

Guiding Principles on Management of Research Information and Data

Final version 11 May 2020

1. Preamble [why a set of guiding principles]

This document sets out the guiding principles from the Dutch Research Institutions in their collaboration with commercial parties in developing new services related to research intelligence and scholarly communication. 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 the collective responsibility of all actors in the scientific and scholarly system to manage these issues in accordance with research ethics and public values. And as such agree on principles and conditions under which information about research can be stored, (re)used and enriched, avoiding undesired network or platform effects.

2. General principles

Metadata related to scholarly communications has been managed in discrete, unconnected and sometimes closed, commercial systems. Such collections of data have been closely tied to the interface to query the data. The collective ambition for the Dutch Research Institutions is that this metadata is open for others to access, reuse, enrich and described according to existing open standards, identifiers, ontologies and thesauri. In general the Dutch Research Institutions want to engage with players in a healthy market, fostering joint innovation, while protecting the sovereignty of scholarly capital and preserving academic autonomy through a community-owned governance system. The following general principles for the management of research information are designed to guide the future collaboration with commercial third party providers.

GP1. Ownership of (meta)data

The ownership of research output (articles and research data) and related metadata resides with the institutions and researchers. Knowledge institutes maintain control over any (meta)data stored into services related to research intelligence and scholarly communication. Furthermore, knowledge institutes have the responsibility to manage any data that is functionally part of information e.g. research policy or evaluation, including derived, enriched or transient (meta)data.

GP2. Enduring access

The scholarly enterprise is constantly producing new outputs and interactions. Data and (metadata) descriptions are being added, through manual data entry but also through (algorithmical) enrichment. Without the availability of this data, full (information) integrity is lost and transparency and accountability is impaired.

Services that collect, interact with or use scholarly output in any shape or form must facilitate complete, non-discriminatory and enduring access to primary metadata and derivative data without functional, technical, legal or financial limitations.

GP3. Trusted and transparent provenance

All metadata held must have transparent, trusted provenance. It must be clear to stakeholders in any service how and where metadata was derived, especially metadata that has been collected algorithmically. Technical, legal and operational agreements for metadata sharing should be laid down in a transparent agreement system.

GP4. Interoperability as part of community owned governance

Interoperability is a precondition for realising a trusted, transparent, decentralised infrastructure. Continuous innovation of information services and technology make for an ever-changing framework. The rigorous enforcement of agreements made in advance cripples innovation while agreements that are too loose create a false sense of trust.

The conditions for interoperability take shape in an open, inclusive dialogue between all involved parties. This includes both public knowledge institutes as well as any party that delivers a service which handles scholarly output or (meta)data on research dynamics of publicly funded research performed by Dutch knowledge institutes. By engaging in this dialogue, parties commit themselves to implement agreements and sharing in the costs that invariably come with making and maintaining infrastructures interoperable. Standardized scholarly metadata which is accessible and separated from associated services and tools allows for competition without platform or vendor lock-ins.

GP5. Open collaboration with the market

It is appreciated that services from market players can add value to research. An open collaboration model allows the creation of network effects between services when they are outsourced to different market players while preventing (de facto) monopolies such as gravitational hubs and “winner-takes-all” scenarios. This requires a strong information position of and collective action by knowledge institutions.

Knowledge institutions and market players engage in an open collaboration where innovation, competition and public value are recognised and respected cornerstones.

GP6. Community owned governance

The data sets and (decentralised) infrastructure is considered a common pool resource. Ownership of this resource lies with public knowledge institutes. However no single party can hold absolute control over the entire model. This requires a structure fostering an inclusive sustainable decision-making process, to manage information between stakeholders, deal with conflict and steer new developments.

The principles, standards and collaborations underpinning Dutch scholarly capital are governed principally by knowledge institutes representing the academic community. The governance will be open to market players and other stakeholders. The model will adopt the subsidiarity principle and include arrangements for collective choice, monitoring and enforcement of agreements and conflict resolution mechanisms.

4. Towards an Open Knowledge Base

The principles above are part of the ambition to come to a federated *metadata* infrastructure that would connect all sorts of entities, each with clear identifiers. This ensures the findability of researchers, articles, books, datasets, research projects, research grants, organisations, organisational units, citations etc. 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.

As for now, the Dutch research institutions have the ambition to create themselves an Open Knowledge Base (OKB) with two main features. Firstly, it separates the interface from the data. Secondly, it opens up and connects the underlying metadata to other sources of metadata. Such an approach allows much greater freedom - users are no longer restricted by the specific manner in which the interface was designed nor restricted 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 moment that suits them. The openness also allows third parties to build specific interfaces and different services on top of the OKB.

An upcoming feasibility study will clarify the needs, opportunities and threats, finances and governance for such an open, community driven infrastructure of connected metadata in relation to principles as set above. Projects with commercial third party providers must contribute as building blocks to the development of an OKB according to public values and principles.

5. Implementation for development projects with commercial parties

Currently some Dutch Research Institutions undertake a number of projects with commercial vendors to further develop and implement new services related to research intelligence and scholarly communication. It is our aim that each party adheres to the above principles during the execution of existing and potential future projects. And that as such both the Dutch Research institutions and any commercial third party provider commit themselves to implementing the above principles and ambitions and to ensuring that those projects will contribute to the creation of a common open infrastructure on connected metadata. The OKB is a possible long-term aim/condition.

Governance

The implications of the principles as outlined are explicitly not elaborated in detail mainly because of:

- a) lack of one single uniform standard and
- b) shifting implications over the intended contract period.

Therefore a strong governance structure is needed to guide the implementation of these principles into specific conditions for projects developing such services.

The principles, standards and collaborations that support the Dutch scholarly capital require management by means of community owned governance (GP6). As such, elaborations and

tightening of the above principles will have to apply to all contracts. Consequently, a relation must exist between the foreseen community owned governance structure of the OKB ambition and the steering committee or executive board during the execution of each contract with a commercial third-party provider. This amongst other include that

- Project proposals will be based on an open consultation of the Dutch research community under elaborated conditions set by the foreseen national community owned governance structure.
- A steering committee in which knowledge institutes have the casting vote decides if a project can start based on description of work amongst others describing the implementation of principles above and how projects from different research areas contribute to the ambitions above related to OKB.
- A steering committee should also be able to approve a project if not all of the above conditions apply.
- Before a project can be approved by the steering committee, all participating organisations should have agreed to collaborate in advance.
- Long term development agreements need a solid legal exit plan. Such an exit plan should be included into the contractual agreement. The exit plan will have to come into effect especially if the above principles are not sufficiently implemented or if sovereignty of scholarly capital cannot be guaranteed in a way. Exit plan amongst others includes continuity of access, data portability and ownership of metadata.

In addition, the following rules apply as well:

1. Any party needs to be able to take part in a project
2. At least three research institutions need to take part in a project before it can take off.
3. Depending on the focus of the project a competitor needs to partake.

Scope for the development of new services

It is the intention of Dutch Research Institutes to work towards an overview of products and services needed, based on the public ambitions (e.g. OKB). Hereby a level playing field will be created such that the development of new services is open to all third parties. The intention with new services and systems is to aim to make science more transparent, efficient, inclusive and collaborative, allowing for the broadest possible audiences to have the opportunity to participate, to make use of and to contribute to the scientific process.

The development of such services will focus on a variety of themes related to research intelligence and scholarly communications, such as

- *Easy and rewarding engagement for researchers to open their research output.* Minimise effort for researchers. Services that assist researchers with this task by providing as much as possible all relevant outputs and correct metadata in which each datapoint is only entered once in one of the connected systems.

- *Enabling overviews, analytics and 3rd party use cases in an open way.*
Create maximal value. Any available datapoint that is relevant to a stakeholder is made available to them in the most current and most useful way. For example alternative impact metrics, strategic research intelligence, e.g. emerging research fields, funding opportunities
- *Connect different CRIS systems in a seamless way.*
Aggregating service to create a complete overview of Dutch outputs and relevant identifiers to enable specific facet-based analytics. Open to ingest many different data models, for example those from Scopus, Web of Science, PubMed, Dimensions, Crossref but also CERIF.
- *New workflows, eg. preregistration, open peer review.*

Acknowledgements

This document is advice from the [Dutch Taskforce on Responsible Management of Research Information and Data](#). This Taskforce was established early 2020 by the Association of Universities in the Netherlands (VSNU), The Netherlands Federation of University Medical Centres (NFU) and The Dutch Research Council (NWO) to address issues around the responsible use of research information and the role of commercial third party providers in particular.