



Research Intelligence DataSearch in Pure

Version 5.16

Empowering Knowledge™
© 2019 Elsevier B.V. All rights reserved.



1. DataSearch Integration

1.1. About DataSearch

DataSearch is a search engine for datasets. It allows scientists and researchers to search for many different data types and formats across a variety of domain-specific and cross-domain institutional data repositories and other data sources. Researchers can quickly preview and assess datasets before accessing them in the destination repository.

DataSearch indexes both metadata and the data itself, making it easier to discover datasets. Researchers can preview available datasets and select which to import.

1.2. DataSearch in Pure

Through the integration with DataSearch, datasets can be periodically synchronized into Pure. These datasets can then be validated and are available in Pure for other purposes, such as being shown on the Pure Portal or available in reports.

This supports researchers in their need to access and review the data behind research, and research administrators to keep track of the volume of datasets created at their institution.

The key features of DataSearch available in Pure are:

- Datasets are automatically synchronized from a wide range of external repositories.
- Dataset content in Pure is kept up to date with changes from the source repository using the 'available updates' feature.
- Metadata is imported into Pure, with links to the repository where the dataset is stored.

The following fields are retrieved from DataSearch, and are available in the Pure Dataset record as read only:

- Title
- Description
- Persons (with the role 'Creator' assigned by default)
- Organizational unit
- DOI
- Files (represented as links back to the original record)
- Keywords

1.3. Access options for DataSearch

There is an option to use a paid (enriched) version of the API, which is available from Pure version 5.16.2 and requires an API key. To obtain a key, your institution should contact the Pure Support Team, and will be eventually able to register directly via the [DataSearch website](#).

API type	Description
Free API	Research managers can search for datasets using their institution's Mendeley Institution ID.
Paid API	Research managers can search for datasets using their institution's Mendeley Institution ID, Scopus ID or SciVal affiliation ID. Once datasets are matched to an institution, further matching is performed on the author, using the Scopus, Scival or ORCID author ID.

1.4. Using DataSearch in Pure

Enable the DataSearch integration

To enable using DataSearch to search for datasets in Pure:

1. Navigate to Administrator > Datasets > Workflow step configuration.
2. Ensure that Workflow steps are enabled for datasets.

Tip

If they are not enabled, you can use the **Change workflow step** button to launch the **Workflow step configuration window**, and set the **Workflow step** setting to **Enabled**.

3. Navigate to Administrator > Datasets > DataSearch.
4. Set the **DataSearch integration** toggle to **On**.
5. In the **Institution IDs** field, enter your institution's Mendeley ID.

Note

If you are using the paid API, you can also enter a Scopus ID or SciVal ID.

6. In the **Default managing organization** section, select an organizational unit. This will be used when a managing organization cannot be determined from the metadata imported from DataSearch.
7. In the **DataSearch sources** section, enter the exact names of the data sources that should be searched via DataSearch, separated by commas (.). If you leave this field blank, all of the available sources will be searched by DataSearch.

Tip

You can find the exact names of the sources by clicking **Show available sources**. The full list of available sources is shown.

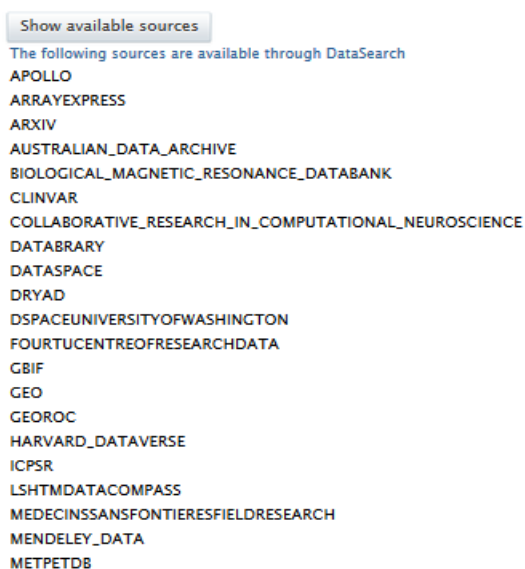


Figure 1. A screenshot of some of the sources returned. You can view the exact source list in your Pure.

8. Click **Save**.

The DataSearch integration has now been configured. Each time the DataSearch job runs, Dataset import candidates are presented to the user on the Editor > Datasets tab (when in table view).

View dataset import candidates

To see datasets (or updates to datasets) that have been brought into Pure by the *DataSearch: Import Datasets* cron job:

1. Navigate to Editor > Datasets.

If you are using the content matrix view (the default view), all datasets are grouped in a table by organizational unit. You can see the number of import candidates in the **Import candidates** column.

Last DataSearch import occurred: 23 Oct 2019 08:07 - the next import will happen at: N/A

	Import candidates	Entry in Progress
[Redacted]	3	2
[Redacted]		1
[Redacted]		1
Department of Architecture, Design and Media Technology		9
Department of Business and Management		2
Department of Chemistry and Bioscience		2
Department of Communication and Psychology		2
Department of Culture and Global Studies		6
Department of Electronic Systems		5
Department of Law		11
Department of Materials and Production		2
Department of Political Science		3
Department of Sociology and Social Work		1
Global & Transnational Law Research Group		1
Signal and Information Processing		1
TBRP – Theory Building Research Programme		2
The Faculty of Engineering and Science		1
The Faculty of Social Sciences		2
The Technical Faculty of IT and Design		1
Total	3	55

Tip

On this screen you can also see when the last DataSearch import occurred and when it is next scheduled.

- Click on one of the numbers in the **Import candidates** column. The **Import candidates from DataSearch** window is opened.

Import candidates from DataSearch for [Redacted] University

List of new Datasets imported from DataSearch for [Redacted] University

4 results Limit result: All Sort by: As returned by source

Geometric measurement of the surface of a v-bend during multi-scan laser forming
 [Redacted] 2019 DOI.
 Updated Found: 23 Oct 2019 08:07
 Update Source data | DOI

Images and videos_ Experiment_% plastics waste in municipal mixed solid waste in the Municipality of Aalborg
 [Redacted] 2019 DOI.
 Updated Found: 23 Oct 2019 08:07
 Update Source data | DOI

Single grain kernel model results
 [Redacted] 2018 DOI.
 Found: 23 Oct 2019 08:07
 Import Source data | DOI Reject

Test problems for metamodeling comparison: 5 building performance metrics and 8 theoretical problems
 [Redacted] 2017 DOI.
 Found: 23 Oct 2019 08:07
 Import Source data | DOI Reject

Figure 2. The newly detected or changed datasets for that particular organizational unit are shown.

Tip

You can limit to show only new import candidates or updated import candidates. You can sort by the source, the time they were found, or by title.

- If you want to investigate more details about the dataset candidates, you can click **Source data** to expand a JSON snippet with metadata about the dataset, or click **DOI** to navigate to the data publisher's website.

Import a candidate dataset

1. Open the list of candidate datasets for the organizational unit you want to import to. See [Section View dataset import candidates](#) above.
2. Click **Import** under the name of the dataset candidate that you want to import.

The **Import from DataSearch** window opens.

People match

People in DataSearch	People in Pure
1 [Redacted]	<p>2 matches</p> <p>[Redacted] Person: VIP (Staff)</p> <ul style="list-style-type: none"> Department of Materials and Production The Faculty of Engineering and Science Organisational unit: Institute The Faculty of Engineering and Science Aalborg University Organisational unit: Faculty Physics and Mechanics Department of Materials and Production Organisational unit: Research Department/section Materials Science and Engineering Physics and Mechanics Organisational unit: Research group <p>+/-</p> <p>↕</p>

3. Follow the fields in the **Import from DataSearch** window to enter all required information.

In the **People match** section, for each author of the dataset, you need to (where possible) identify an internal person who corresponds to this author. Pure tries to match people in DataSearch with people in Pure, and indicates with a label whether there are multiple matching names, a single matching name, or no matching names.

If there are multiple people with the same name at your institution, you need to select the relevant person from the drop-down list. If there is no person with this name, importing the dataset will create an External Person.

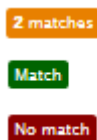
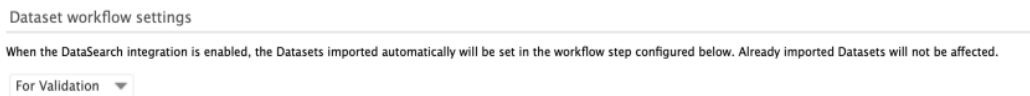


Figure 3. Yellow, green and red labels appear indicating whether you need to review the matched persons, can be confident in the match, or whether there is no match.

You can add to the organizational unit affiliations of this person to this dataset using the **+/-** button or switch person using the **↕** button.

4. Click **Save**.

The dataset is entered into Pure in the workflow step configured on the Administrator > Datasets > DataSearch tab.



If the dataset has a specific ID in its metadata, any subsequent updates made to that Dataset in DataSearch (e.g. metadata edits, new data files) will be automatically synchronized in Pure when the *DataSearch: Import Datasets* cron job next runs.

Import a dataset update

If a Dataset that has been previously imported into Pure with DataSearch has been edited in an external repository with updated Persons and/or Organizations, then this Dataset will re-appear in the import candidate list, with the label "Updated". To import this update:

1. Open the list of candidate datasets for the organizational unit you want to import to. See [Section View dataset import candidates](#) above.
2. Locate the dataset candidate that has been updated.

Tip

You can use the **Limit result** dropdown to show only updated datasets.

3. Click **Update** under the name of the dataset that you want to update.

A window is opened, allowing you to update the person and organizational unit affiliation details.

4. Click **Save**.

The changes are added to the Dataset record in Pure. If there have been any changes to the Dataset that were entered directly in Pure prior to this update step where the field was not included in the source data, these are retained.

Reject a dataset candidate

When processing a Dataset candidate, you can also *reject* a given candidate and choose to not import it into Pure. If a candidate is rejected it will not appear in the list of import candidates again.

Note

If you delete an imported dataset, it also appears in this list of rejected candidates. This is to prevent it from being re-imported into Pure.

To reject a dataset candidate:

1. Open the list of candidate datasets for the organizational unit you want to import to. See [Section View dataset import candidates](#) above.
2. Click **Reject** in the row of the dataset candidate that you want to import.



- To re-introduce rejected dataset candidates, navigate to Administrator > Datasets > DataSearch and select the Clear rejected candidates button.

Rejected candidates

<input type="checkbox"/>	1 result		Sort by: ↑ Title
<input type="checkbox"/>	Single grain kernel model results		Created: 23/10/2019
	[REDACTED] DOI.		Rejected: 23/10/2019

[Clear rejected candidates](#)

Note

Clearing rejected candidates will re-introduce all rejected candidates to the list of import candidates.