

Profile Refinement Service and Pure

March 2017

This document describes the Profile Refinement Service (PRS) and how it works in conjunction with Pure. It also describes how data flows into the system and how to share feedback or report issues regarding publication data.

PRS is a service that enables your instance of Pure to ingest disambiguated Scopus publication data relating to researchers at your institution. It is a service that is offered on top of Pure and integrates seamlessly with the system.

The overall process for PRS is described in Figure 1, which illustrates how master data from Pure (such as Persons) are fed into the PRS cycle.

Master data updates:

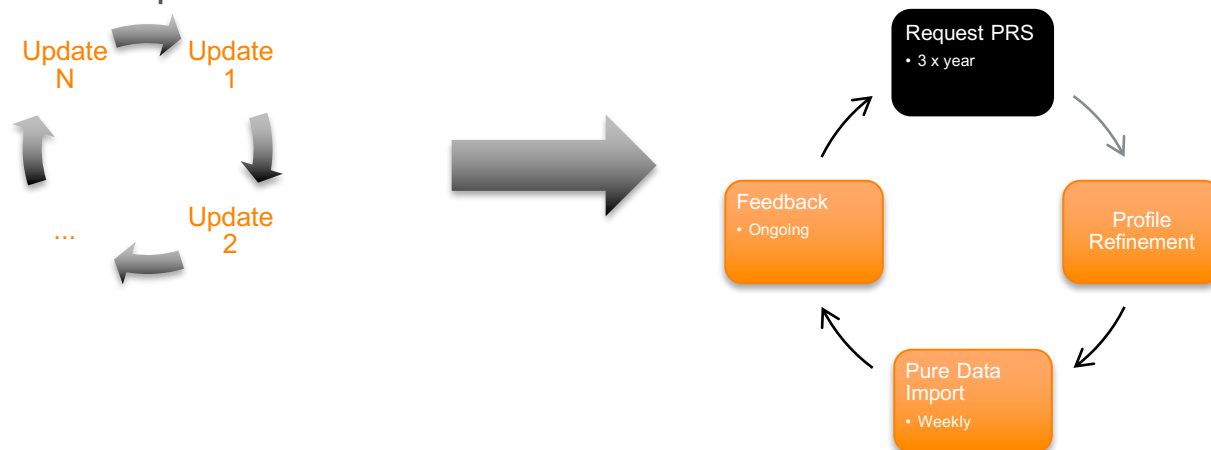


Figure 1: The first step in the PRS process is the submission of a Pure master data snapshot to the PRS. This can be initiated three times a year by contacting the Pure Support Team. Note that you can still update master data as frequently as needed, while the PRS process allows data to be submitted up to three times a year. PRS will commence after data submission and a completion date will be provided upon request. Once PRS is completed, the weekly update will automatically pull updates into Pure. Feedback (data corrections, additions, etc.) may be reported to the PRS Helpdesk on an ongoing basis.

Profile Refinement Service

The purpose of the Profile Refinement Service (PRS) is to build a correct and complete profile for a researcher, making use of their publication information in Scopus. This process populates Pure with a disambiguated publication list for the researcher in question. Disambiguation happens on an ongoing basis and any new Scopus publications that are disambiguated with high confidence are added to the Pure profiles on a weekly basis. To achieve a high degree of accuracy, an automated algorithm is supplemented by manual validation which occurs three times a year.

For successful profile refinement, it is crucial that accurate metadata is provided in Pure, in particular the researcher's full author name, email and organizational affiliations. This metadata is used to perform deep web searches for additional information, such as CVs or publication data, to be certain that the correct author has been identified. Metadata is entered into Pure using a master list (Excel spreadsheet) or using a person/organization synchronization (XML).

The PRS process consists of two parts: the first is an automated algorithm and the second consists of manual checks by Profile Specialists. The initial automated disambiguation is done by running each Scopus publication record through a series of algorithms that identify publication sets with high accuracy. These publication sets are then manually verified and merged by a Profile Specialist into a single profile. In certain cases, ambiguous authors of publications may be sent back to the customer for clarification and further refinement.

When a profile is manually verified, it is fed back into Scopus to merge all the affected publications under a single Scopus author profile. Having a verified Scopus profile supports the automatic assignment of new publications in Pure in the future. As the algorithm assigns new publications to the profile, this data is automatically fed into Pure through a weekly import process. This yields an updated Scopus publication list in Pure. The details of this process are described in the following section.

Profile Refinement Service Data in Pure

PRS data is fed into Pure through the weekly import process, which runs on a fixed schedule. The weekly import process augments the publications with additional information (e.g. affiliation data) as they are imported, handing author and organizational affiliations as follows:

- Every author on the publication in Scopus is reflected on the publication in Pure. If the author is in the list of persons to be profiled, the author is represented as an internal person. If the author is either not on the list to be profiled or is an author not affiliated with your institution, the author is represented as an external person.
- For internal authors, if the publication was written while the author was associated with your institution, an internal organizational unit is related to the publication through the author. Which internal organizational units are related depends on the following rules in this particular order:
 1. If the publication date of the publication falls within the start and end date interval of any affiliation of the person, then this affiliation is selected.
 2. If an affiliation of the person is marked as primary, this affiliation is selected.
 3. If none of the above apply, all current affiliations of the person are selected.
- For internal authors, if the publication was written while the author was associated with another institution, that institution is added as an external organization.
- For external authors, the institution of the author is added as an external organization.

As noted above, publications are first identified by algorithms, and then manually verified three times a year. Both types of publications are loaded into Pure, but by default, they are handled slightly differently in order to allow you to distinguish these:

- Publications that have been manually verified are loaded into Pure in the *Validated* workflow step.
- Publications that have been identified algorithmically, but have not yet been manually verified, are loaded into Pure in the *For Validation* workflow step.

- When a publication is manually verified (during one of the three yearly updates) this will automatically be reflected in Pure, with the workflow changing from *For Validation* to *Validated*. In certain circumstances the manual verification can also lead to the publication being deleted in Pure (if manual verification finds that the publication was wrongly assigned).

By default, the weekly update imports publications from Scopus and makes no further changes to the publication in Pure (besides validation/deletion as described above). However, it is possible to enable updates on already imported publications:

- With this option turned on, Pure will automatically apply updates to publications, e.g. title/abstract corrections, host publication, journal, etc.
- Pure will check each publication for modifications made by end-users (except the *root* user) and preferentially keep these changes over the updates provided by Scopus.

To enable this feature, please contact the Pure Support Team.

In addition, Pure imports citation metrics from Scopus once a month. Citations are assigned to any publications that exist in Scopus and these metrics can be used in Pure in a variety of ways, including showing displaying the h-index in the Pure Portal, as well as for reporting purposes (this requires the Report Module).

As the PRS process requires processing of data, there will be a difference between when the publication is made available on Scopus.com and when it is imported into Pure. In general, a delay of around a month should be expected.

Providing Feedback

If you discover incorrect or missing publication data after the PRS process, please report this back to us by sending an email to sveprofiles@elsevier.com. In order to help us identify and address your issues as fast as possible, please include the following information whenever possible:

- the name of your organization and Pure Person ID to identify the researcher
- a short description stating the nature of the issue
- identifying information for any publications, i.e. title, journal, DOI and any unique identifiers (e.g. Scopus EID, PubMed ID, etc.) in a standard citation format

If submitting batch corrections (10+ publications), we recommended you provide the data in a spreadsheet format to facilitate processing.

The PRS Helpdesk will respond within (1-2) business days and advise on the next steps to handle the inquiry. Feedback is initially assessed and triaged by the PRS Helpdesk. Depending on the nature of the inquiry, it may be forwarded to a different team, however, the PRS Helpdesk will provide regular status updates on the issue. For instance, an issue may be routed to the Scopus Helpdesk if it is determined that corrections need to be made in the Scopus.

Helpdesk Timeframes

Helpdesk guidance timeframes are listed in this section. Note that there is a distinction between PRS issues and Scopus issues in the two tables below. The reason for this is that not all issues can be resolved by the PRS Helpdesk, while others require additional assistance from the Scopus Helpdesk. Timeframes are generally longer when the Scopus Helpdesk must be involved, as fixes sometimes have to be handled at external publisher level.

As a rule of thumb, the following rules apply in terms of Helpdesk involvement:

- If the issue is related to attribution of a publication (e.g. incorrectly linked to a profile or misclassified as internal/external), then it is generally handled by the PRS Helpdesk.
- If the issue involves a fix of publication data (e.g. misspelled title/author, duplicate, incorrect citation, etc.), then it will need to be fixed in Scopus.com by the Scopus Helpdesk.

Generally, any fixes applied by the PRS Helpdesk may take up to a week to become visible in Pure due to the weekly import. The following table lists common PRS issues.

PRS issue	Timeframe
Time to first response	1 business day
Forward to Scopus Helpdesk	1-2 business days
Add/remove publication	≤ 1 week
Internal/external attribution	≤ 1 week
Metadata fix	≤ 1 week

Common Scopus issues are remedied by the Scopus Helpdesk and will become visible after the weekly import in Pure, which is factored into the timeframes listed in the following table.

Scopus issue	Timeframe
Author name correction	4-5 weeks
Incorrect title	4 weeks
Citation correction	4 weeks
Missing/duplicate publication	6-8 weeks
Batch duplicate submission (10+ publications)	6-12 weeks

FAQ

1. What if I am missing a publication or have found a publication that is not mine in Pure?

Please forward the issue to the PRS Helpdesk for validation. They will be in contact with you to solve the issue.

2. Why is my recent publication not listed in Pure?

The publication may not yet be available in the database or it might be due to the publication not being covered by Scopus. Note that from the time a publication becomes visible in Scopus, it will take up to 2-4 weeks before the publication is automatically imported to an existing profile. In the event that you are missing a publication, we welcome you to reach out to the PRS Helpdesk for clarification.

3. What if I identify a content issue in my publication?

Common content issues are missing author name, title corrections and affiliation corrections, etc. Please forward these types of issues to the PRS Helpdesk for validation, where you will be assisted further.

4. What if I discover duplicate publications in Pure?

Publications can be entered into Pure from different data sources other than PRS (e.g. online sources, bulk import or manually), and in some cases this may introduce potential duplicates which are flagged if the title is similar.

Before submitting a list of publications to the PRS Helpdesk, we recommend that you first verify the origin of the duplicates, such that only PRS duplicates are forwarded to the PRS Helpdesk. The origin can be determined by viewing the publication metadata in Pure.

We recommend you handle duplicates as follows depending on origin:

- Non-PRS duplicates can be manually resolved using the resolve/merge feature in Pure.
- If duplicates have different origins, we recommend keeping the PRS version.
- Clear PRS duplicates should be submitted to the PRS Helpdesk for resolution.

5. What if there are any metadata errors (e.g. name, title, email, affiliations) in the Pure profile?

The metadata must be modified in the Pure master data source and reloaded into Pure. Afterwards these changes will be visible in the profile in the Pure Portal.

6. What if I only want to see validated publications in the Pure Portal?

The Pure Portal has configuration options that support the display of only the publications that are in the *Validated* workflow step. Please contact the Pure Support Team to enable this feature.

7. What is the data source for PRS?

Publications that are covered in Scopus, see www.elsevier.com/solutions/scopus/content.

Contact Information

PRS Helpdesk: sveprofiles@elsevier.com

Pure Support: pure-support@atira.dk