Framework document Steering group VSNU/NFU/NWO – Elsevier

Name Pilot/Service: Preprints Monitor
Short description of the pilot/service: Preprints have been a key part of the publishing process ever since 1991 (when the arXiv started) and form a fundamental pillar in Open Access publishing and Open Science. While preprints are openly available for reading, their use for other purposes such as meta data ingestion in a CRIS system, alerts or impact tracking is less common and not straightforward to support. In order to make such use cases possible, preprints need to be harvested from multiple preprint servers/services, their metadata harmonised and authors as well as their affiliations identified. • This pilot would aim at creating such a harmonised and enriched meta data feed that can be ingested by participating institutions' CRIS systems, repositories, an Open Knowledge base, etc. • A long-term objective of this pilot (not part of the intended outcomes of the pilot) would be to enable linking of preprints to their final published versions (version of record). Such a link would enable tracking of scholarly communication from early stage to final stage, and combine the online attention often received by preprints versions to the academic impact generally received by final publications; an often heard and much wanted need within the academic community. • An important characteristic of preprints is that they are being edited over time and that there are updates and subsequent versions of preprint papers. We aim to include links to the latest version at all times (given acceptable update intervals). • 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.

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	This project idea was submitted through an anonymous survey to the Dutch Pure user group in February 2022. 12 people responded, out of which 5 said they would be interested in such a pilot, and

		4 more could be. The following institutions formally shared their interest in this pilot: Amsterdam UMC, TU Delft, University of Groningen, Utrecht University and Wageningen U&R
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	Most if not all harvested preprints have a DOI. Elsevier's 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 or csv
Ability for other commercial parties to join	YES	Any other 3rd commercial party can make use of the preprints data 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	Preprints are currently harvested from 7 of the largest preprints servers: arXiv, bioRxiv, ChemRxiv, medRxiv, SSRN, TechRxiv and Research Square.
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	Harvested preprints go through Elsevier's content enrichment pipeline, which includes such services as: • Matching of authors to Scopus (and, if available, ORCID) Author profiles • Matching affiliation to Scopus Institution profiles
Insights and lessons published with Open Access licence	YES	This is at the discretion of the partners. We would love to be a co-author to the publishing of these insights and learnings.
Will the pilot contribute to Open Science?	YES	Publishing via preprints is already common practice in disciplines such as Physics and gained a lot of momentum in Biology and Medicine since the start of the pandemics. Preprints enable quick and free access to research to everyone.

		This pilot would enable additional use cases linked to preprints: • track and promote Open Science behaviors (alongside OA publishing); by tracking pre-prints as an accepted and established form of scientific communication, this could also impact the publishing behavior of researchers • enable institutions to support researchers who want to showcase their preprints and/or use them in their CVs (for grant proposals) • help better assess the overall impact of publications (preprints + final published version)
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 sources (all harvested preprints servers are open). 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	At the start of the project, institutions will receive a data dump of their preprints' metadata. Once developed, 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 CRIS providers to integrate the API with their systems.
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: when finalized