

TERMS OF REFERENCE - FEASIBILITY STUDY ON AN OPEN KNOWLEDGE BASE FOR THE DUTCH RESEARCH COMMUNITY

Date of ToR: 13 July 2020

Assignment

The VSNU, represented by the [Taskforce on Responsible Management of Research Information and Data](#), seeks a contractor to undertake a Feasibility Study on an Open Knowledge Base (OKB) for the Dutch research community, and to inspire and sharpen institutional perspectives and support for the concept of an Open Knowledge Base

The key goal of the Feasibility Study is to assess the feasibility of an Open Knowledge Base within the context of different options and make related recommendations pertaining to specific factors (e.g., governance, technical architecture, scope).

Expected deliverable

The ultimate deliverable is a report that addresses the key goal. It should take a bird's-eye perspective, that helps senior decision makers and stakeholders to take decisions about progressing with an Open Knowledge Base. The report should include an assessment of the feasibility of an Open Knowledge Base in the context of the different options and specific recommendations according to the factors outlined in Section 2. It should also include an initial evaluation of the institutional support for the concept of an Open Knowledge Base.

Furthermore, we expect the following deliverables:

Deliverable **01**: discussion document

Minimal requirements: methodology, agree on exact scope and deliverables, outline of study report, planned interviews list of interviewees.

Planning: deadline M1 (September = month 1), online meeting for presentation and discussion with supervisory committee, by 15th September 2020.

Deliverable **02**: first progress report (document and optional slide deck)

Minimal requirements: analysis of three options; indication SWOT

Planning: M2, by 30th October 2020, online meeting with supervisory committee.

Deliverable **03**: slide deck as input for the 18th November Rectors Conference meeting

Minimal requirements: outline of study scope, methodology, preliminary results.

Planning: M3, by 9th November 2020.

Deliverable **04**: second progress report (document and optional slide deck)

Minimal requirement: draft final report

Planning: M3, online meeting with supervisory committee by 30th November 2020.

Deliverable **05**: final report

Minimal requirement: study report according to index agreed upon at the start meeting in September 2020, 1.5 pager executive summary, final layout.

Planning: M4, by 14th December 2020.



Timelines for Proposal

Please send your **proposal by email** to both [Darco Jansen](#) and [Robert van der Vooren](#) by Sunday 9th August 2020 at the latest. Selection of the contractor will be end of August 2020 latest. In the meantime, the following members of the Taskforce can be contacted for clarification on these Terms of Reference:

[Sarah de Rijcke](#) and [Magchiel Bijsterbosch](#) between 15th and 24th July and [Darco Jansen](#) from 3rd August onwards. Questions and responses will be shared among all consultants.

Supervisory Committee

Chair: Wilco Hazeleger, Dean Geosciences, University Utrecht

Members

- Sarah de Rijcke, Professor of Science, Technology and Innovation Studies / Scientific Director at Centre for Science and Technology Studies (CWTS), University of Leiden
- Magchiel Bijsterbosch, Advisor, SURF
- Alastair Dunning, Head, Research Services at TU Delft Library
- Maurice van der Feesten, Innovation Manager Research Services, Vrije Universiteit Amsterdam
- Darco Jansen, manager Open Science & Open Access, VSNU

Core Competencies

We expect the contractor to be familiar with:

- the challenges in scholarly communication and the management of responsible use of research information and data, both in the Netherlands and abroad;
- business modelling based on support level;
- the principles and functioning of underlying conventional and linked data technologies.

Requirements

The proposal should clarify the methodology, the stakeholders that need to be contacted, corresponding timeframe and a break-down of costs. The methodology should minimally include

- elaborated options according to the factors outlined in section 2.
- a SWOT analyses and business modelling for the different chosen options.

Payment

Maximum budget available is 58,000 EUR for the whole assignment including all costs, excluding 21% VAT. VAT reverse charge to VSNU might apply.

Location

All activities can be done online.

Language

Deliverables should be in English. Interviews, if part of the methodology, can be in English or Dutch.



Section 1: Background

Background on an Open Knowledge Base

Open Science is having a dramatic effect on scholarly communications. Publishing models are changing; new forms of scholarly output are created; FAIR data offers opportunities for new forms of data sharing and analysis. Above all, Open Science is propelling deep-seated change in how scientific endeavour is recognised and rewarded.

Embedding Open science demands appropriate infrastructures. Sustainable infrastructures are one part of this. But they must also respect the transparency of interactions between public and private partners in scholarly communications, particularly as the publishing industry focuses more on data analytics services. How data related to publications and other scholarly output is handled and analysed has a crucial impact on judgements about scientists' research success.

Traditionally, universities collect publication data, data from their CRIS and (alt)metrics data from heterogeneous sources. They then store this data in the data infrastructures they are currently familiar with. Maintaining this diversity of practices and systems is hugely expensive; especially as universities use the same data but the knowledge and expertise is scattered in sub-niches catering for custom solutions. Secondly, metadata related to scholarly communications has been managed in discrete, unconnected, closed, commercial systems. Such collections of data have been closely tied to the interface to query the data. This restricts the power of the data – whoever creates the interface determines what types of questions can be asked.

An Open Knowledge Base (OKB) counters both these practices. Firstly, it proposes that universities start using a single data structure. This is very challenging, but there are also strong advantages from building on both the shared human knowledge and the technical resources at universities. Secondly, an OKB also separates the interface from the data. It opens up and connects the underlying metadata to other sources of metadata in a publicly governed infrastructure. Such an approach allows greater freedom – analysis of the data is no longer restricted by the specific way the interface was designed nor limited to querying one set of metadata. Such openness makes the OKB flexible about the type of data it incorporates and when – other data providers with different datasets can connect or incorporate their data at a date that suits them. The openness also allows third parties to build specific interfaces and different services on top of the OKB.

For the field of scholarly communication, such an ambitious metadata infrastructure would connect all sorts of entities, each with clear identifiers. Researchers, articles, books, datasets, research projects, research grants, organisations, organisational units, citations etc could all form part of a national OKB that connects to other OKBs. It would also help create enriched data, which could then be fed back into the OKB.

Such a richness of metadata would be a springboard for an array of services and tools to provide new analyses and insights on the evolution of scholarly communication in the Netherlands. At a more fundamental level, it offers the public domain the opportunity to take responsibility for the data that has a crucial bearing on the progress of the domain of science and the scientists that work within it. The [principles](#) proposed by the [Dutch Taskforce on](#)



[Responsible Management of Research Information and Data](#) attempt to enshrine that responsibility - giving universities, medical centres and other research institutes the agency to determine how data may (or may not) be used.

The best current example of an Open Knowledge Base for scholarly communication is developed by [OpenAIRE](#).

The Netherlands Context

Multiple trends within the Netherlands point to the advantage of creating an Open Knowledge Base. Policy makers and researcher communities require transparency of data and algorithms for responsible decision-making and evaluation; libraries are exploring how infrastructure can offer greater agency in their missions; publishers wish to explore innovative services for fairer metrics for research intelligence and scholarly communication services with high quality Dutch (meta) content; and researchers expect rapid and trusted access to research outputs and related metrics.

However, the idea of an Open Knowledge Base remains ambiguous and needs further elucidation. It is clear (from the above-mentioned principles) what some basic aspects of an Open Knowledge Base might contain. Collecting and aligning metadata (and perhaps full text) from the research output of Dutch authors, enriched with linked data relationships, ontologies and identifiers from Dutch research and funding institutions would form a part of such an OKB.

Yet the feasibility of embedding an OKB within the Dutch research landscape needs to be assessed. Who would pay for, govern and manage such an OKB? Could it maintain widespread support from institutions? How would it be constructed technically? Would it be a trusted source for third parties (publishers and others) to build reliable services upon? Such questions underpin this call for a Feasibility Study.

Section 2: Options and Factors for an Open Knowledge Base

Three options for an Open Knowledge Base

This feasibility study should not result in a yes/no answer. Rather, it should assess the feasibility of an OKB in the context of three options. These options outline the organisational nature of the OKB, marking different degrees of public and private sector involvement. They should not be taken as absolutes, but rather as mileposts on a continuum. The most feasible solution may sit between the stated options, depending on the different factors.

Option 0 – Do not create an OKB; rely on current institutional systems provided by commercial organisations. Should this option appear most likely within the feasibility study, the consequences of following this option need to be articulated.

Option 1 - The OKB as a single shared public / private venture with collective responsibilities between the public sector and one or more commercial organisations

Option 2 - A strong public OKB steered and financed by public organisations only



Factors to be addressed

Some factors relating to the creation and development of an OKB are described, along with some specific questions. These are for guidance only. Not all questions need to be answered. Moreover, there may be other factors and related questions that the feasibility study should address. It is expected that during the first month the scope is set between contractor and supervisory committee (as part of the first deliverable).

Governance and finances

Who would steer the direction of the OKB? How would any board reflect the multiple research institutions contributing to the OKB? Who would pay for such an OKB, for both start-up and ongoing costs? To make an OKB effective, it would require the ongoing participation of every research institution in NL – how would the business model ensure that? And who would do the day-to-day management of the OKB? What should be the role of commercial organisations contributing to the OKB and its underlying principles. Should they have a stake in the governance of an OKB? To what extent are universities already contributing to the development of aspects of an OKB? (e.g. via staff expertise and input to existing CRIS systems), and to what extent could this be rechannelled into creating an OKB?

Scope of Metadata Collection

The potential scope of an OKB is huge. Each content type has their own specific metadata schemes. These schemes evolve over time. How are different metadata types incorporated over time? Article metadata first? Then datasets, code, funding grants, projects, organisations, authors, journals? What about different versions of metadata schemes, need all backlog records be converted?

Quality, Provenance and Trust

Would the metadata in the OKB be sufficient to underpin high-quality services? What schema would need to be created for the different sorts of metadata? What critical mass of metadata would be required to create engaging services? What kind of metadata alignment and enrichment would need to be undertaken? Would that be done centrally or by institutions and publishers? What costs would be associated with that? Would the costs be ongoing? Should provenance to the original supplier of the metadata and metadata enrichments be attributed?

Service development and commercial engagement

What incentives would there be for commercial partners to a) provide metadata and b) build services on top of the OKB? Would the investment to develop such services simply lead to one or two big companies dominating the service offer? Would they compete with services not relying on the OKB? What would happen to enriched data created by commercial companies? Would it be returned to the OKB? Would the resulting services be of use to all contributing members? Could the members develop their own services independent of commercial offerings?

Technical architecture of an OKB

There are two approaches to the technical construction of an OKB. A centralised data store offers certain advantages, e.g. in speed of data access, simpler management structure. But a federated structure, where different parts of the data are held in managed in multiple



locations, can be a better way of increasing buy in from partners and helps avoid creating a single point of failure.

Implementation timeline: Lean or Big Bang

When implementing the OKB, should we first carefully design the full stack of the infrastructure, and solve all the questions within the grand information architecture? Or let it grow organically, and start with collecting the metadata in the formats that is already legally available according to the publishing contracts? Can we do both in parallel; start collecting, and start designing?

International Connection

Is it feasible to create a successful OKB on a national scale? Or does an OKB require an international approach to create a critical mass of data? Under what conditions would other countries be willing to connect to a Dutch OKB?