An Open Knowledge Base for the Netherlands

Report of a Community Workshop

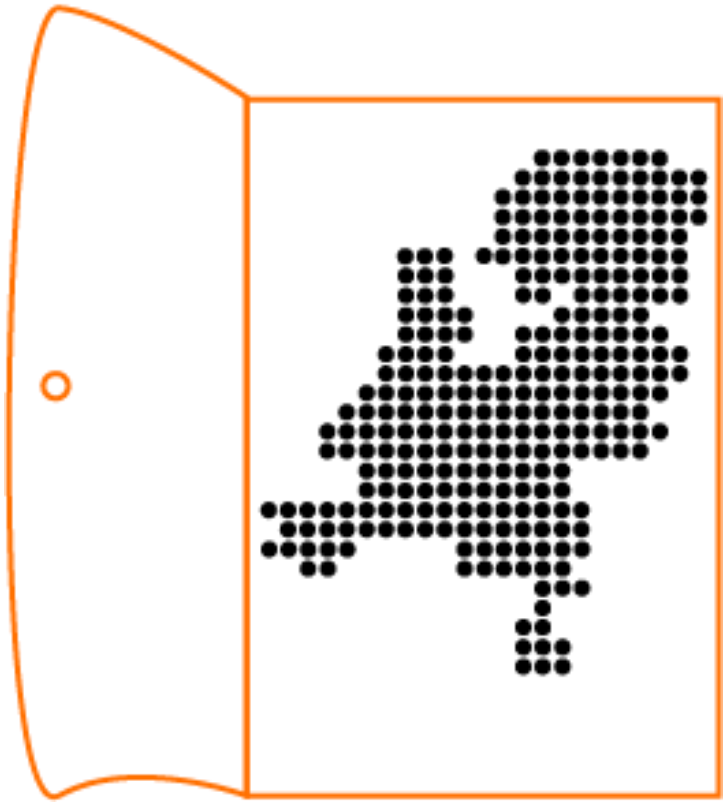


Image adapted from original by Fien Robbe at the Noun Project under a CC BY License

1 Executive Summary

The concept of developing a national Open Knowledge Base for the Netherlands (NL-OKB) has been proposed in response to the strategic needs of the research community in the Netherlands. In parallel with the work of the Dutch Taskforce on Responsible Management of Research Information and Data addressing these opportunities and the preparation of a feasibility study by Dialogic, interested stakeholders were convened with the goal of identifying expert and user-community interests in and need for an NL-OKB. The goal in convening a workshop was:

1. To gather evidence on the feasibility of an NL-OKB
2. To test the community interest and appetite for developing an NL-OKB
3. To identify a practical pathway forward towards startup and implementation of an NL-OKB

Over two days, 19-20 November 2020, 35 participants representing national and international organisations met in a virtual workshop. This included representatives of VSNU, NWO, NFU, SURF, DANS, CWTS and a range of Netherlands institutions alongside international stakeholders such as Crossref, ORCID, OpenAIRE, DataCite, SPARC North America, Jisc, UKRI and others. This workshop provided a parallel strand of convening to the work of the VSNU Task Force and the Feasibility Study for an OKB commissioned from Dialogic.

There was strong support for an NL-OKB amongst the assembled group. The group as a whole was strongly in favour of the development of an NL-OKB run on behalf of and controlled by the academic community. Of those present, virtually all indicated they had a direct stake and interest in supporting the development of an NL-OKB. The level of enthusiasm and good-will for supporting such a development was remarkable. International participants were also keen to see efforts in the Netherlands succeed as an exemplar to be drawn upon.

The assembled group reached a series of consensus conclusions, that taken together provide the beginnings of a roadmap for further development.

- An NL-OKB is technically feasible
- A range of data resources and technology systems are already available, that could form the basis of a rapidly developing prototype
- Data integration projects in the startup phase should be built around selected use-cases, and there was broad consensus that two useful use cases for an initial focus were:
    - Demonstrating a capacity to provide information on open access performance at the institutional (including medical centre) and discipline level, with deduplication of outputs across Netherlands institutions
    - Supporting the future development of the SEP 2021-2027 through providing frameworks for more diverse indicators, tracking a wider range of outputs and activities
- The core of the value for an NL-OKB is in providing rich research information that enables stakeholders to evaluate based on their own values. Initial steps should focus on this “values” proposition
- Development efforts should be rapid but can be phased. A number of initial activities could take place in an agile manner in parallel and do not require full community consensus on issues like organisational form prior to commencing
- However, in such an agile process it will critical to identify up front a consensus on the transition conditions to move from startup to production/maintenance is crucial

2 Table of Contents

3 Motivation for an Open Knowledge Base in the Netherlands

The Netherlands is unique in terms of its research culture, the coordination amongst its research institutions. Alongside the scale and influence of its research the Netherlands is also a leader with respect to its focus on reforming research practice to support open science and scholarship. In addition, the Netherlands has an ambitious agenda for reforming the evaluation of research, as demonstrated by its new Strategy Evaluation Protocol (SEP) and a widely supported national initiative to improve the recognition and rewards of academics.

The Netherlands has its own specific issues where socially robust research can provide solutions, including specific environmental challenges as a result of climate change, the need for intensive agriculture on saline land, and a unique cultural heritage as a result of the intertwined histories of language, religion, culture and immigration. All of these require a research enterprise well-placed to respond to both international and local needs.

Research can contribute to better solutions to these and other complex problems if knowledge is shared as widely and as quickly as possible. The National Open Science Plan (2017) notes that "...society as a whole should benefit from publicly funded research…" and that "...[an] innovative open model needs to be developed to enable target audiences such as SMEs, municipalities and the 'ordinary citizen' to access research results". Key stakeholders have noted that putting these "...shared ambitions into practice requires a modernisation of the system of recognition and rewards" involving "...a culture change as well as national and international coordination between all parties involved". Such coordination requires shared, trusted and transparent information. In turn the Dutch Taskforce on Responsible Management of Research Information and Datasets notes that the "...increasing interwovenness of information about research (research intelligence) and research itself raises a number of challenging issues both for users and for producers of this information".

It is in this context that the draft guidelines of the Dutch Taskforce on Responsible Management of Research Information and Datasets set out an ambition for a national Open Knowledge Base (OKB). We believe that it is technically, financially, and politically feasible for such a resource to be built today, for that resource to be sustainable and cost-effective and for control to be vested with representatives of the Dutch research and wider community. More specifically, the Netherlands is amongst the best-placed countries internationally to deliver a well-managed, innovative and user-focussed resource. Achieving this requires vision, coordination amongst the key players, and a commitment from a sufficient group of Dutch and international stakeholders to work collectively.

To complement the other efforts in scoping this problem CWTS in collaboration with staff from the Curtin Open Knowledge Initiative, TU Delft, University of Utrecht Library and SURF convened a workshop of national and international experts to define pragmatic next steps forward for a community-governed NL-OKB. Over two days, 35 experts representing national and international expertise on data, governance and sustainability, and user needs identified the important issues and first steps to take on realising this goal. A particular focus was to survey the context, institutions, and user needs of the Netherlands specifically to identify opportunities for efficient and effective development and implementation and to showcase Dutch capacities to the rest of the world.

3.1 Background

Amidst broad international efforts in addressing the evolving role of research evaluation, the Position Paper by the Dutch Knowledge Coalition provides a roadmap for the Netherlands. This contribution outlines a modernization strategy to refocus recognition towards a broader range of contributions to science and society, to widening the net for academic talent, and to promote quality over quantity.

Stimulating open science is one of the other main goals of the initiative. As the VSNU "Guiding Principles on Management of Research Information and Data" from the Responsible Research Information Taskforce note, the "...increasing interwovenness of information about research (research intelligence) and research itself raises a number of challenging issues...It is the collective responsibility of all actors in the scientific and scholarly system to manage these issues in accordance with research ethics and public values". Importantly, this component is always about both enhancing open knowledge creation itself as well as about enacting open infrastructures.

This will be achieved by ensuring the creation and retention of information about research. This information is available in the form of metadata, small bits of descriptive information embedded in resources and outputs. Ownership of this metadata by public institutions, enduring access, trusted provenance, interoperability and community governance, as articulated in five of the six Guiding Principles, can underpin the efforts taken in the Netherlands to reform research evaluation.

The sixth principle, open collaboration with the market, requires an open and honest conversation about the requirements of Dutch stakeholders, the capacities of relevant institutions and providers and the management of power relationships with third-party providers. The framework agreement between the Association of Universities in the Netherlands (VSNU), the Netherlands Federation of University Medical Centres (NFU), the Dutch Research Council (NWO) and Elsevier envisions a substantial step forward in realising a key objective in the Netherlands' open science plan--full open access to Elsevier journals for articles published by Dutch authors. This progress on open access is indeed an achievement that breaks a previously long-held negotiations stalemate.

The agreement between Dutch research organisations and Elsevier has also focussed the attention of the Dutch research community on issues of governance and control of infrastructures for research information. A set of stakeholders including lead organisations (VSNU, NFU, NWO), existing infrastructures (SURF, DANS, NARCIS), individual institutions and researchers have argued for the importance of an OKB at the centre of a national infrastructure for research information in the Netherlands. Our workshop reinforced the critical importance of carefully considering the scope, governance and control of the various aspects of an OKB as well as supporting in broad terms the importance of strong guiding principles building on those provided by the Taskforce.

The agreement between VSNU, NFU, NWO and Elsevier envisions a series of pilots for open science analytics and these are defined in broad terms. It is essential to define how these potential pilots relate to the vision for an OKB and to ensure that the final infrastructures provided adhere to, and build on the detail of, the Guiding Principles, so that the final product is a coherent infrastructure system for the Netherlands and not a set of competing parallel services.

Central to this is the question of how the characteristics of an operational NL-OKB will align with Dutch Open Science ambitions? At this crucial point in the implementation of open science, the stakes are high. Flexible and responsive resources are required that provide the intelligence resources to support strategic decisions in the planning, funding and operations of research. This requires a level of ownership, influence oversight, and control of the data and analytics for all key Netherlands stakeholders. Delivering and exceeding the requirements set out in the Guiding Principles requires that we directly address the question of whether a community-led Open Knowledge Base is feasible. What, if any, form of industry/academy relationship would ensure community ownership and control of our core capacities for research intelligence? What does this look like from an academy standpoint? What are the strategic considerations of all stakeholders involved (university leadership, funders, libraries)?

Broadly speaking, we therefore need to address the following questions. Is it technically, financially and politically feasible for the community of stakeholders to design, deploy and maintain a national OKB? What scope does this have? Who are its stakeholders? How are its relations to third-party providers (including the proposed pilots) managed? Several of these questions make up the scope of the independent feasibility study commissioned by the Taskforce. The present document reports on an explicitly community-led initiative to address needs, use cases, existing capacities, and possible routes forward.

3.2 Issues and Use Cases

A wide range of concerns have been raised in relation to control over a central infrastructure for research as an OKB. Here we describe three main issues and concerns that relate to important use cases for an OKB.

3.2.1 The 'Values' Proposition

Research is defined by a set of values held by practitioners, trainees, and the institutions that support them. At the core of these are working in the wider public interest to improve society, transparency, openness, reliability, and increasingly

recognised the importance of creativity and diversity. These values are held and protected through public institutions, including KNAW, VSNU, NWO, universities, scholarly societies and others in the context of the Netherlands.

A further important value and quality of publicly supported work is ensuring transparency and quality of information to key stakeholders on the different ways in which their investment leads to societal benefits. This is far beyond simple return on investment in financial terms and requires trusted, credible, transparent, flexible and responsive information that can be deployed for evaluation that faces both inwards and outwards. In an increasingly data-driven world, there is an increasing risk of being misled by biased or inappropriate data fed into unmonitored black-box algorithms. In order to be capable of expressing and acting on their core values, Dutch institutions must be in control of the underlying data, the prioritisation of new kinds of data, the processing and presentation of that data and internal critique of its biases.

> *An OKB must be open, transparent, and have priorities and strategy set by the community to deliver the capacity for expressing and acting on the values of the Netherlands' public research institutions and to inform on the national societal benefit.*

3.2.2 SEP 2021-2027: The Evaluation Use Case

The Strategy Evaluation Protocol (SEP) is a world-leading progressive approach to research evaluation, providing flexibility, contextualisation and a commitment to a diversity of approaches. As a result it also faces challenges of scaling, consistency and credibility. These are common criticisms of approaches that are perceived to be primarily narrative driven. The shift to the next generation in the Strategy Evaluation Protocol (SEP 2021-2027) focuses on defining strategy and identifying qualitative and quantitative indicators that can inform on that strategy. There is an opportunity to combine the SEP with the benefits of quantitative and robust evaluation indicators *provided* that such indicators can be obtained through a system that flexibly allows those being evaluated to select, design and adapt the indicators that fully represent their work. International best in class evaluation approaches take a similar approach in co-design of evaluation indicators and face similar challenges in a lack of infrastructure to support such a flexible approach.

An OKB that is built to reflect the diversity of practice, approach and research targets is not just a useful means of supporting the SEP approach, but is critical to enabling it reach its full potential and address its critics. Such a system must be capable of enhancing its data capabilities and possible approaches to indicator development. It must also be responsive to the needs of innovative researchers engaging in new types of research activities and generating new forms of outputs. It cannot be rooted in systems that are backwards looking and prestige driven if it is to support the transformative cultural change envisaged by the *Room for everyone's talent* paper and demanded by innovative researchers internationally. Netherlands' research and education institutions have led international progress on developing more comprehensive and inclusive visions for recognition and rewards. Realising these ambitions requires the flexible and responsive evidence base that an OKB could provide.

> *An OKB must support researchers and their organisations in describing how they want their research to be evaluated, providing flexibility in indicator development and design, so as to capture the diverse qualities of research activities and outputs, and enable the continuous improvement of the scope and diversity of activities and outputs collected and monitored.*

3.2.3 Open Access Evaluation at the Institutional Level

Institutions in the Netherlands are amongst the global leaders in the shift to open access and open research practices more generally. They are also seeking to provide data to demonstrate the progress made in delivering open access. Even in the narrow space of information to evaluate progress towards open access and inform policy design and implementation, it has been difficult to provide a consistent view across and between institutions. As goals broaden from the narrow focus on open access to open science practices more generally, this will only become more challenging. There are a number of issues contributing to this, but at the centre of those are a lack of coordination across research institutions on gathering data on research outputs, and the limitations of information on open access provided by traditional commercial information providers.

An OKB must provide two capacities to address the immediate need for institutional needs for open access evaluation: 1) It must successfully coordinate amongst institutions to integrate and deduplicate the full diversity of Dutch research outputs. 2) It must have the capacity to determine open access status across the full diversity of outputs. Even modest developments of this use case require a general capacity to enrich and aggregate data about a wider diversity of research outputs.

4 The Workshop

Responding to the current environment a group of national and international researchers and research supporters came together to identify how to advance the agenda for a community-owned OKB in the Netherlands. We believe that it is technically, financially, and politically feasible for such a resource to be built today, for that resource to be sustainable and cost-effective and for control to be vested with representatives of the Dutch research and wider community. Our goal in convening a workshop was:

1. To gather evidence on the feasibility of an NL-OKB
2. To test the community interest and appetite for developing an NL-OKB
3. To identify a practical pathway forward towards startup and implementation of an NL-OKB

We convened 35 participants representing national and international organisations, and a wide range of expertise in data, technical architecture design, user needs and use cases, sustainability and funding, and community governance. This included representatives of VSNU, NWO, NFU, SURF, DANS, CWTS and a range of Netherlands institutions alongside international stakeholders such as Crossref, ORCID, OpenAIRE, DataCite, SPARC North America, Jisc, UKRI and others.

Over two days in November 2020 we worked in three main streams to identify user needs, existing capacities and opportunities and next steps. This work proceeded across three main streams: one focussing on user needs and architectural design; one focussed on existing data capabilities and future needs alongside technical design; and finally one focussed on the development of community governance and sustainability frameworks, with a specific focus on existing frameworks and success stories in the Netherlands.

4.1 User Stream

The User Stream focussed on developing use cases, through identifying for different stakeholder groups what broad themes and specific questions an OKB could support. The initial goal was to develop specific questions for stakeholder groups and to map these with respect to the data sources required and the impact that answering them would have. The members of the stream would then explore the various data sources needed to answer those specific questions, and the effort required to bring those data sources together.

Seven stakeholder groups were identified alongside a set of characteristic questions that are summarised here. These were not comprehensive and future work should expand the coverage of stakeholder groups more broadly to ensure systems the support an increasing engagement across society. A more detailed collection of questions was made for each group that was discussed. Those questions and themes uncovered by the group for each stakeholder group are available in this Miro Board.

4.1.1 Individual Researchers, Research Groups & Departments

Researchers work at an international and community level. A national base has less interest for them outside of required evaluation systems (which are generally driven by University management or other stakeholder groups, see below). Many of the use cases for researchers related to placing their work, or the work of others in a global context and therefore showed a need for an OKB to be linked to other data sources that provide an international perspective.

4.1.2 Deans and University Senior Management

There was clear agreement that the current shift in recognition and rewards (led by senior management) within the Dutch university system requires the research information infrastructure provided by an OKB. This provides an important policy driver for the OKB. How can an OKB provide the information that is needed to support and continually develop the evaluation of research, education, leadership and social impact?

4.1.3 Research Support (Libraries, Funding Support)

The clearest set of immediate use cases sat in this group. Many of the use cases overlap with current use cases that are made via CRIS and other systems. One example of this is the open access monitoring use case discussed above. There is an urgent need for movement here - will an OKB provide a national conglomeration of scholarly metadata that currently sits siloed in university CRIS systems?

4.1.4 Publishers, Commercial Companies

There were no representatives from commercial companies amongst this stream. Third parties with access to an OKB could clearly provide specialised analysis services and could additionally see an opportunity in data enrichment services. Indeed an OKB might support an increase in innovation in this space by reducing barriers to market entry.

4.1.5 Funding Bodies, Government

Funders have a range of use cases, including evaluation of funding programs, setting strategic priorities, and assessing policy implementation. Each of these requires improved data provision, and new and improved mechanisms for enriching internal data sources. Government requirements are generally assumed to be based around reporting, and often the primary use case is regarded as a reduced reporting burden for research organisations. However, it is also possible to imagine an opportunity for providing more timely, directional and strategic information to the government and encouraging deeper interactions between research activities and government where appropriate.

4.1.6 Meta-researchers

Meta-researchers do research that aims to understand science system dynamics and/or to support the various stakeholder groups discussed above to make better decisions. Thus, all the use cases presented above are of interest for this group.

4.1.7 Conclusions

An OKB would need to grow with time; serving the straightforward use cases first and then supporting more complex ones. The different use cases listed above could be analysed in more detail to map the impact (how useful would answering this question be?) against the required effort and to identify, categorise and prioritise the use cases that should be tackled in each phase of the development of an OKB. It is important not to limit the capabilities of an OKB by architectural decisions (e.g., federated vs. centralised models) made in an early stage of development.

The group was concerned to ensure that existing sources, systems and standards were fully used, and additionally noted dependencies on the sustainability of the web of other providers of data, systems and standards. In focussing on specific use cases there was a clear demand for managing risks, both in terms of data availability, but also dependencies on software and tools, resulting in a clear preference for open approaches. It is important to clarify what data and tools need to be in the public domain for an OKB to succeed.

Central to the group's conclusions was the need for a clear focus on specific use cases, and a clear means of selecting and prioritising those use cases that would engage stakeholders and particularly those stakeholders that would need to invest in the success of an OKB. Ensuring a link between the development of the OKB and national policy, via the selection of new use cases to direct development was seen as one route to maximising engagement.

4.2 Data Stream

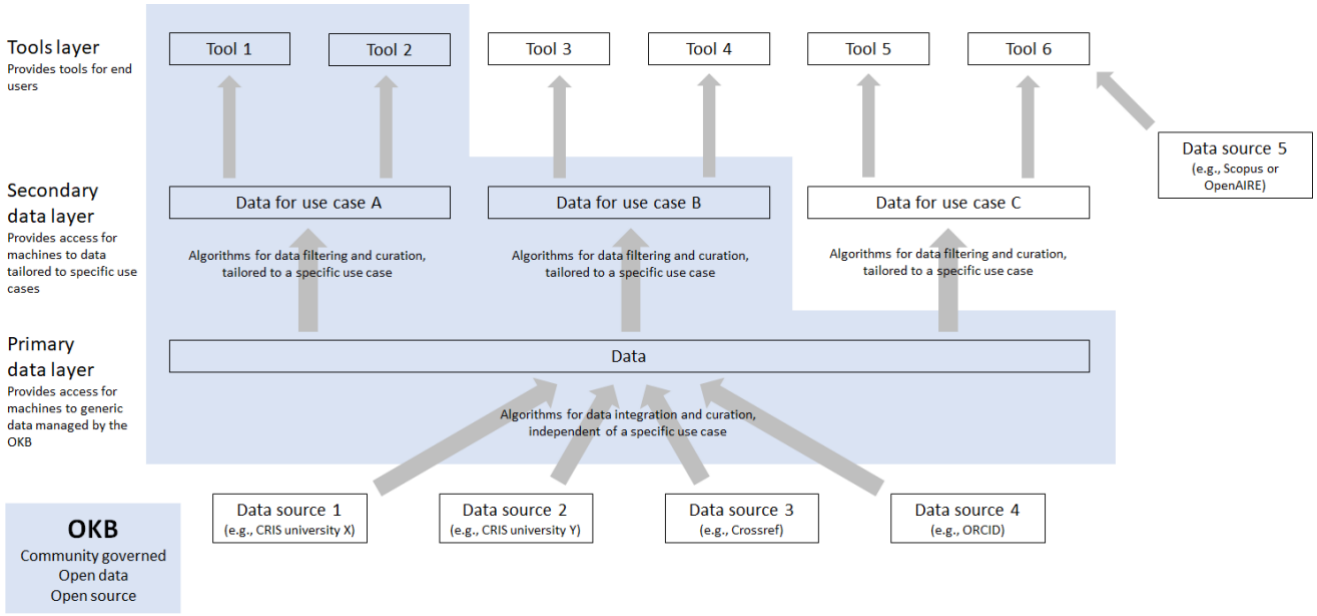
In the Data Stream in the workshop, group members extensively discussed the conceptual and technical design of an OKB and the data to be included in an OKB. Below we report the main conclusions.

4.2.1 Conceptual Design of an OKB

An OKB is an infrastructure that makes information about the Dutch research system openly available. The information is made available in a machine readable way. It can be integrated into other infrastructures. The information can also be ingested by tools that make it available to end users. Such tools may be part of the OKB, but they do not need to be. The OKB is governed by the Dutch research community.

Based on the discussions in the workshop, a conceptual design of an OKB has been prepared, shown in the figure below. In this figure, the area colored in light blue represents the OKB. The rest of the figure provides a partial representation of the external environment of the OKB. The conceptual design can be summarized as follows:

- The OKB consists of two (or more) data layers and may also include a tools layer.
- The OKB ingests data from a number of data sources, such as Crossref, DataCite, ORCID (and the PID graph more generally), and CRIS systems of universities. The data is fed into the primary data layer of the OKB. In the process of feeding the data into the primary data layer, the data is integrated and curated only to the extent that this is possible without assuming a specific use case for the data. This may mean that inconsistencies in the data cannot always be resolved. Data provenance is essential in the primary data layer. It ensures that the origins of each data point are clear.
- To prepare the data for specific use cases, the data is fed from the primary data layer into the secondary data layer. This involves filtering and additional curation of the data. Filtering and curation can be done separately for each use case covered by the OKB.
- The OKB may also include tools for end users. Each tool addresses a particular use case and obtains the relevant data from the secondary data layer.
- The data layers of the OKB make their data openly available in a machine readable way, ideally under a CC0 public domain waiver (while being mindful to avoid undermining the sustainability of upstream data providers). The data in the primary data layer offers the highest flexibility, since the data has not yet been tailored to a specific use case. The data in the secondary data layer is easier to use, since it has been tailored to a specific use case, but it offers less flexibility. The data sources from which the OKB ingests data must allow the data to be made openly available. Ingesting data from proprietary data sources is therefore difficult. There may be conditions under which this is possible, but this needs further investigation. For instance, CRIS systems of universities are often enriched with data from proprietary data sources, and it might perhaps be possible to make this data openly available in the OKB.
- The tools in the tools layer of the OKB are free to use.
- The algorithms used by the OKB are open source, so that all data processing performed by the OKB is fully transparent. Anyone may reuse the algorithms and build on them.
- Third parties may use data obtained from the OKB in their own infrastructures and tools. These infrastructures and tools are not part of the OKB. They do not need to be community governed and do not need to make their data openly available, and their algorithms do not need to be open source. Tools do not need to be free to use. They may integrate data from the OKB with data from other data sources. These other data sources do not need to be open.



4.2.2 Technical Design of an OKB

There are different ways in which the conceptual design presented above can be translated into a technical design. A full assessment of the different ways in which an OKB can be technically implemented could not be made in the workshop. Nevertheless, we offer the following recommendations:

- Based on the discussions in the workshop, a Semantic Web approach seems most suitable for implementing the primary data layer of the OKB. Using such an approach, the primary data layer can be relatively easily extended over time in order to cover new types of data. We suggest a centralized approach for storing the data in the primary data layer, although a partly federated approach may also be considered.
- The secondary data layer aims to make data easily available for specific use cases. The technical implementation can be done separately for each use case. For many use cases, a relational approach with centralized data storage seems preferable. Such an approach decreases the complexity of querying the data for downstream users.

4.2.3 Data Included in an OKB

An important aim of an OKB should be to provide data not only on journal articles (typically with a DOI), but also on other types of outputs (e.g., data sets, software, policy reports, books, articles in Dutch language journals and news media, etc.), activities (e.g., in public engagement and education) and contributorship roles (e.g., through the CRedIT taxonomy). This will support the broader evaluation approaches currently being developed and implemented by Dutch research organisations.

As much as possible, data sources should be expanded to include non-traditional research outputs and activities. Where these are not currently covered in external sources (and there are limited incentives for third party providers to address them), an OKB should provide agency and incentive to researchers to report on and be credited for these, e.g. through their universities' CRIS systems where appropriate and through open systems to support researchers not affiliated with universities.

Two aspects deserve specific attention:

- Where existing taxonomies and identifiers exist for non-traditional research output, these can be used in an OKB. Where they do not exist, an OKB should provide at least basic identifiers to allow inclusion of all outputs and activities in the semantic or relational structure of the OKB.

- Collection and inclusion of data on non-traditional research outputs and activities should be accompanied by a critical discussion on the use and use cases of these data. Rather than aiming to include as much data as possible, researcher agency in the selection and prioritization of data to be included might be a better way to support both responsible evaluation and concerns around privacy and openness, especially around data that are not otherwise openly available.

4.3 Governance and Sustainability Stream

The key to a functioning community-based OKB that serves the interests of stakeholders across the Netherlands is achieving productive engagement across those stakeholders. This requires trust, both in the data and services provided by the OKB, and also in the strategic directions and priorities that "the OKB" and its various stakeholder groups choose to invest in. Appropriate governance and community building arrangements are therefore crucial.

"Governance" means many things to different people. In the context of the current discussion we take governance to mean *the concrete set of arrangements: including legal form, community representation, financial sustainability and management, processes for setting priorities, and for decision making; that creates trust and engagement in the community of stakeholders*.

Important to note is that this definition encompasses how revenue is raised (and from whom) and how decisions are made to spend it, as well as questions of legal form and representation. Also crucial is that trust and engagement is a goal beyond those directly represented in governance structures or that contribute financially.

The *Principles for Open Scholarly Infrastructures* cover some aspects of these concerns, specifically focussing on three areas: governance, finances, and "insurance" which covers aspects of the decision making process. The six Guiding Principles proposed by the Research Information Taskforce provide high level requirements that would need to be satisfied for community acceptance in the Netherlands. The Values and Principles Framework and Checklist developed by Educopia provides a set of values-driven areas for focus as well as specific signals or processes. Most of these are positioned at a high level, although the Educopia Checklist has a range of very specific suggestions on practice.

4.3.1 Workshop Goals

Our key goals for the workshop were to identify key stakeholder who would need to be engaged on issues relating to governance, identifying existing organisations and frameworks in the context of the Netherlands that might take a lead role in the development and maintenance of an OKB and identify possible sustainability models, again with a focus on previous successes in the Netherlands.

The group included international experts, representatives of sustainable international and national infrastructures, as well as key stakeholders from the Netherlands.

4.3.2 Thinking in terms of roles and responsibilities, rather than organisations

A key finding for the governance group was that there is a range of roles that require some form of governance, coordination and/or sustaining:

1. The data and technical infrastructure systems that make up the core of an OKB
2. The set of services offered that utilise, or contribute to enriching, the data resource. This may include non-community contributions, in which the governance question is one of how relations between the community and non-community actors are managed
3. The forum or organisations where the rules and regulations that govern all these interactions, internal and external, are set
4. A framework for gathering input on the prioritisation of future technical developments and resource allocation, as well as ensuring ongoing efforts on inclusion and diversity

5. An oversight role that ensure that the rules and regulations are being followed by all relevant actors, both those within the community and those outside

The value in separating and articulating these different roles is that it helps us to avoid thinking in terms of specific or imagined organisations. While some of these roles could clearly be combined into a single organisation, and efficiencies might be gained by doing this, there is also value in a separation of responsibilities, for instance with the oversight role envisaged in #5 being taken by a separate organisation to that implementing relations on services or maintaining a technical infrastructure.

Such a separation of responsibilities may also be particularly appropriate in the context of the Netherlands, where existing bodies including SURF, DANS, NPOS, NWO, KB and UKB have specific expertise and experience in differing aspects of these roles. More specifically there is a history of these bodies taking particular roles and existing networks of trust and relationships between them.

4.3.3 The feasibility of startup projects and moving rapidly towards startup

There was a strong consensus on the need to move rapidly on the development and feasibility testing for a community-led OKB. There was additionally some concern about institutional inertia and that building a complete consensus amongst all stakeholders would take too much time. There was enthusiasm for developing a set of startup projects that could tackle specific challenges, with the goal of resolving the most challenging issues to build towards a consensus. This would involve identifying the key issues that would need to be resolved, scoping them as projects, with the aim of delivering initial findings that would provide the evidence needed for the leadership of key institutions to support a more formal deployment phase.

The feasibility of this approach in the context of the Netherlands builds on the experience of a number of projects that have developed into longer-term systems. Three examples to draw from are NARCIS, HBO Kennisbank and in particular Studiekeuze123. In each case a development project, including work packages that examined governance and sustainability models, identified and built consensus around technical, governance and sustainability models that were carried into a deployment and maintenance phase. Both Studiekeuze123 and HBO Kennisbank show some level of the separation of responsibilities described above.

4.3.4 Defining the transition from startup to maintenance

An agile approach as proposed above requires that there be clarity and consensus on the conditions to be met for a transition from a series of startup phase projects to develop a “minimal viable product” into a deployment and maintenance phase. These conditions will need to be defined in advance and some more work will be required on consensus building amongst them but it is clear that they will include the following in some form:

1. A model (or models) for the governance of the different roles and relations between them
2. A credible resourcing model that is sustainable in the medium-long term
3. Value propositions articulated sufficiently for a critical mass of diverse stakeholders to support the proposed resourcing model (and an outline of the need for value propositions for a broader range of stakeholders)
4. A technical proof of feasibility for sufficient use cases to support those value propositions
5. Initial draft rules of engagement (building specificity and implementation details from the Task Force Guiding Principles) that define the scope of the community of stakeholders, principles for interaction with non-community members, and some definition of future scope

While there are interdependencies between these deliverables, significant work can be carried out on them in parallel. In addition the Netherlands context creates some strong guidance for specific routes to take and experience to draw on given the existing set of institutions, frameworks and sustainability models. Within the group many concerns were raised about particular pitfalls, conflict of interest, or other traps that similar efforts, or the academic ecosystem more generally, have suffered from. Therefore gathering international expertise on these issues, including contract and

procurement design, forms of oversight, separation of responsibilities, and critical principles of community ownership will be a key aspect of this phase.

4.3.5 Inclusion and Exclusion - Stakeholders and orientation

The group noted that there were significant issues involving inclusion and exclusion. Within the stream and across the workshop participants there was a concern to ensure that inclusion was a core value. This covers both a concern to ensure engagement with the full range of stakeholders, including those who are not commonly included (broader civil society, students) and those who experience exclusion (for example based on gender, socioeconomic background, immigration status, ethnic background etc). The overarching message was to build inclusion into the governance model and operating strategy. The opportunity to address inclusion in the beginning, while a complicating factor, is much easier than fixing it later.

Issues of exclusion and inclusion also relate to the objects considered in scope for an OKB. The dominant bibliographic data sources have continuing issues with: (a) disciplinary bias (b) diversity of output types c) diversity of contributor roles (contributors who are not listed as authors) d) mission orientation and selectivity that is opposed to comprehensive coverage. Many of these issues reflect or compound systemic gender and socio economic inequalities.

The orientation towards inclusion is not however universal. By definition the idea of “community” governance requires exclusion of those who are not part of the community. Inclusion does not imply a lack of standards or requirements for community members. The Principles for Open Scholarly Infrastructures uses the term “non-discriminatory membership”, which draws on the approach adopted by Crossref where members must be a publisher - which is defined broadly - and agree to abide by the set of member responsibilities. These responsibilities are explicit, and members are required to adhere to them. They are also designed and implemented with a goal of minimizing and mitigating any exclusionary effects they may have.

In the context of an NL-OKB community membership will mean agreeing to a set of conditions. These conditions may be difficult for third party commercial suppliers to agree to, although there is no a priori reason to exclude them on principle. Nonetheless two major players (Clarivate and Elsevier) have such significant market dominance that relations with them would likely need to be on the basis of strict vendor-procurement basis, meaning that they were explicitly excluded from community governance. Careful design of procurement and contractual arrangements for relations with third parties (which are not limited to Clarivate and Elsevier, but could include other players such as Microsoft, Google, and Digital Science, and which should also include a consideration of dependency on cloud providers such as Amazon, Microsoft, Google, Digital Ocean and others) will be an important part of OKB design, and is an area where community initiatives are traditionally very weak, both in terms of expertise and delivery.

Such an approach enables an OKB to design for inclusion from a market competition standpoint, by clarifying the basis and scope of competitive offerings and innovation. This could involve e.g. reducing barriers to participate in developing services ‘on top’ of the OKB. This can also address issues of market dominance through developing more accessible and inclusive procurement workflows and procedures to support the engagement and innovation of smaller suppliers.

4.3.6 Governance and Sustainability - It is possible to move fast and to move together

The group as a whole felt that it was possible to make a credible case for appropriate resourcing of a set of startup projects to demonstrate the possibilities of a community-led NL-OKB and to define the next phase. The group was strongly supportive of the concept of an NL-OKB. While there are many issues to be addressed, the community support, the demonstrated value propositions and the existing frameworks and experience in the context of the Netherlands led the group to be confident that a well designed and coordinated set of startup projects could provide a strong consensus route to the creation of a community governed OKB for the Netherlands.

5 Synthesis

5.1 Strong support for an NL-OKB

Amongst workshop participants there was strong support for the need and value of a community led NL-OKB. In addition, almost all participants indicated a direct personal or organisational commitment to taking part in a startup project with the goal of creating an NL-OKB. Amongst those with less interest in direct participation (largely those representing international interests) there was nonetheless an interest in tracking progress and learning from the experience of the Netherlands. The assembled community had a strong view that such a system is important and would be critical to realising the open science aspirations of the Netherlands.

5.2 An NL-OKB is technically feasible

The data group developed a conceptual design that provides both flexibility and scalability for data management and handling. While the details of technical design will require further work in the startup phase, likely through a series of tests and demonstration projects tied to specific use-cases, the conceptual framework was broadly agreed by participants to be achievable and to provide substantial value and future opportunity for further development.

5.3 A range of data resources and technology systems are already available

The data and user streams identified a range of data sources and technology systems that are, or could be, fully community owned and controlled. These build on open bibliographic data sources (Crossref, DataCite, Microsoft Academic), existing technology and information initiatives (OpenAIRE, Curtin Open Knowledge Initiative, Wikidata) and a range of emerging standards and approaches for data integration.

5.4 Data integration projects in the startup phase should be built around selected use-cases

There was a strong view that startup phase demonstration and feasibility projects must be designed to demonstrate significant concrete value propositions to key stakeholders, and should be built around existing needs. There should also be an effort to ensure that some of these use-cases include forward looking needs, such as enhancing the diversity of research outputs and activities for which useful information can be captured.

There was a broad consensus on two use cases for initial projects:

- Demonstrating a capacity to provide information on open access performance at the institutional (including medical centre) and discipline level, with deduplication of outputs across Netherlands institutions. This should include at minimum due diligence on identifying gaps and issues for specific forms of output (such as books), for outputs in Dutch, and for coverage of disciplines.
- Supporting the future development of the SEP 2021-2027 through providing frameworks for more diverse indicators, tracking a wider range of outputs and activities. These could include peer review, public engagement, various alternative forms of outputs and communications as well as tracking use and engagement with research. In the context of an initial startup project this should focus on providing flexibility to end users with respect to the choice and construction of indicators associated with traditional outputs, in particular articles in scholarly journals. In addition, the initial focus should be on providing data and indicators on a carefully selected set of non-traditional outputs or activities.

5.5 Focus on the “values” proposition

Although not strictly a use case, representatives of several lead organisations emphasised the importance of making the case for community capacity to enact its values. A core benefit of an NL-OKB is that control and influence over its design, data holdings and default indicators is crucial for the Netherlands research community and society more broadly to be able to control and guide its own evaluation processes.

5.6 Development efforts should be rapid but phased

There was a strong view that moving forward to an NL-OKB should be done rapidly and that a number of startup projects could be carried out in parallel. Technical feasibility projects should be connected to specific use cases and aim to identify future challenges and issues for data integration and enrichment in practical and concrete terms. Building stakeholder engagement will be an important early-phase project, alongside developing a concrete proposal for governance, management and funding arrangements. These can build on the successful experience of previous projects to infrastructure developments in the Netherlands and follow a similar path.

5.7 Consensus on the transition conditions to move from startup to production/maintenance is crucial

Critical to success through a series of parallel projects will be defining early the conditions for community acceptance, and therefore the process for moving from the startup phase to the maintenance phase. While some of these requirements can already be outlined it will be an important early goal to build a consensus amongst key stakeholders on the detailed acceptance criteria for a transition, as well as the conditions under which a decision not to proceed might be taken.

An important aspect of this is to take an intelligently staged approach to stakeholder engagement. Defining first the set of stakeholders that are needed to move such a project forward and defining early on a pathway that continues to broaden the set of engaged stakeholders, with the goal of creating a truly national system with benefits for a wide range of interests and groups across the country.

6 Colophon

This report was prepared by the workshop conveners: Cameron Neylon (Curtin University), Magchiel Bijsterbosch (SURF), Alastair Dunning (TU Delft), Bianca Kramer (Utrecht University), Sarah de Rijcke (CWTS), Clifford Tatum (SURF, CWTS) and Ludo Waltman (CWTS).

This report may be re-used under a Creative Common Attribution v4 License.

The report text was based on contributions and edits from the workshop participants. The list of participants is below.

Alastair Dunning	TU Delft
Andrea Mannocci	OpenAIRE
Armand Guicherit	TU Delft
Bianca Kramer	Utrecht University
Cameron Neylon	COKI
Chris Baars	DANS
Christian Hauschke	TIB
Claudio Aspesi	Independent consultant
Clifford Tatum	CWTS/SURF
Darco Jansen	VSNU
Hans de Jonge	NWO
Herbert van de Sompel	DANS
Ivan Heibi	OpenCitations
John Doove	SURF
Lucy Montgomery	COKI
Ludo Waltman	CWTS
Magchiel Bijsterbosch	SURF
Maria Cruz	NWO
Martin Fenner	DataCite
Matthew Buys	DataCite
Maurice Vanderfeesten	VU Amsterdam
Max Kemman	Dialogic
Michele de Bonis	OpenAIRE
Mogens Sandfær	TU Denmark
Najko Jahn	OpenAIRE
Nees Jan van Eck	CWTS
Neil Grindley	JISC
Neil Jacobs	UKRI
Nick Veenstra	Eindhoven University of Technology
Petra Drankier	NFU
Rachael Lammey	Crossref
Rachel Bruce	UKRI/JISC
Robbin te Velde	Dialogic
Rodrigo Costas	CWTS
Sander Bosch	VU Amsterdam
Sandro La Bruzzo	OpenAIRE
Sarah de Rijcke	CWTS
Silvio Peroni	OpenCitations
Simon Porter	Digital Science
Thed van Leeuwen	CWTS
Tom Demeranville	ORCID