# The Scholarly Communication Platform Framework

A systematic approach for exchanging information about journals and other platforms for scholarly communication

**Bram van den Boomen**
Centre for Science and Technology Studies (CWTS), Leiden University, The Netherlands
b.van.den.boomen@cwts.leidenuniv.nl
https://orcid.org/0009-0005-5127-1200

**Nees Jan van Eck**
Centre for Science and Technology Studies (CWTS), Leiden University, The Netherlands
ecknjpvan@cwts.leidenuniv.nl
https://orcid.org/0000-0001-8448-4521

**Ludo Waltman**
Centre for Science and Technology Studies (CWTS), Leiden University, The Netherlands
waltmanlr@cwts.leidenuniv.nl
https://orcid.org/0000-0001-8249-1752

**Tony Ross-Hellauer**
Institute of Interactive Systems and Data Science, TU Graz, Austria
ross-hellauer@tugraz.at
https://orcid.org/0000-0003-4470-7027

**Serge P.J.M. Horbach**
Danish Centre for Studies in Research and Research Policy, Aarhus University, Denmark
s.horbach@ps.au.dk
https://orcid.org/0000-0003-0406-6261

# 1. Introduction

Keeping track of the many ongoing developments in scholarly communication is a major challenge. While most research is still communicated through peer-reviewed scientific journals, there are lots of developments that aim to innovate the communication of scientific work. Some of these developments focus on enriching or complementing the journal publishing model (e.g., preprinting and open peer review). Others are challenging this model in more fundamental ways by offering alternative approaches to scholarly communication (e.g., F1000, Octopus, and Peer Community In).

Only time will tell which approaches to scholarly communication best suit the needs of the scientific community and, more broadly, of all beneficiaries of scientific work. But to make sure new approaches to scholarly communication get a fair chance to demonstrate their value, it is essential that researchers and other stakeholders have access to information about the many different ways in which scientific work can be communicated.

There are quite a few sources that provide such information. Tools and databases are available that help researchers understand how to make their research openly accessible (DOAJ), how to make sure their work can be posted in a repository or on a preprint server (Sherpa Romeo), how to ensure compliance with funder requirements (Plan S Journal Checker Tool), and how to pick a journal that offers a particular peer review approach (Transpose), that adopts particular open science practices (TOP factor), or that is considered to have a sufficiently high citation impact (Journal Citation Reports).

Unfortunately, however, the landscape of tools and databases has become almost as complex as the landscape of scholarly communication itself. Given this complexity, it seems unrealistic to expect researchers to know which tools or databases to use to obtain specific information. And if researchers do not make use of the information that is available, it is also unattractive for publishers and other actors in the scholarly communication system to invest in making more and better information available.

To address this challenge, we aim to make it easier both to use information that is already available and to make new information available. We do this by introducing the Scholarly Communication Platform Framework. This framework provides a systematic approach for exchanging information about platforms for scholarly communication, more specifically platforms for the publication and/or evaluation of research articles, such as scientific journals, preprint servers, and peer review platforms (in particular platforms for preprint peer

review). The Scholarly Communication Platform Framework can be used to bring information from different databases together and to present this information in an integrated way, for instance on a website. When new databases become available, providing new information about scholarly communication platforms, these databases can be integrated with relative ease.

The Scholarly Communication Platform Framework consists of two core elements. First, the framework includes a data model, the Platform Assertion Document (PAD) data model, that facilitates the exchange of information about scholarly communication platforms in a machine-readable and traceable way. This data model is discussed in Section 2. Second, the Scholarly Communication Platform Framework includes ontologies for describing specific properties of scholarly communication platforms. We introduce one such ontology, the Scholarly Communication Platform Ontology (SCPO). This ontology, discussed in Sections 3, 4, and 5, models publication and evaluation policies of scholarly communication platforms. Additional ontologies can be added to the Scholarly Communication Platform Framework to model other properties of scholarly communication platforms, such as the individuals and organizations that manage a platform, the publication services provided by a platform, the usage of a platform, and so on.

The Scholarly Communication Platform Framework aims to meet the following requirements:

- *Extensible:* The framework is able to cover a wide range of platforms for scholarly communication (e.g., journals, preprint servers, peer review platforms, etc.) and diverse types of information about these platforms (e.g., information about publication and evaluation policies, information about individuals and organizations managing a platform, statistics about the usage of a platform, etc.). The framework can be easily extended to accommodate new developments in scholarly communication.
- *Machine-readable:* The framework enables information to be exchanged in an interoperable way that allows for easy processing and interpretation by computers.
- *Traceable:* The framework enables information to be traced back to the original source (e.g., who provided the information and when was the information provided?).

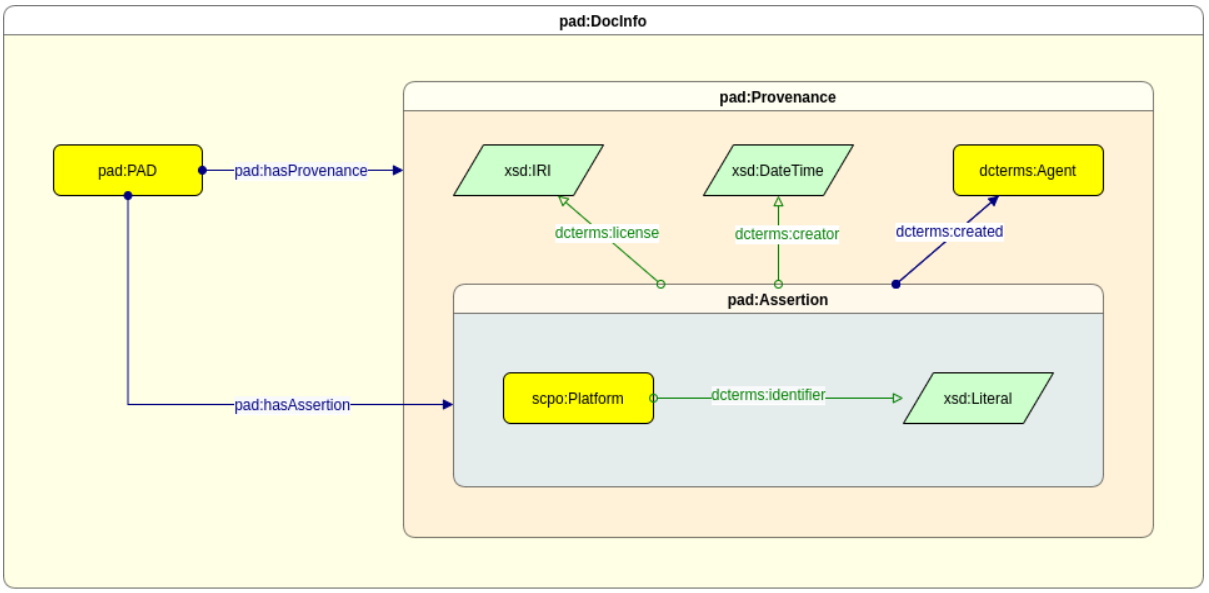
To meet the above requirements, we adopt a Semantic Web approach for the Scholarly Communication Platform Framework. Such an approach aligns closely with the goals of our project. Semantic Web approaches offer a high level of flexibility, providing freedom in modeling the domain of interest and extending the model when necessary. In addition, federation and interoperability are at the core of Semantic Web approaches. This helps to bring information from different databases together and to prevent further fragmentation.

Semantic Web approaches can also be used to cross-check information from different databases. For instance, by linking information about scholarly communication platforms to information about individual research articles, the formal policies of the platforms can be validated against the actual practices at the level of individual articles.

## 2. Platform Assertion Document data model

The goal of the Platform Assertion Document (PAD) data model is to add context to the information about a scholarly communication platform, for instance regarding the origin (provenance) of the information. This context is important to determine whether information can be considered reliable and up-to-date and to understand the conditions under which the information can be used.

The PAD data model is inspired by the model introduced for nanopublications. It uses named graphs to add provenance information. A named graph enables adding a unique identifier to a statement (i.e., to a subject-predicate-object triple). This allows statements to be made about other statements.




| Prefixes | |
|---|---|
| pad: | http://purl.org/job/pad/ |
| scpo: | http://purl.org/job/scpo/ |
| dcterms: | http://purl.org/dc/terms/ |
| xsd: | http://www.w3.org/2001/XMLSchema# |

As shown in the above diagram, a PAD consists of three named graphs:

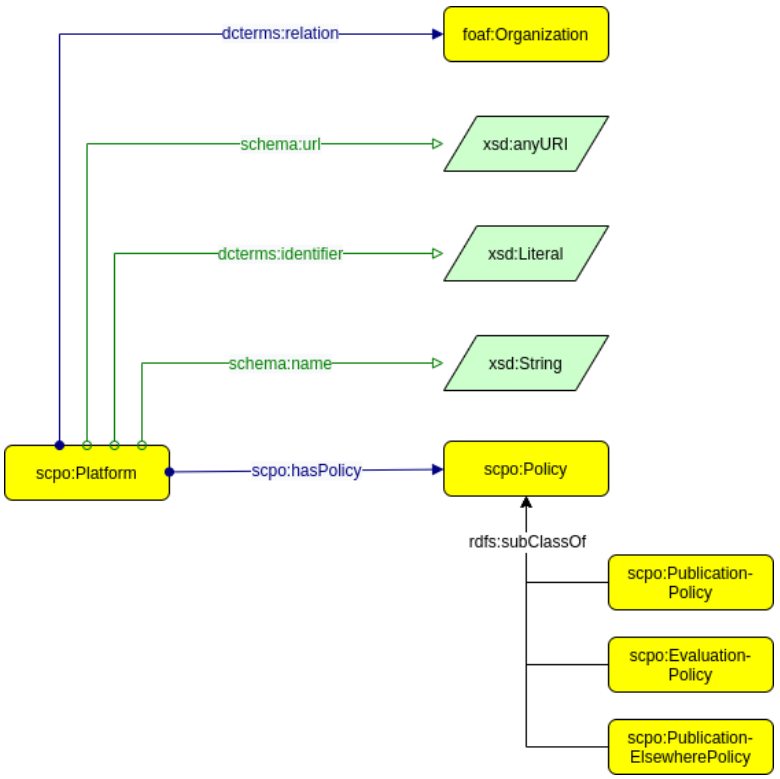
- The assertion graph contains statements about a scholarly communication platform. These statements may be subject to error, interpretation, and change. We therefore refer to these statements as assertions.
- The provenance graph contains statements about the assertion graph. These statements may for instance provide information about the source of the assertions in the assertion graph, the date and time at which the assertions were made, and the conditions under which the information in the assertion graph can be reused.
- The docinfo graph contains statements about the PAD itself. These statements are primarily used to link the named graphs that make up the PAD together. They may also provide information about the code that was used to create the PAD, the person or organization that created the PAD (which may be different from the source of the assertions in the assertion graph), etc.

The PAD data model provides users of a PAD with the information needed to reason about the reliability of the information in the assertion graph, to resolve conflicting information, and to deal with updated information. Example use cases include:

- Discarding information from a particular database because the information is considered unreliable.
- Aggregating information about a platform, and in case of conflicting information, using the most recent information.
- Correcting or updating a PAD by referring to the deprecated PAD in the provenance graph of a new PAD.
- Confirming the information in a PAD by referring to the assertion and provenance graphs of that PAD.

## 3. Scholarly Communication Platform Ontology: Modeling scholarly communication platforms

In this section and the next ones, we introduce the Scholarly Communication Platform Ontology (SCPO). We first discuss how we model a number of general properties of scholarly communication platforms. We then discuss how we model publication and evaluation policies of these platforms.




As shown in the above diagram, an assertion asserts a property of a scholarly communication platform. Examples of properties include:

- An identifier of a platform.
- A name of a platform.
- A URL of a platform.
- An organization managing a platform.
- A policy of a platform, for instance a publication policy (e.g., covering the organization of open access publishing) or an evaluation policy (e.g., covering the organization of peer review). In the next sections, we discuss how we model these policies.

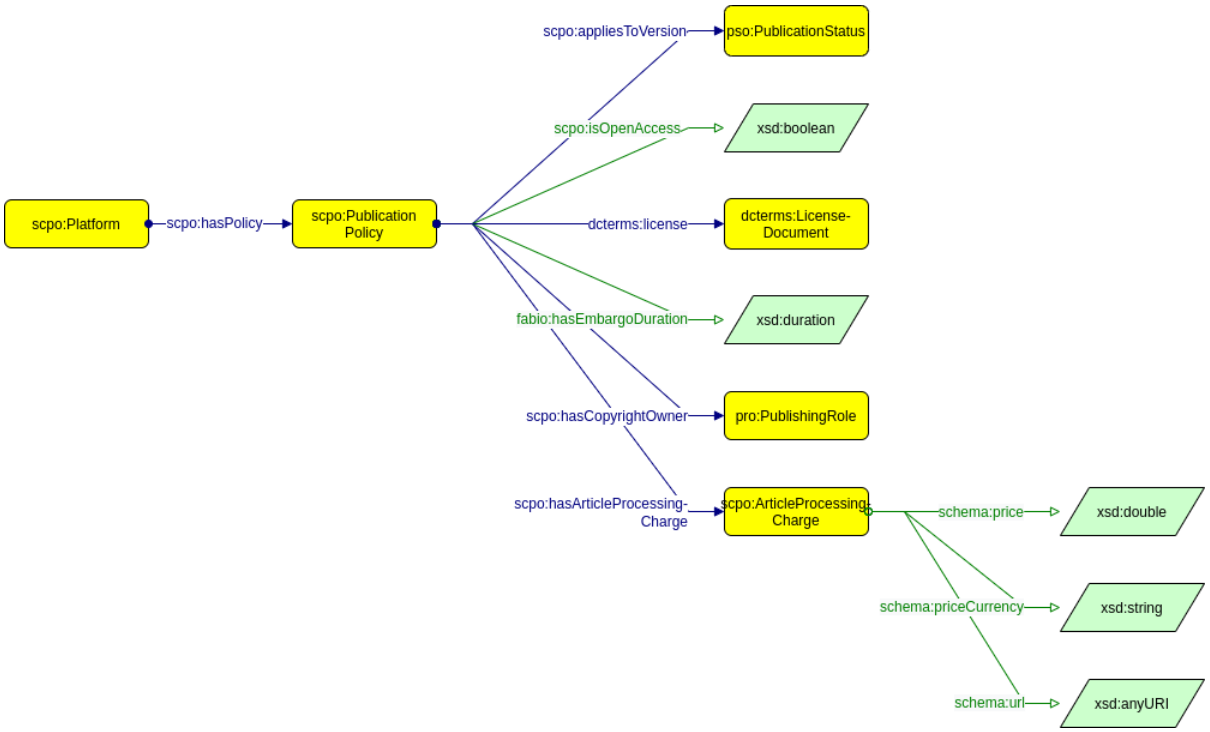
In addition to the properties displayed in the above diagram, there are many other properties that one may want to assert for a scholarly communication platform. Examples include:

- The editorial board of a platform.
- The adoption of the TOP guidelines by a platform.
- Usage and impact statistics for a platform (e.g., turnaround times, number of research articles, number of views, number of citations, and number of social media mentions).

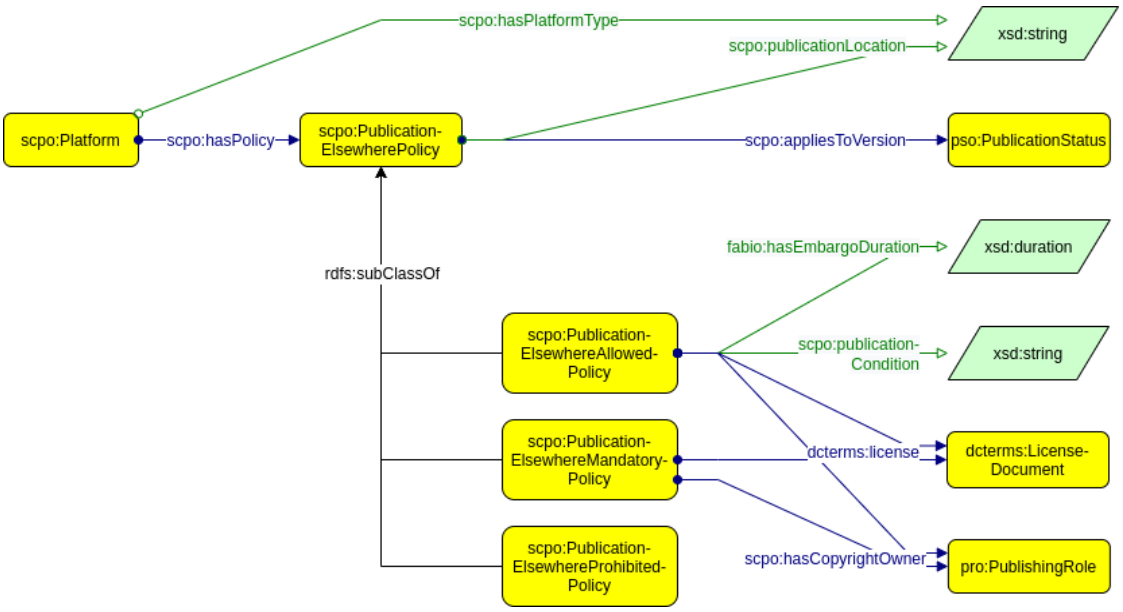
These properties fall outside the scope of the SCPO. Additional ontologies can be created to handle these properties.

# 4. Scholarly Communication Platform Ontology: Modeling publication policies

We now discuss how the SCPO models publication policies of scholarly communication platforms.




As shown in the above diagram, a publication policy describes a policy of a scholarly communication platform for publishing research articles. Some platforms have multiple publication policies. For instance, a hybrid open access journal has a publication policy for closed access articles and a publication policy for open access articles. Other platforms, for instance peer review platforms, do not publish research articles and therefore do not have a publication policy.



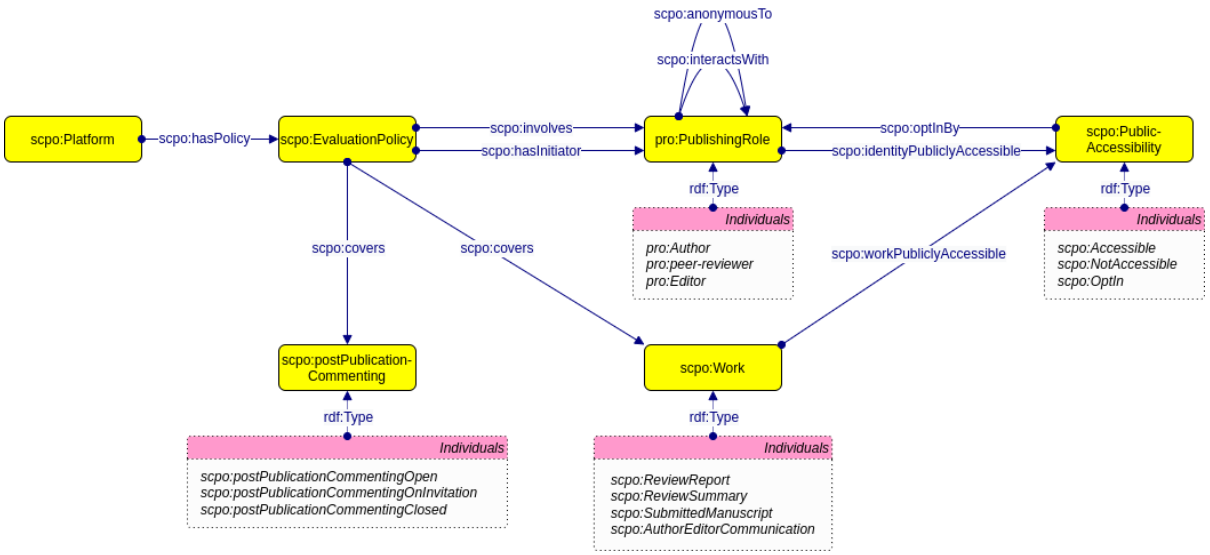

In addition to publication policies, a scholarly communication platform may also have publication elsewhere policies. As shown in the above diagram, a publication elsewhere policy is a policy of a platform that prohibits, allows, or mandates the publication of a research article on other platforms. A platform may have different publication elsewhere policies for different versions of a research article.

The following examples illustrate the idea of a publication elsewhere policy:

- Some journals prohibit the publication of some or all versions of a research article on a preprint server (‘Ingelfinger rule’).
- Some journals allow the publication of some or all versions of a research article on a preprint server.
- Some journals mandate the publication of some or all versions of a research article on a preprint server. *eLife* is a well-known example. Overlay journals are another example.
- Peer review platforms typically mandate the publication of a research article in a journal or on a preprint server. It is not possible to publish a review of an article on these platforms if the article itself has not yet been published.

# 5. Scholarly Communication Platform Ontology: Modeling evaluation policies

Finally, we discuss how the SCPO models evaluation policies of scholarly communication platforms.




As shown in the above diagram, an evaluation policy describes the policy of a scholarly communication platform for evaluating research articles. Evaluation may range from in-depth peer review performed by many journals and peer review platforms to more superficial screening performed by many preprint servers. Some platforms, for instance repositories such as Zenodo, do not perform any evaluation and therefore do not have an evaluation policy. The approach taken in the SCPO to model evaluation policies builds on and extends the STM peer review terminology.

The evaluation of a research article may involve individuals with different roles, for instance authors, reviewers, and editors. These individuals may or may not know each others' identity. In addition, the evaluation of an article may involve different types of documents. In addition to the article itself, the STM peer review terminology for instance distinguishes between review reports, review summaries, and author/editor communication.

Both the identities of the individuals involved in the evaluation of a research article and the documents involved in this may or may not be made public, or they may be made public only with the consent of specific individuals.