Pilot initiation template

This template serves to assess an early pilot idea to be executed under the VSNU/NFU/NWO-Elsevier services agreement.

Pilot idea can be submitted by any member of the agreement.

Pilot name:	Preprints Monitor
Date:	March 2022
Submitted by:	Guillaume Warnan (Elsevier), after validation with the Dutch Pure user group and the support of 5 institutions (Utrecht University, Wageningen U&R, University of Groningen, Maastricht University and Amsterdam UMC)
Time Required:	9-12 months

Short description of the project:
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 subsequent objective of this pilot (not part of the first phase of the pilot) would be to link 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.

This project idea originated from the community and interest was confirmed through an anonymous survey to the Dutch Pure user group (which represents 12 out of the 14 research universities, the KNAW, and a few smaller institutions) in February 2022. 12 people responded, out of which 5 said they would be interested in such a pilot, and 4 more could be; i.e. 75% of the respondents showed interest for such a pilot.

What OS need/opportunity will this address?
Publishing via preprints is already common practice in disciplines such as Physics and Econometrics, and gained a lot of momentum in Biology and Medicine since the start of the pandemic.
Preprints enable quick and free access to research to everyone and allows the author to claim the date of the initial results regardless how long the publication process may take until final publication in a journal.

In the same survey previously mentioned, respondents were asked about potential use cases / services such a preprints feed would make possible:

- 9 respondents said such a feed would enable to track and promote Open Science behaviours (alongside OA publishing); one respondent added that “By taking pre-prints more seriously as an accepted and established form of early scientific communication, we hope this will also impact the publishing behaviour of our researchers”
- 7 said it would enable institutions to support researchers who want to showcase their preprints and/or use them in their CVs (for grant proposals)
- 7 said it would help better assess the overall impact of publications (preprints + final published version)

In addition to those use cases, the creation of an alert (based on this feed) would enable institutions to identify research yet to be published and proactively reach out to authors to verify that their research data is/will be made openly available, FAIR compliant and will be properly linked to the final version of the article. As such, this pilot would also help support the Open Data ambitions in NL.

What type of services or outputs will it deliver?
This service will deliver:

- An API enabling to feed CRIS systems or other types of systems
    - Initially data will be delivered via a dump in order to jointly optimise the schema and identifiers
    - API will be developed in a parallel process.
- Tracking the impact of preprints
    - Online impact (media mentions, social media, etc.)

What are the major risks (such as complexity or dependency) in the project?

Risk	Impact (1 to 5)	Possibility (1-5)	Overall risk (I * P)	Mitigation
Insufficient quality of the indexing (precision and recall)	5	1	5	Scopus already indexes preprints from the 5 largest preprints servers and attributes preprints to researchers
Ability to link preprints and final published articles	4	2	8	Similar matching processes already used by some of Elsevier’s solutions
Lack of interest from institutions	5	1	5	Ongoing discussions with stakeholders to confirm their interest