Framework document Steering group VSNU/NFU/NWO – Elsevier

Name Pilot/Service: funding data flow

Expected outcome:

The aim of this pilot is to improve the open flow of data regarding research investments of Research Funding Organizations (awarded grants) and resulting research outputs enabling universities participating in the project to inform key stakeholders, such as funders, research administration and researchers.

The "open flow" will increase the quality/coverage of the data, such that better and more complete information can be given and shared publicly (and subsequently better decisions can be made)

There are no restrictions for the institutions on how the enriched and stored data can be used; they can make it publicly available through their public portals, through a future Open Knowledge base, etc.

Following a review of the enrichment and ingestion processes and the available data, the partners may decide to start a follow-up pilot which will focus on:

- For institutions & researchers: researcher engagement and validation, further matching services and subsequent automated delivery of the associated linked outputs (both traditional and non-traditional) to institutional portals or any other destination as determined by the applicable participating institutions
- For the RFO: the automated delivery to and ingestion of research outputs associated with research funded by the RFO increasing coverage of those associated research outputs.

1. (a) Participating institutions Participation in the Professional Services is at each Institution's sole discretion and a pilot shall only commence if there is a minimum participation by at least three Institutions *	Evaluation YES NO	Evidence and Comments
Are at least 3 institutions involved in the pilot?	YES	5 institutions have been actively participating in the development of this pilot: TU Delft, TU/e, UU, RUG and Amsterdam UMC

Evidence of how and when other institutions can join	YES	Any institution can join the pilot at any point in time.
1. (b) Interoperability and vendor neutrality Elsevier shall use all reasonable efforts to ensure that the Professional Services are interoperable, both on the input (uploaded) and output side (created) *	Evaluation YES NO	Evidence and Comments
Use of open identifier systems	YES	Grants are identified via their source persistent identifiers (grant ID). The enrichment pipeline is designed to maximize author and institutional attribution, as well as enabling linking with entities in Current Research Information Systems such as Pure. During the enrichment process, at the moment the Scopus institutional identifiers are used. Scopus Author ID and ORCID, when available, are also added.
Use of standardised metadata schemas		To be investigated with the partners. If there is one, we will use it. Outcome of the pilot could be a recommended format.
Existence of a well-documented API and open data-dump function		This is one of the deliverables of the pilot and developed in close collaboration with the participating institutions.
Ability to export data in a variety of formats		Json
Ability for other commercial parties to join	YES	Any other 3rd commercial party can make use of the funding data and associated output available in the CRIS systems of the institution, at the discretion of the institutions.

2. Transparency, inclusion and collaboration The Services and resulting Deliverables are aimed to make science and research more transparent, efficient, inclusive, openly and freely accessible, and collaborative. *	Evaluation YES NO	Evidence and Comments
Provenance on how and where metadata was derived	YES	The funding stream is based on the CORDIS Platform which is open to everyone.
Descriptions of workflows that result in indicators, metrics and/or other relevant outcomes will be open and transparent. These will demonstrate, for example calculation steps, search strings used to define entities, etc.	NA	-
Description of the services used to create metadata	YES	Harmonized data set will go through the content enrichment pipeline, which includes such services as: • Matching of awardees to Scopus (and, if available, ORCID) Author profiles • Matching affiliation to Scopus Institution profiles • Conversion of EUR to USD for benchmarking analysis • NLP text mining to assign concepts and ASJC research categories • Other enrichment services when these are built
Insights and lessons published with Open Access license	YES	While this is at the discretion of the partners, Elsevier staff members will be happy to act as co-author or support in other ways the dissemination of insights and learnings from this pilot.
Will the pilot contribute to Open Science?	YES	This pilot would enable research institutions to get a constant homogenized flow (potentially with

		feedback loops to improve quality) of funding data that could be then linked within their systems to researchers, projects, pieces of equipment and outputs. At a later stage (when a majority of institutions agree to join) an national integrated dataset could be generated to be stored and made available in a national knowledge base.
Demonstration of connection to non-Elsevier products	YES	The API will make it possible to integrate with any CRIS (not just Pure) or any other system (such as a middleware) available at the institution.
3. Access to research data and metadata Elsevier will give enduring access during the Term to all (research) data, including metadata, analytics and information*	Evaluation YES NO	Evidence and Comments
Describe the ownership / licensing of data made as part of the service	YES	The Metadata links to a public data source (CORDIS). Once metadata has been provided, it will be put into the system of record of the institution. The institution then decides what do with this data. This way the data is also available for the NL OKB under development, if and when this is required.
Describe how access (institutional and / or public) to the data will be set-up during the term; this section will also indicate cases where certain data is not publicly access.	YES	Access will be provided through an API. The API key will be provided to the participating institutions once the SoW is signed. The service will also integrate with RIM systems. Elsevier will offer to work with RIM providers to integrate the API or the data with their systems. In the case of Pure this is already possible through an existing capability.

4. Data portability Institution shall be entitled to transfer the data provided, uploaded or created to its own or to a third party host environment *	Evaluation YES NO	Evidence Comments
Evidence on how data transfer is possible.	YES	The data is portable and/or remains in the institutional information system when the service is cancelled/discontinued.
How can an institution withdraw data?	NA	NA
5. Intellectual property *	Evaluation YES NO	Evidence Comments
Details on IP related to data provided, created or enriched	NA	NA
6. Additional considerations	Evaluation YES NO	Evidence Comments
What processes will be put in place to evaluate the service during and at the end of pilot	YES	The evaluation of the service is part of the pilot itself and one of the deliverables as identified in the SoW; there will be an ongoing conversation about the usability of the data and the functionality and documentation of the API.
Terms of use of the deliverables during and after contract period	YES	The deliverables of this pilot are stated in the SoW. If participating institutions so wish, service is available as part of the agreement until December 31, 2024. In case participating institutions do not want to continue to use the service, all public data serviced via the API or Pure will remain in their system (if participating institutions so wish)
Pilot project team	YES	Names of the partners' contacts detailed in the respective SoWs

For the full text, please refer to the contract.

| Approved by the VSNU/NFU/NWO-Elsevier steering group | Date: 27 October 2021 |