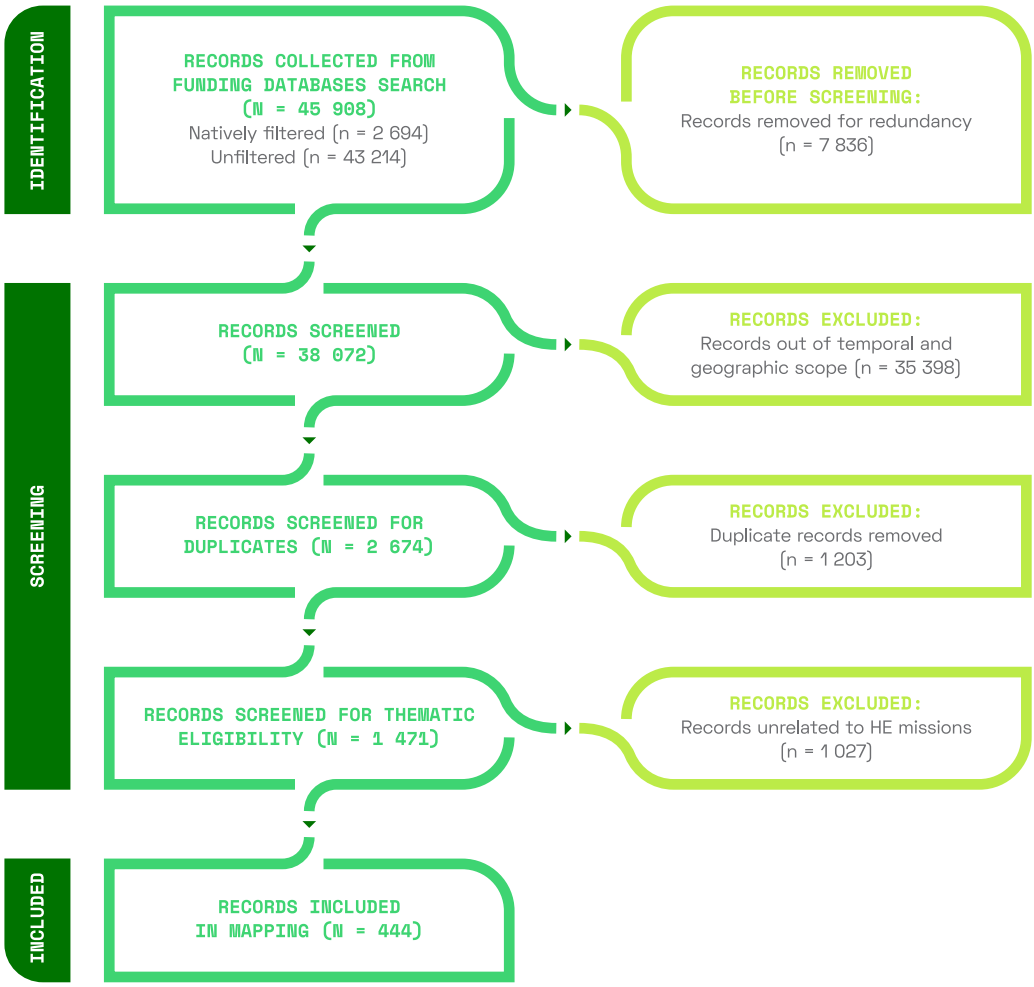


figure 1.
Flow diagram of steps for the assembly of the **projects** included in the database, including the number of records identified, included and excluded, and reasons for exclusions. It is based on the template provided in the PRISMA 2020 flow diagram for new systematic reviews, which included searches of databases and registers only.

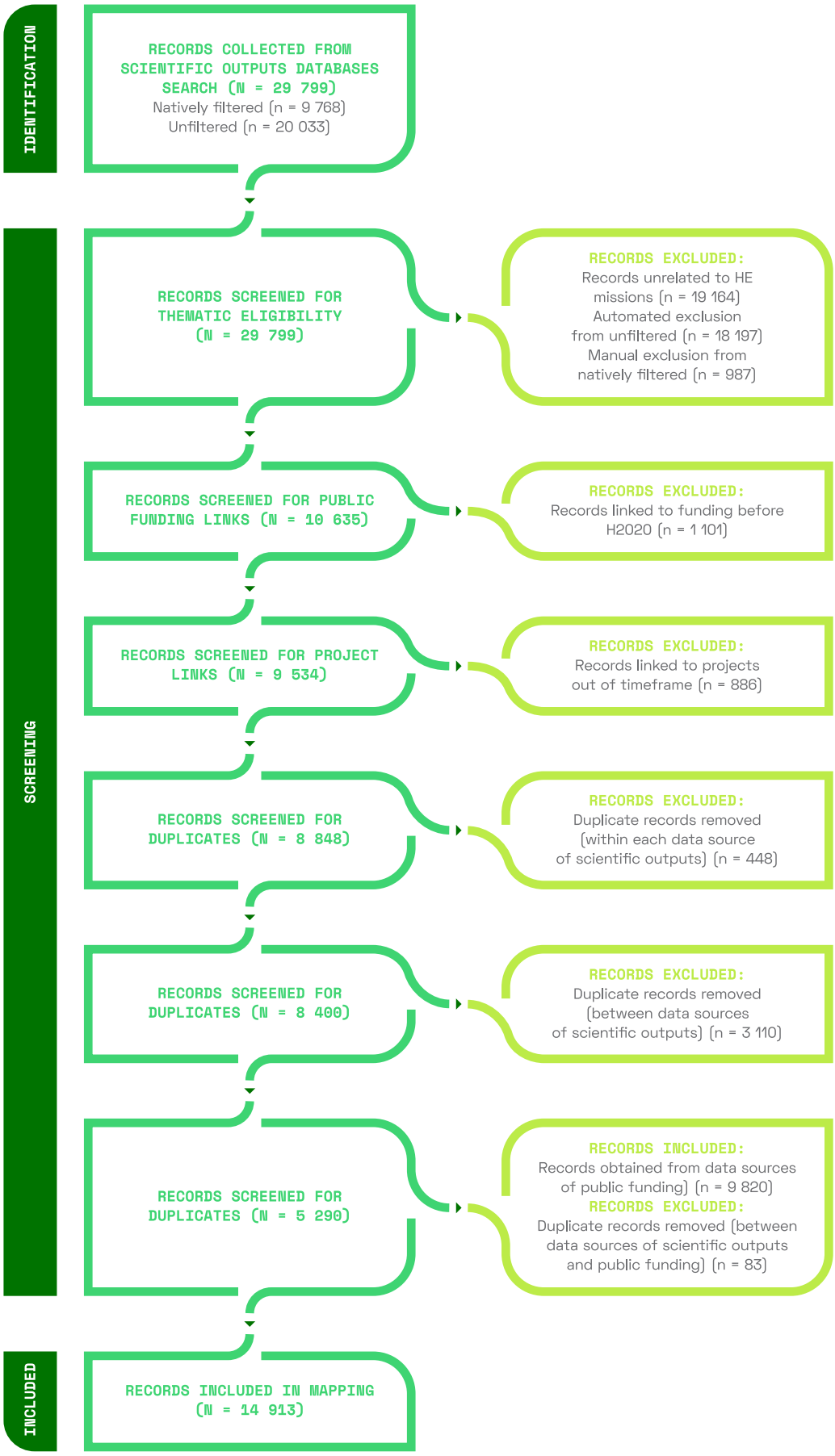


figure 2.
Flow diagram of steps for the assembly of the **publications** included in the database, including the number of records identified, included and excluded, and reasons for exclusions. It is based on the template provided in the PRISMA 2020 flow diagram for new systematic reviews, which included searches of databases and registers only.

D. inter-rater reliability protocol

inter-rater reliability protocol

Inter-rater reliability is the degree to which two or more raters agree, which tackles consistency issues in the implementation of a rating system. A variety of statistics can be used to evaluate inter-rater reliability; percentage agreement is one of the most popular. A high degree of agreement between two examiners is referred to as strong inter-rater reliability (Anastasi & Urbina, 1997).

The team set a target of achieving a high degree of agreement, exceeding 90%. In case the level of agreement was lower than 90%, the researchers analysed each disagreement, checked for the divergent criteria, and attempted to come to a consensus. This circumstance prompted an overall methodology correction, applied to the current and next data sources to be addressed. A second filtering of the HE missions in the current disputing dataset was performed, using the new methods, and the level of agreement was calculated again, this time with a sample of 20%, ensuring that the new methods guarantee a level of agreement equal to or greater than 90%.

Due to the size of the data, the researchers took a random sample of 10% of each dataset of public funding to determine the IRR. Since each data source provides different information, this was done to each data source's data. The succeeding scientific output data was much larger in size, and thus the 10% sample included only publications containing the keyword 'sustainab', rather than the entire dataset; all other publications were collected using the predefined keywords, and were already covered by the previously agreed-upon modifications that had been made in aforementioned IRR exercises.

The IRR was measured at a point in the workflow in which indicators required the researchers' judgement, i.e.: filtering the projects by HE missions. Because, as previously stated, the initial set of keywords did not necessarily work for all data sources. The only data source that the researchers did not deem necessary to undertake the IRR exercise for was FCT, since none of the projects that were collected used a problematic keyword (such as 'sustainab') and were also very well described and incorporated into one or more of the HE missions, rendering this exercise superfluous.

To find the percent agreement for the two researchers, it is necessary to:

1. Count the number of ratings in agreement.
2. Count the total number of ratings.
3. Divide the total by the number in agreement to get a fraction and convert to a percentage.

	CORDIS	UKRI	SCIVAL	SCOPUS	WOS
Population (nr projects/publications)	160	1050	132	379	966
10% sample	16	105	13	38	97
1st Percentage agreement	80%	94%	92%	92%	96%
2nd Percentage agreement	92%	-	-	-	-

E. interview script

guião - português

O meu nome é Diogo Correia e a minha colega é a Beatriz Varandas e queremos, antes de mais, agradecer-lhe por ter aceite o nosso convite. Esta entrevista insere-se no âmbito de um Projeto da PARSUK que visa mapear Colaborações Científicas entre Portugal e o Reino Unido, tendo por objetivo estudar as suas características e fatores de sucesso e risco, compreender os mecanismos pelos quais se formam e retirar ilações em prol de colaborações futuras. Salientamos que todas as respostas fornecidas serão apenas aproveitadas para fins de investigação. Todos os nomes de indivíduos serão anonimizados. A entrevista será gravada, porém apenas a gravação em áudio será armazenada para fins de transcrição.

Esta entrevista servirá para falarmos acerca de um projeto ou outro tipo de empreendimento científico no qual tenha, enquanto afiliado a uma instituição portuguesa/britânica, trabalhado em colaboração com um ou mais investigadores sediados no Reino Unido/em Portugal. Pedimos que considere a experiência mais memorável ou significativa em que já se envolveu com estas caraterísticas. Posso iniciar a gravação? *(aguardar confirmação do participante)*.

Obrigado pela sua participação e por consentir com as condições da entrevista

1. **Poderia descrever o âmbito científico da sua colaboração?**
   (questão desnecessária no caso BRF)

Sugestões:

- Ciências Naturais
- Engenharia e Tecnologia
- Ciências Sociais
- Ciências Agrárias e Veterinárias
- Ciências Médicas e da Saúde
- Humanidades e Artes

2. **Em que nível de carreira se encontrava?**

Sugestões:

- First Stage Researcher (up to the point of PhD);
- Recognised Researcher (PhD holders or equivalent who are not yet fully independent);
- Established Researcher (researchers who have developed a level of independence);
- Leading Researcher (researchers leading their research area or field).

3. **Em que instituição se encontrava a trabalhar aquando do início da colaboração?**
   (questão desnecessária no caso BRF)

4. **Que outras instituições estiveram envolvidas, seja no Reino Unido ou em outros países?**

Nota:
- Definir se UK ou PT
- Definir colaboração como multi ou bilateral
- Definir setor da instituição

Sugestão:
- Se preferir, enumere antes os indivíduos que participaram.

5. **O *lead researcher* era você? Era o seu colaborador português/britânico?**

6. **A iniciativa de estabelecer a parceria proveio do lead researcher, de um co-investigator, ou de uma terceira entidade?**
   - A. Houve condições institucionais ou protocolos pré-existentes que agilizaram o estabelecimento da colaboração, ou ela formou-se graças a iniciativa puramente pessoal?
     - I. Se estes fatores de agilização existirem, quais são?
   - B. Como foi transmitido o convite para estabelecer a colaboração?

   Sugestões:
   - Conferência/Seminário/Evento Científico
   - Convívio Informal
   - No Local de Trabalho
   - Chamada/Videochamada
   - ResearchGate/LinkedIn/Twitter/Outras Redes Sociais
   - E-mail
   - Website Pessoal
   - Carta

7. **Quando se iniciou a colaboração?**
   - A. Está em decurso ainda?
     - I. Se sim, tem duração expectável?
       1. Se sim, qual?
     - II. Se não, quanto tempo durou?
   - B. A duração esperada foi a que se sucedeu?
     - I. Se não, porquê?

8. **Que expectativas tinha no início da colaboração?**
   - A. Em que medida essas expectativas foram cumpridas?
   - B. Que obstáculos enfrentaram?

9. Quais os principais motivos para ter participado nesta colaboração?

Sugestões:

- Prestígio académico
- Admiração profissional
- Perspetivas de financiamento futuro
- Oportunidades de aprendizagem
- Aprofundar relações com uma universidade ou organização
- Fomentar novas parcerias no futuro
- Interesse comum num ramo de investigação de nicho
- Aumentar a visibilidade do ramo de investigação
- Uma ideia de investigação difícil de concretizar sem dividir esforços
- Cumplicidade/ Boas relações com o colaborador

10. A colaboração recebeu financiamento?
(questão desnecessária no caso BRF)

Nota:

- Definir o caráter informal ou formal da colaboração em paralelo com o trajeto de financiamento (independente ou adjacente?)

A. Se sim, poderia divulgar a quantia atribuída?
   I. Esta atribuição de financiamento foi documentada online?
      1. Se sim, este montante proveio integralmente da Comissão Europeia, FCT, UKRI ou ERC?
         A. Se nenhuma das anteriores, consegue divulgar que outras fontes de financiamento existiram?
      2. Se não, de onde proveio este financiamento?

B. Se não, que impacto isso teve na colaboração?

11. Foram gerados conteúdos académicos ou culturais em resultado desta colaboração?

A. Foram publicados artigos em coautoria?
   I. Se não, qual o principal motivo na sua opinião?

   Sugestões:

   - O artigo está em desenvolvimento/por publicar
   - Essa nunca foi uma prioridade/um objetivo da colaboração
   - A parceria terminou antes disso ser possível
   - Foram feitas outro tipo de publicações (capítulos de livros, editoriais, etc.)

B. Foram publicadas patentes ou registos de propriedade intelectual?

C. Foram criados outro tipo de produtos culturais?
   I. Se sim, quais?

   Sugestões:
   - Conteúdos em Redes Sociais
   - Blogues
   - Notícias
   - Conteúdos em Website
   - Livros
   - Conteúdos Audiovisuais Online
   - Podcasts
   - Exposições
   - Palestras
   - Workshops/Formações
   - Software

12. Qual foi o impacto industrial, económico ou comercial da sua colaboração?

13. O que preferia que tivesse sido diferente acerca da sua colaboração?

14. Que ferramentas considera que poderiam facilitar mais colaborações deste tipo?

Sugestões:
- Fundos de Pump Priming
- Incentivos à mobilidade
- Programas de coordenação e co-orientação internacional de doutoramentos
- Plataformas digitais especializadas
- Iniciativas de Open Science

15. Que impactos teve esta colaboração sobre o resto da sua vida profissional?

16. O que é que a comunidade académica pode aprender com a sua experiência colaborativa?

Não temos mais questões, muito obrigado pela sua contribuição.

script - english

My name is Diogo Correia, and my colleague is Beatriz Varandas. First of all, we thank you for accepting our invitation. This interview aims to collect data for a PARSUK Project on Scientific Collaborations between Portugal and the United Kingdom, with the objective of studying their characteristics and success and risk factors, understanding the mechanisms by which they are formed, and drawing conclusions in favour of future collaborations. We stress that every response obtained shall be used for research purposes only and, if necessary, depicted in anonymity, without naming any individual or entity. All names of individuals will be anonymised. The interview will be recorded, but only the audio recording will be stored for transcription purposes.

In this interview, we will ask you to talk about a project or other type of scientific endeavour in which you have, while affiliated with a Portuguese/British institution, worked in collaboration with one or more researchers based in Portugal/the United Kingdom. We ask that you consider the most memorable or meaningful experience you have ever been involved in with these characteristics. Can I start recording? *(Wait for confirmation from the participant)*.

Thank you for your participation and for agreeing to the conditions of the interview.

1. **Could you describe the scientific scope of your collaboration?**

Suggestions:

- Natural Sciences
- Engineering and technology
- Social Sciences
- Agricultural and Veterinary Sciences
- Medical and Health Sciences
- Humanities and Arts

2. **What career level were you at?**

Suggestions:

- First Stage Researcher (up to the point of PhD).
- Recognised Researcher (PhD holders or equivalent who are not yet fully independent).
- Established Researcher (researchers who have developed a level of independence).
- Leading Researcher (researchers leading their research area or field).

3. **In which institution were you working at the beginning of the collaboration?**

4. **What other institutions were involved, either in the UK or in other countries?**

- Define whether UK or PT.
- Set collaboration to multi- or bilateral.
- Define the sector of the institution.

Suggestions:
- If you prefer, list the individuals who participated beforehand.

5. **Were you the lead researcher? Or was it your collaborator?**

6. **Did the initiative to establish the partnership come from the lead researcher, a co-investigator, or a third party?**
   A. Were there institutional conditions or pre-existing protocols that facilitated the establishment of collaboration, or was it formed thanks to a purely personal initiative?
      I. If these enabling factors exist, what are they?

   B. How was the invitation to establish collaboration transmitted?

   Suggestions:
   - Conference/Seminar/Scientific Event.
   - Informal socialisation.
   - At the Workplace.
   - Call/Video call.
   - ResearchGate/LinkedIn/Twitter/Other Social Networks.
   - Email.
   - Personal Website.
   - Letter.

7. **When did the collaboration start?**
   A. Is it still going on?
      I. If yes, is it expected to last?
         1. How long is the expected timeframe?
      II. If not, how long did it last?

   B. Was the expected duration the one that followed?
      I. If not, why?

8. **What expectations did you have at the beginning of the collaboration?**
   A. To what extent were these expectations fulfilled?

   B. What obstacles did you face?

9. **What are the main reasons for having participated in this collaboration?**

Suggestions:

- Academic prestige.
- Professional admiration.
- Future financing prospects.
- Learning opportunities.
- Deepen relationships with a university or organisation.
- Foster new partnerships in the future.
- Common interest in a niche area of research.
- Increase the visibility of the research.
- A research idea that is difficult to implement without dividing efforts.
- Complicity/good relations with the collaborator.

10. **Has the collaboration received funding?**
    - A. If yes, could you disclose the amount awarded?
        - I. Has this funding been documented online?
            1. If yes, did this amount come entirely from the European Commission, FCT, UKRI or ERC?
                - A. If none of the above, can you disclose what other sources of funding existed?
            2. If not, where did this funding come from?
    - B. If not, what impact did it have on collaboration?

11. **Were academic or cultural content generated as a result of this collaboration?**
    - A. Have co-authored articles been published?
        - I. If not, what is the main reason in your opinion?

          Suggestions:
          - The article is under development/unpublished.
          - This was never a priority/a goal of collaboration.
          - The partnership ended before that was possible.
          - Other types of publications (book chapters, editorials, etc.).
    - B. Have patents or intellectual property records been published?
    - C. Were other types of cultural products created?
        - I. If yes, which ones?

          Suggestions:
          - Content on Social Networks.
          - Blogs.
          - News.
          - Website Content.
          - Books.
          - Online Audiovisual Content.
          - Podcasts.
          - Exhibitions.
          - Lectures.
          - Workshops/Training.
          - Software.

12. What was the industrial, economic or commercial impact of your collaboration?

13. What would you rather have been different about your collaboration?

14. What tools do you think could facilitate more collaborations of this type?

Suggestions:

- PUMP Priming Funds.
- Mobility incentives.
- International PhD coordination and co-supervision programs.
- Specialised digital platforms.
- Open Science Initiatives.

15. What impacts did this collaboration have on the rest of your professional life?

16. What can the academic community learn from your collaborative experience?

We have no further questions, thank you very much for your time.

F. survey

survey

general information

The goal of this questionnaire is to identify scientific collaborations between Portugal and the United Kingdom. These may be formal or informal, visible or invisible, bilateral or multilateral with relevant individuals and institutions.

Data from this questionnaire will give us give us a fuller picture of the scientific landscape between the two countries and how we can best support it in the future.

By participating in this questionnaire, you are consenting the use of your data for research purposes only. Filling out this survey should take approximately 10-15 minutes.

Thank you.

Add name/e-mail

PART 1 - IDENTIFICATION

* Mandatory

Name: [short answer]

E-mail: [short answer]

Institution*: [short answer]

City/Country*: [short answer]

Scientific discipline*: [select one]

- Natural Sciences
- Engineering and Technology
- Social Sciences
- Agricultural and Veterinary Sciences
- Medical and Health Sciences
- Humanities and the Arts
- Other

Career level*: [select one]

- Early-career researcher
- Mid-career researcher
- Senior-career researcher

PART 2 – SCIENTIFIC COLLABORATION

Name of collaborator: [short answer]

E-mail of collaborator: [short answer]

Collaborating institution*: [short answer]

City/Country of collaborating institution*: [short answer]

Start date*: [dd/mm/yyyy]

Duration*: [months]

Funding*: [yes/no/I don't know or not applicable]

Amount: [Euros/GBP]

Source: [short answer]

Other countries involved: [short answer]

PART 3 – START AND MOTIVATORS

How was this scientific collaboration initiated*? [long answer] You may wish to describe the format (e.g. face-to-face event, remote, etc.), how exactly it started, and whose initiative it was. Please be as specific as possible in the context of your research field.

What motivated this scientific collaboration*? [long answer] You may wish to describe what your personal and/or group expectations were when starting this collaboration. Please be as specific as possible in the context of your research field.

PART 4 – ENABLERS AND BLOCKERS

What factors have facilitated the start and maintenance of this scientific collaboration*? [long answer] You may wish to refer to personal, institutional, and/or external factors. Please be as specific as possible in the context of your research field.

What factors have hindered the start and maintenance of this scientific collaboration*? [long answer] You may wish to refer to personal, institutional, and/or external factors. Please be as specific as possible in the context of your research field.

PART 5 – RESULTS AND IMPACT

Publications*: [yes/no/I don't know or not applicable]

Type: [short answer]

Number: [short answer]

Details: [long answer] Please provide references where possible.

Intellectual property*: [yes/no/I don't know or not applicable]

Number of patent applications: [short answer]

Details: [long answer] Please provide references and applicant country where possible.

Other results: [long answer] You may wish to describe other social, economic, political, commercial or industrial results of this scientific collaboration. Please provide references where possible.

What was the impact of this scientific collaboration in your career?* [select one]

- Very positive
- Positive
- Neutral
- Negative
- Very negative
- I don't know or not applicable

Please explain why*: [long answer]

END / THANK YOU

Please share your suggestions on how we can best support the international scientific collaboration landscape between Portugal and the United Kingdom. [long answer]

Do you wish to be contacted about future research on this topic or collaboration opportunities with us?* [yes/no]

G. participant information sheet

participant information sheet

PUMP: Portugal – United Kingdom Mapping Project
Project duration: March - August 2022

introduction

PUMP is a multidisciplinary project led by PARSUK with the goal of mapping the recent evolution of scientific collaborations between Portugal and the United Kingdom. The main goals of this project are to deliver a comprehensive database of scientific collaborations between Portugal and the United Kingdom, and a final report. This report will include, besides a quantitative description, a more qualitative assessment of the researchers' individual experiences on these collaborations, in order to develop case-studies of emerging partnerships and hidden connections. This information will be crucial to develop feasible methods to proactively detect interactions in the field and inform multiple stakeholders involved in science policy, higher education, and research fund management (e.g.: the Ministry of Science, Technology and Higher Education in Portugal).

Your participation in this interview is voluntary, and you have the right to withdraw at any point without consequences. This document aims to clarify all questions that might arise regarding your participation in this project.

interview description

If you choose to participate in the research, you will be asked to commit to an online interview which should take around **30 minutes**. The interview will be **conducted and recorded (via Zoom)** for ease of transcription.

The interview will delve into your experience of a specific scientific collaboration between the two countries, Portugal and the United Kingdom, in order to study its characteristics, success and risk factors, understand the mechanisms underlying its emergence, and draw conclusions in favour of future collaborations. For these purposes, you will be asked to share names of institutions, projects, and other collaboration-related data, which will only be used with your consent for this study. However, **you and all individuals mentioned during the interview will remain anonymous. The data archives for each interview will also be anonymised** (with numeric identifiers only accessible to the two researchers working on this study). If you feel unable to answer a question, it will be omitted: participation is voluntary and there are no consequences to omitting a question.

data management plan & ethical considerations

Collected data will be stored securely and anonymously in PARSUK's system. All data provided may be featured in the final report, except for the name of individuals – namely people referenced by the interviewee, or the interviewee themselves. Participants may ask to retrieve the data they have provided and/or require the elimination of the data at any moment, by contacting the research analysts, Beatriz Varandas or Diogo Correia, through the emails disclosed below.

Ethical safeguards will be implemented to ensure best research practices are always followed in this project; to guarantee best research practice is always respected, the research team will be in constant contact with the other elements of PARSUK, including the project team and advisory group. For any requests or questions, whether a priori or a posteriori, please don't hesitate to contact the team of research analysts at **beatriz.varandas@parsuk.pt** or **diogo.correia@parsuk.pt**.

parsuk portuguese association of researchers and students in the uk

www.parsuk.pt