ELSEVIER

Research Intelligence

Populating Pure in Bulk

Version 5.16

Empowering Knowledge™

Table of Contents

1. Introduction 1
1.1. In this manual... 2
1.2. Before you begin 2
1.2.1. Populating content in the correct order 2
1.2.2. Additive import strategy in brief 4
1.2.3. Alternative import strategies 4
2. Preparing data sources 6
2.1. Data formats 6
2.2. Data mapping 6
2.3. Recommended workflow for mapping and transforming data 6
2.4. Pure's XML import format and the XSD 6
2.4.1. Converting your legacy data and external data to XML 7
2.4.2. Handling multiple XML files 7
2.4.3. Finding example XML and the XSD 7
2.5. The Pure data model: content types, templates and types 8
2.5.1. Content types 8
2.5.2. Templates 8
2.5.3. Types and sub-types 8
2.5.4. Content types, templates and types in action 8
2.6. Mapping types and fields to the Pure data model 9
2.6.1. Type mapping 9
2.6.2. Field mapping 9
2.7. Mapping external values to classifications 10
3. Migrating legacy data into Pure 12
3.1. Importing or migrating data with the Bulk Import Wizard 12
3.2. Viewing the status of an import from the Bulk Import Wizard 12
3.2.1. Using the log entry details area 14
3.3. Bulk populating content in Pure - detailed steps 14
4. Integrating external data into Pure 17
4.1. Creating an integration job 17
4.2. Viewing and editing an integration job 19
4.3. Configuring an integration job 19
4.3.1. Opening the Configuration area 19
4.3.2. Configuration area fields and buttons 20
4.4. Scheduling an integration job 24
4.4.1. Opening the Schedule area 24
4.4.2. Schedule area fields and buttons 25
4.5. Running an integration job on an ad hoc basis 25
4.6. Viewing the status of an integration job 26
4.6.1. Opening the log entry details area 26
4.6.2. Using the log entry details area 28
4.6.3. Integration job error messages 28

4.7. Using the Bulk Import Wizard to import external data 29
4.8. Integrating data sources with Pure - detailed steps 29
5. Evaluating the quality of imported data 31
5.1. Using the job logs to evaluate import quality 31
5.2. Fixing incorrectly integrated records 31
5.3. Deleting individual records 31
5.4. Bulk deleting records 32
6. Importing data into Pure 34
6.1. Populating Organization data 34
6.2. Populating User data 37
6.3. Populating Person data 40
6.4. Populating Supervisor data 43
6.5. Populating Publisher data 45
6.6. Populating Journal data 47
6.7. Populating Event data 49
6.8. Populating External organization data 52
6.9. Populating External person data 54
6.10. Populating Research output 57
6.11. Populating Activity data 61
6.12. Populating Funding Opportunity data 63
6.13. Populating Application data 65
6.14. Populating Ethical review data 67
6.15. Populating Award data 70
6.16. Populating Project data 73
6.17. Populating Dataset data 75
6.18. Populating Press/Media data 77
6.19. Populating Prize data 79
6.20. Populating Impact data 82
6.21. Populating Course data 84
6.22. Populating Student Thesis data 86
6.23. Populating Facilities/Equipment data 89

This document is Version 1.5 of the manual *Populating Pure in Bulk*, produced on Monday, July 8, 2019 at tag 898ed51 .

1. Introduction

The goal of a Pure population is to populate Pure with data before going into production. Your data may come from:

1. **Legacy data sources**
   A legacy data migration moves data into Pure from legacy systems that will soon be decommissioned. Once you go into production with Pure this data is maintained *exclusively* in Pure.
2. **External data sources**
   An external data integration regularly copies data into Pure from systems where the data is actively maintained. In Pure, the data is linked to other records, and you can configure whether aspects of this external data are maintained in Pure or are read-only.

Populating data always consists of one or more of the following tasks within Pure:

1. **Populating data with the Bulk Import Wizard**

   You will use the Bulk Import Wizard to migrate legacy data into Pure.

   You can also use the Bulk Import Wizard to initially populate data that you want to later integrate into Pure, such as data from an external HR system.

   Tip

   To open the Bulk Import Wizard, click the Administrator tab on the top navigation, then the Bulk import tab on the navigation. Select the relevant wizard name listed under the Bulk import tab.

2. **Populating or synchronizing data by running a job**

   If you are integrating external data, you will set up and run a job to populate or synchronize the data in Pure from an external system.

   Tip

   To view or create a scheduled job, click the Administrator tab on the top navigation, then the Jobs tab on the navigation. You can add a new job by selecting its name from the Create new job drop-down menu.

   Integrations from an external data source are normally run periodically, for example, every night. However, you can also run integrations on an ad hoc basis, for example, when you are initially populating Pure with data.

   Jobs can also used to periodically import data from online data sources, though these processes are not described in this manual.

3. **Manual entry**

   You can also enter data manually into Pure by creating and editing records using the Pure interface, though we recommend that you attempt to automate the entry of records while populating Pure before going into production.

   Tip

   A quick way to manually enter most types of records is to use the Add new shortcut button available on most screens.

   Note

   After you go into production, manual entry is the normal way of entering records that are not synchronized from an external data source.

The input for the Bulk Import Wizard and integration jobs is an XML file containing the records to be populated or synchronized. See Section 2. Preparing data sources.

Warning

You can only use the Bulk Import Wizard and integration jobs if your institution has purchased the *Import* module.

1.1. In this manual...

This manual contains both general information about how to populate Pure, as well as specific tutorials for populating each specific content type.

It is intended that you begin by reading the introductory sections (see Section 1. Introduction), followed by the instructions specific to the content type you are populating (see Section 6. Importing data into Pure). Refer to the other sections as required for more information about how to prepare data and how to access and use specific parts of the Pure interface.

For information about:

- choosing the order in which to populate content, see Section 1.2.1. Populating content in the correct order.
- an overview of the population process, see Section 1.2.2. Additive import strategy in brief.
- preparing the data before importing it into Pure, see Section 2. Preparing data sources.
- how to populate data using the Bulk Import Wizard, see Section 3. Migrating legacy data into Pure.
- how to integrate data using an integration job, see Section 4. Integrating external data into Pure.
- how to evaluate the success of the import process, see Section 5. Evaluating the quality of imported data.
- how to populate (and optionally integrate) a specific content type, see Section 6. Importing data into Pure.

1.2. Before you begin

1.2.1. Populating content in the correct order

If you are populating Pure with multiple content types, there is a specific order in which different data sources must be imported into (and for some content types, synchronized with) Pure. This is due to the highly-linked nature of data in Pure.

The dependency of some data on other data means that the process of populating Pure can be described as three stages: Master data integration, Research output migration, and migrating or integrating the remaining content types.

Master data integration

In Pure, Master data refers to a small set of records that are frequently linked to by other items. One strength of Pure is that it maintains a high data quality by maintaining this small set of records in one place. Only users with the appropriate permissions can edit these important records.

For example, Organizational units (a specific university, faculty etc.) and Persons as well as Publishers, Events and Journals are Master data. These records are linked when a new publication is added which references a Person as the author, the Organization (such as the specific faculty) where the research took place, and a Publisher and Journal (or, if the publication is a conference submission, an Event). If the publication has a connection to other researchers or institutions that are not part of your institution (such as co-authors from a different university), these are linked as External persons or External organizations respectively.

The user entering such a publication does not have to create these associated Master data records, as they already exist or are created automatically. Nonetheless, new records, such as a previously unseen journal, can still be added.

It is important that this Master data is of particularly high quality, because problems with these records affect many other records.

The first step in a Pure implementation is therefore to work on an integration of HR data from your institution's external systems. HR data includes:

- **Organization data** such as records of university faculties or research groups.
- **User data** such as a list of email addresses of users who require login credentials.
- **Person data** such as records of researchers and staff (including Supervisor information).

For more information, see Section 4. Integrating external data into Pure, or access the relevant content type in Section 6. Importing data into Pure.

Recommendation
Your institution may choose to not explicitly import some content types that are created on the basis of other imports, such as *Event*, *Journal*, *Publisher*, *External person*, or *External organization*. For example, Journal records are created as a by-product when Research output records are imported. Nonetheless, if your institution maintains good and valid data on these subjects you should import them at this stage.

Research data migration

Following the Master data integration, research data can be migrated into Pure.

Research data refers to the content types *Research output*, *Activities*, *Applications*, *Awards*, and *Projects*.

Of these, Research output content is generally migrated first as these are typically the most numerous and complex records.

For more information see Section 3. Migrating legacy data into Pure or Section 6.10. Populating Research output.

Migrating and integrating remaining content types

After Master data have been integrated, depending on the needs of your institution you can begin to import other content types, such as:

- Dataset
- Press/Media item
- Prize etc.

Recommendation
We recommend that you begin importing research data before populating the remaining content types, though it is also possible to populate these at the same time if they are not closely linked to research output records.

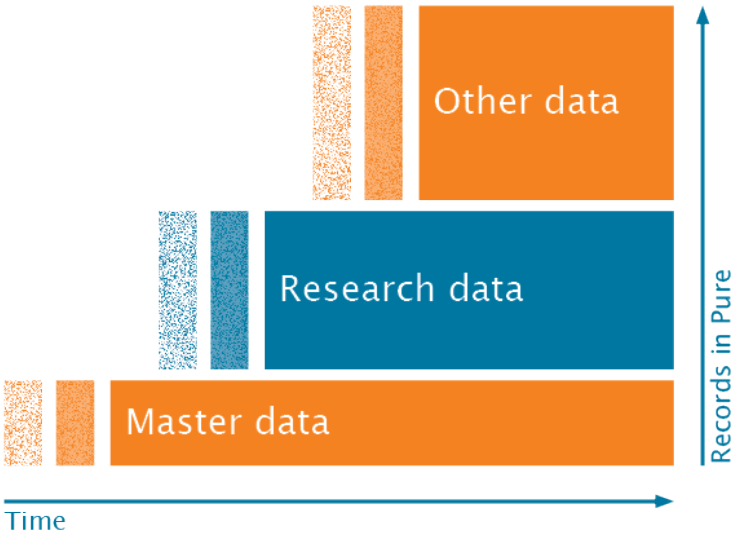


Figure 1. During each of the data import steps you experiment with importing and deleting content until the data is of a sufficient quality. This data then forms the basis for importing further data.

1.2.2. Additive import strategy in brief

While importing each content type, we recommend an *additive* import strategy in which you import, inspect, and (optionally) delete certain records, then improve the mapping and transformation based on the results, beginning with a few simple records and repeating this process until all records have been imported and are correct.

> **Tip**
> Keep in mind that the quality of your data in Pure cannot exceed its quality in the original format, and that the purpose of the improvement step is to reduce noise introduced by your transformation processes and mapping.

Importing data

To import data into Pure:

1. Extract the appropriate data from the legacy or external system.
2. Transform the data into an XML format compliant with the Pure XSD using the mapping between the types and fields in your legacy or external system and the Pure data model. See Section 2.4. Pure's XML import format and the XSD and Section 2. Preparing data sources.
3. Migrate or integrate your data. Consider that you can also use the Bulk Import Wizard to initially populate data when you are performing an integration. See Section 3. Migrating legacy data into Pure and Section 4.1. Creating an integration job.

Inspecting data quality

To inspect the quality of your imported data:

- Check the log of the import. See Section 5.1. Using the job logs to evaluate import quality.
- Search for the type of the items that were imported and manually view items.
- Follow the advice listed for each specific content type in its tutorial. See Section 6. Importing data into Pure.

See Section 5. Evaluating the quality of imported data.

Deleting noisy items

If you have found that the import introduced items with errors, either:

- Search for the data type you imported and bulk delete all of the items of this type.
- Search for the data type you imported, filter to the items that have errors and bulk delete all of the items.
- If the import is the result of an integration job, change the configuration settings so noisy items are updated the next time the job is run.

See Section 5.4. Bulk deleting records and Section 4.3. Configuring an integration job.

> **Tip**
> If your data is correct, proceed by keeping the items that were imported, and importing a different set of records (with different IDs) during the next iteration.

Improving the mapping and transformation

To improve the quality of imported data:

- Review the example XML and XSD for the data you are importing and revise the processes that bulk transform your data into XML. See Section 2.4. Pure's XML import format and the XSD.
- Revise the mapping between the types in your external data and the types in the Pure data model. See Section 2.6. Mapping types and fields to the Pure data model.

1.2.3. Alternative import strategies

If you have access to a higher support tier for Pure Implementation Services, you may consider alternatively

following an *iterative* import strategy, whereby you repeatedly:

- develop your transformation and mapping processes.
- import all of the records at once.
- identify and resolve problems with the transformation and mapping.
- request that your Pure environment be reset to its initial state.

The advantage to this strategy is that you do not bulk delete content, which can be time-consuming and creates an audit trail. However, resetting the environment depends on your hosting options and requires a service request, and also requires manual re-entry of settings configured within Pure, such as user roles or configuration settings.

2. Preparing data sources

2.1. Data formats

You can import records from any data source into Pure, as long as you can extract the data from this source and transform it into XML that conforms to the Pure XML schema definitions (XSD files).

For example, a data source may be a spreadsheet or CSV file, a database, XML, or a different format. A transformation may take place using a transformational language like Perl or XSLT, or any other bulk transform tool of your choice.

XML examples and XSD files are found within Pure on the relevant import screen. For more information, see Section 2.4. Pure's XML import format and the XSD.

2.2. Data mapping

Data mappings describe how data elements in your data source correspond to the Pure data model.

For each record type in the data you want to import you need to choose which type it will receive in Pure, and which fields in Pure correspond to each field in the original system. For more information, see Section 2.6. Mapping types and fields to the Pure data model.

2.3. Recommended workflow for mapping and transforming data

A common workflow for performing this mapping is as follows:

1. Familiarize yourself with the terminology of types, templates and content types in Pure. See Section 2.5. The Pure data model: content types, templates and types.
2. Familiarize yourself with the Pure XML format. See Section 2.4. Pure's XML import format and the XSD.
3. Define mappings and transformations for fields and types. See Section 2.6. Mapping types and fields to the Pure data model.

Recommendation

We recommend you use an iterative process to define the mappings and transformations:

1. Define simple and obvious mappings.
2. Transform several test records to XML using these mappings.
3. Import several records with these mappings.
4. Check the quality of the resulting import.
5. Bulk delete the imported records.
6. Address any problem with the current mappings and transformation, adding more fields for the next import.

2.4. Pure's XML import format and the XSD

XML (eXtensible Markup Language) is a common data format for data exchange between IT systems which supports Unicode characters. XML does not inherently specify the exact elements and attributes a document can have, rather these are defined in a separate XML schema definition (XSD), which is usually referenced in the XML document.

A particular XML document is described as *well-formed* if its syntax is correct and *valid* if it follows all the rules defined in the XSD it references. Pure can import data in XML format if the XML is well-formed and valid against one of the Pure XSDs.

XML files are automatically validated when they are imported into Pure, and invalid XML will raise errors. You can temporarily disable the validity requirements for XML files when configuring an integration job if needed. See Section 4.3. Configuring an integration job. Use a third-party XML validation tool to validate your XML before importing it into Pure to save time.

2.4.1. Converting your legacy data and external data to XML

To populate Pure, you must bulk transform the output of your legacy or external data source into this valid XML format. You are free to transform the data from your legacy and external systems to this XML format using any tools or scripts that are most convenient, such as a Groovy or Perl script, an XML transformation language, etc.

Keep in mind that you need to maintain the semantic correctness of your data by defining the correct mapping between the types and fields in the original data and the types and fields used in Pure.

For more information about the Pure data model and defining this mapping, see Section 2.5. The Pure data model: content types, templates and types and Section 2.6. Mapping types and fields to the Pure data model.

2.4.2. Handling multiple XML files

Pure allows you to import multiple XML files at once, for example when using the Bulk Import Wizard.

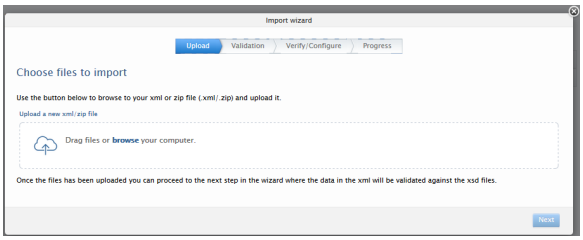


Figure 2. Upload screen of the Bulk Import Wizard.

To import data from several XML files at the same time:

1. Select the XML files you want to import.
2. Convert these files to a *zip* file, using your preferred tool.
3. When you are prompted to select or drag and drop the file you want to upload, select the zip file.

2.4.3. Finding example XML and the XSD

When you are transforming data from an external source or a legacy source into XML, there are several resources that can help you:

- Simple and advanced example XML files for each content type.
- XSD files defining the structure of the XML that can be imported for each content type.

To find these example XML and XSD files:

1. If you are integrating an external data source navigate to the Job Name page. See Section 4.1. Creating an integration job or Section 4.2. Viewing and editing an integration job.

   If you are migrating a legacy data source, navigate to the relevant Bulk Import Wizard. See Section 3. Migrating legacy data into Pure.
2. Click the example XML and XSD files at the top of the page to download them.

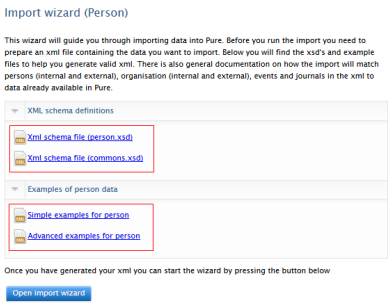


Figure 3. Example files for the Bulk Import Wizard.

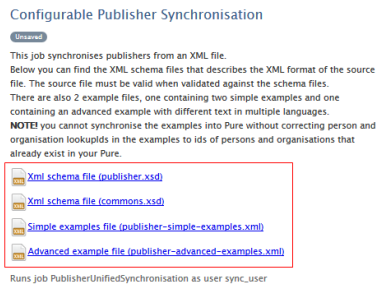


Figure 4. Example files for an integration job.

The example files can be viewed with your preferred tool.

2.5. The Pure data model: content types, templates and types

The Pure data model refers to the structure of the records and their relations as maintained in Pure. This data model is designed in line with international standards to ensure the interoperability of Pure with important external databases.

In Pure, records have different *content types* based on the kind of entity they represent. Within each content type, templates, types and sub-types (where available) further specify the kind of real-world entity that the record represents.

2.5.1. Content types

Content types are the highest-level of distinction between records within the Pure data model. Records with different content types generally refer to entities which (in the real world) are two very different kinds of objects or concepts.

The content types in the Pure data model include *Organization*, *Person*, *Research Output*, *Activity*, *Ethical review*, *Project* etc.

Different content types are generally populated from separate data sources and at separate times, with the exception of External organization and External person which are created when other records link to organizations or people that are not affiliated with your institution and thus do not exist in Pure.

2.5.2. Templates

Templates define the set of fields available for a record. They are specific to a content type.

Many content types, such as *Organizations*, have only one template. Therefore every Organization record has the same fields. Other content types, such as *Research output*, have multiple templates, such as *Contribution to journal* and *Patent*.

Though some fields are present across many templates, other fields are distinct. For example, the *Patent* template has the field *Patent number*, which is not present in other templates.

2.5.3. Types and sub-types

Types and sub-types indicate most specifically what kind of entity a record in Pure refers to. The types and sub-types available for a record depend upon its content type and template.

Records with different types or sub-types still share all the same fields if they have the same template. The only difference is the type value. Not all templates have different types or sub-types available, this is just a level of distinction that may be available to label records.

Nonetheless, it is important to correctly enter a type as this may be relevant for reporting or processing purposes.

2.5.4. Content types, templates and types in action

You can explore the different content types, templates and types available when adding new content using the *Add new content* wizard.

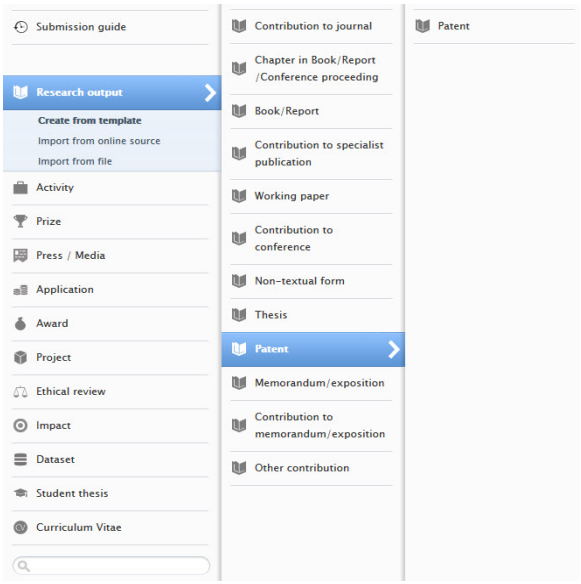


Figure 5. The Add new content wizard prompts you to select a content type in the left column, a template available for this content type in the middle column, and a type available for this template in the right column. To open the Add new content wizard select Add new *in the Pure interface.*

2.6. Mapping types and fields to the Pure data model

2.6.1. Type mapping

For each type of record in your legacy or external data source you need to decide which record type it should map to in Pure. For example, if you use the type *EditorialLetter* in your data you can map it to the type *Editorial* in the Pure data model.

Determining the type mapping is generally straightforward, except in some special cases:

Mapping one type to two different Pure data model types

You may need to split one type from your legacy or external system into two types in Pure if your previous record-keeping conflated two types that are separate in the Pure data model. In this case, you will need to decide which Pure data model type each imported record should receive on a record-by-record basis.

We recommend you devise rules which can perform this classification automatically using the data in the record itself. When you convert your legacy or external data into a form compliant with the Pure XML schema definition, label these records with the relevant new type. See Section 2.4. Pure's XML import format and the XSD.

Determining the type mapping in tricky cases

There may be several cases where the type mapping from your legacy or external system is not clear from the names of the types, and you find several candidate types which are tricky to choose between. In this case, you should choose the type that is most obvious on the basis of:

- the **field mapping** - consider how the fields will be mapped to the available fields for each of the candidate types in the Pure data model. Will a candidate type's fields capture all the relevant data about the records?
- the **Pure interface**, particularly the **Editor window** - consider the context in which each type is available in the user interface. Do you think users will assign the correct type when they enter a record?

2.6.2. Field mapping

For each template in the Pure data model you need to determine a mapping between the fields in your legacy or external system and the template's fields.

You do not have to define a mapping for every field for every type, because in the Pure data model many records have the same field structure. Despite this, it is possible to define a different field mapping for two types if you find it necessary.

Field mapping can be a complex task nonetheless. We recommend you follow these workflows for a smooth process:

Mapping common and mandatory fields first

Begin by mapping common fields like *title*, *subtitle*, *author*, *year* etc. which are straightforward and affect a large number of types. A good rule of thumb is to target fields which are mandatory.

Performing a simpler import then adding more field mappings

Only map more complex fields once you have verified that everything is working as intended with some simple fields. Periodically re-import as you are defining field mappings to check the results of your additions.

2.7. Mapping external values to classifications

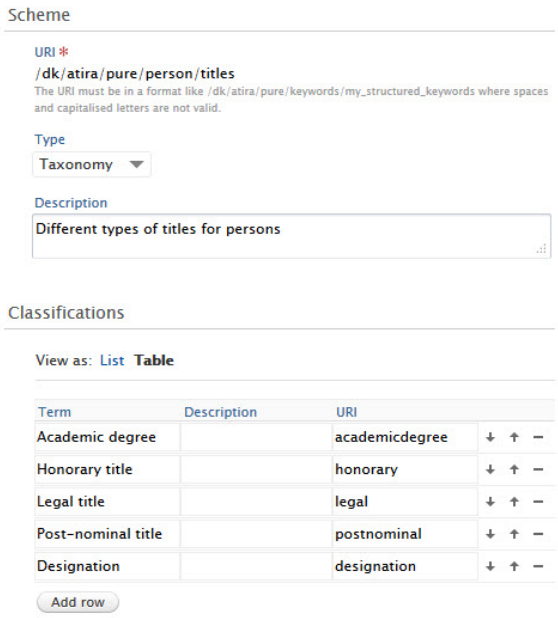
In Pure, certain fields can only hold a restricted number of values. Each of these values is called a *classification*. The hierarchy of classifications that determines which values and sub-values are available for a field is called the *classification scheme*.

When preparing XML for external data where the values correspond to classifications, do not use the interface text of a classification (which can be freely changed), rather use part of the URI of a classification.

To find the value that should be used in XML for a classification:

1. Open the classification scheme that your values correspond to:
    1. Click Classification schemes in the navigation of the Master data tab.
    2. Search for the classification scheme by typing keywords into the search bar.

       The Classification scheme editor window is displayed for the selected classification scheme.



2. View the classifications under this classification scheme in the *Classifications* area, either as a list or table.

3. If you are using the list view, extract the final part of the URI between the final slash and the end of the string.

   For example, from the URI **/dk/atira/pure/person/titles/academicdegree** extract **academicdegree**.

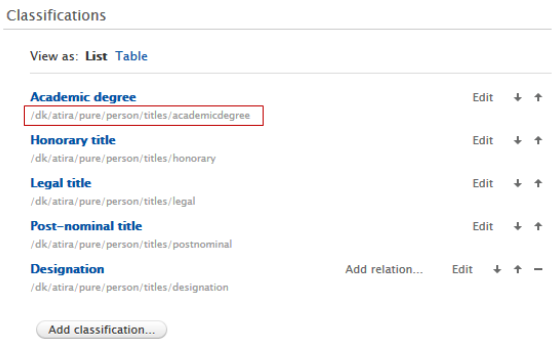


Figure 6. Classifications area of the Classification scheme editor window with a URI highlighted.

4. If you are using the table view, extract the value from the URI field for the relevant classification.
5. Use this value in the XML that you construct with the data from your external system.

3. Migrating legacy data into Pure

You can use the Bulk Import Wizard to import content into Pure.

When the content you are importing will be fully-maintained in Pure in the future, this is called a *legacy data migration*, however you can also use the Bulk Import Wizard to initially populate Pure with data that will eventually be maintained through an integration with an external data source. See Section 4. Integrating external data into Pure.

You can import any content that is in an XML format valid against the Pure XSD.

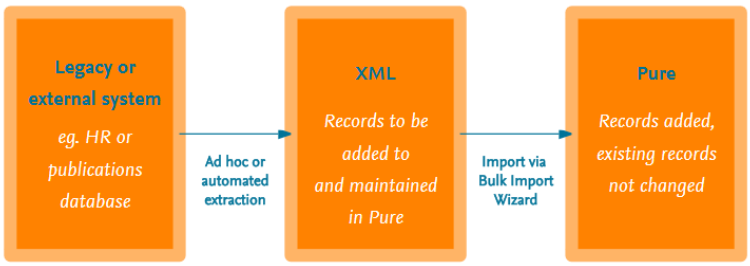


Figure 7. Data formats during a migration.

Warning

Pay attention to the order in which you import content. Importing content in the wrong order will cause errors, particularly when links in the XML to organizations or persons cannot be matched to existing entities. See Section 1.2.1. Populating content in the correct order.

3.1. Importing or migrating data with the Bulk Import Wizard

You can use the Bulk Import Wizard to import data into Pure in bulk, or to migrate legacy data to Pure.

To import or migrate data with the Bulk Import Wizard:

1. Click the Administrator tab in the top navigation.
2. Click the Bulk import tab in the navigation.
3. Click the name of the content type you want to import or migrate in the navigation below the Bulk import tab.

   The Import wizard page is opened for the wizard you selected.
4. Review the following documents available on the Import wizard page:
   - the example data XML files for the content type.
   - the XSD files for the content type.
   - the User Guide PDF for the wizard for information about importing or migrating data.
5. Generate a well-formed and valid XML file from your data source with the records you want to import or migrate.
6. Click Open import wizard.

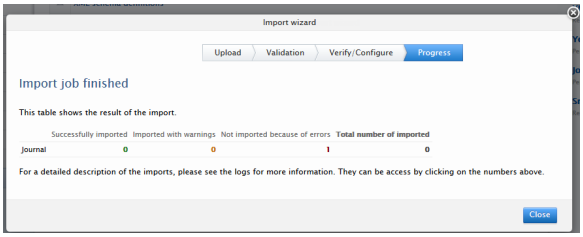
   The Bulk Import Wizard wizard for your chosen content type is opened and guides you through the legacy data migration process.

3.2. Viewing the status of an import from the Bulk Import Wizard

After you have used the Bulk Import Wizard to import data into Pure, you can view a log with details about the success or failure of the import.

To view the log for an import, either:

- Click on the numbers in the table on the *Progress* screen of the Bulk Import Wizard.

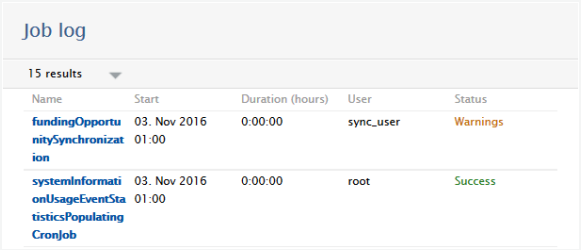


- **View all jobs' logs from the Job log page.**

1. Click the Administrator tab in the top navigation.
2. Click the Jobs tab in the navigation.

   The Scheduled Cron Jobs page is displayed with a list of scheduled jobs.
3. Click the Job log tab in the navigation.

   The Job log page is displayed.



4. Click the name of the job you want to view log entries for.

   A detailed list is displayed with the results of each run.

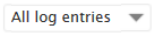
5. Click the *Start* column for the run you want to view details for.

   The log entry detail area is expanded.

> **Tip**
>
> You can also click in the *Success*, *Warnings*, *Errors* or *Fatal errors* columns for the run to show only the log details with that status.

> **Tip**
>
> The job name for an import with the Bulk Import Wizard is **Child job of Import wizard execution - [Content Type]** where [Content Type] is the name of the import wizard selected.

3.2.1. Using the log entry details area

Field/Button	Description
	Filter the log entries shown to: • All log entries. • Entries with a status of success. • Entries with a status of warning. • Entries with a status of error. • Entries with a status of fatal error.
	Enter a keyword and press **Enter** to show only log entries with one or more lines that exactly contain the search query. **Note** You cannot search for multiple keywords unless they appear exactly in the order given in one or more log entry lines. Search for an empty string to show all entries. **Tip** Search for part of the error message that does not relate to a specific record to see how commonly this error occurred during the job's execution.
	Generate and download an *XLS* file with the content of the *Title and description* field for the current filter and search selections.
Status	The status of the log entry.
Title and description	The title of the log entry, followed by warnings and information about the log entry.

3.3. Bulk populating content in Pure - detailed steps

These steps describe in detail how to bulk populate a content type in Pure using the additive strategy described in Additive import strategy in brief (sec. 1.2.2).

In contrast to other tutorials, these detailed guides:

> describe strategies for approaching the amount and type of data to import when testing the import procedure.
> guide you through the appropriate steps to correct errors based on what kinds of errors are visible in the Pure interface.

You can see less-detailed instructions specific to each content type in Importing data into Pure (sec. 6).

1. Begin the population process for this content type by planning which data should be imported into Pure. Consider:
    - what the data source is.
    - how the records can be extracted.
    - where the records will be maintained in the future.
    - which types of content these data link to.
    - how types and fields should be mapped to the Pure data model.
    - which field, or which composition of fields, can be used as the permanent unique ID for each record.

If your data source will continue to be maintained externally to Pure, after you have imported the records into Pure you will need to configure an integration job so that changes made in the external data source are reflected within Pure. See Section 4. Integrating external data into Pure.

Tip

If you are populating multiple content types, consider the order they are populated as some data may need to link to other existing records. See Section 1.2.1. Populating content in the correct order.

2. Extract a subset of the records from your data source.

Tip

Begin initially with just a handful of records that have the most common types in your data source. Each time you re-approach this step include more and different (or more complex) types of records.

3. Define a mapping between the types and fields in the extracted data and the Pure data model for any types and fields not yet mapped. See Section 2. Preparing data sources.

Note

It is possible to start with a minimum field set of only the mandatory fields until you are sure that the import process is working correctly then add more (optional) fields later. If you follow this approach, keep in mind that existing records are not updated, so delete the minimal record first before importing it again with more information.

4. Transform the external data into XML using your preferred process. See Section 2.4. Pure's XML import format and the XSD.
5. Use the Bulk Import Wizard to import this XML file into Pure. If you encounter errors during the import, use the error message to resolve these errors. Proceed to step 2, 3 or 4 depending on the errors.

Note

Error messages appear on the Progress screen of the Bulk Import Wizard, or you can access them from the Job Log Page. See Section 3.2. Viewing the status of an import from the Bulk Import Wizard.

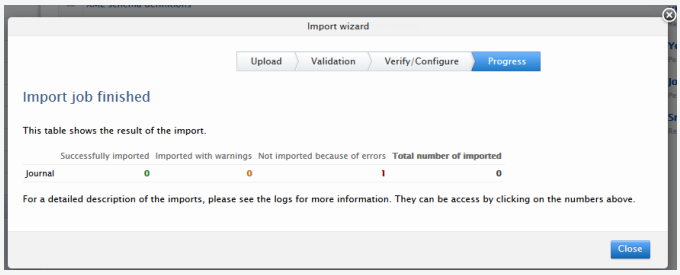


Figure 8. Errors displayed in the Bulk Import Wizard.

6. Inspect the data in the GUI for correctness by opening individual records from the Editor or Master data tabs.
    - If the imported data is accurate, proceed to the following steps.
    - If the imported data is not accurate:
        - If there is a problem with the type and field mapping, you need to revise the mapping. Delete all records with this content type and proceed to step 4.
        - If the problem is not related to the field mappings, it may lie with your source data or the transformation processes. Delete all the inaccurate records, clean your source data if necessary, and revise your transformation processes. Proceed to step 2.

7. Inspect some records in the GUI for completeness.
    - If all the fields from your data source are present and correct, proceed to the following steps.
    - If some fields from your data source are missing, repeat the import process by including a few more fields and records from your data source and defining transformations for these new fields. Proceed to step 2.
8. Consider how you will add new records in the future.
    - If you will continue to maintain this data externally to Pure, you can either import the remaining records now using the Bulk Import Wizard, or allow them to be imported the first time you configure the integration between Pure and the external data source.

      The migration process is complete, proceed to the instructions for setting up an integration. See Section 4. Integrating external data into Pure.
    - If you will add records of this type exclusively in Pure in the future, ensure that you extract the complete set of (remaining) records of this type, and proceed to step 4 to transform and import them.

      Tip

      Pure does not import a record twice, so you can freely include records that have already been imported.

All your data of this content type has been populated in Pure, or will be populated once you have configured an integration job.

4. Integrating external data into Pure

You can integrate external data into Pure by setting up an integration job that updates Pure to reflect changes in some external data source.

An integration job can be used to populate Pure with its initial data, where the data comes from sources maintained outside Pure. If you are populating Pure initially and want to test the format of the data before configuring how existing content should be synchronized, you can use the Bulk Import Wizard. See Section 4.7. Using the Bulk Import Wizard to import external data.

You also can only integrate content that is in an XML format valid against the Pure XSD with a unique and stable ID for each entity.

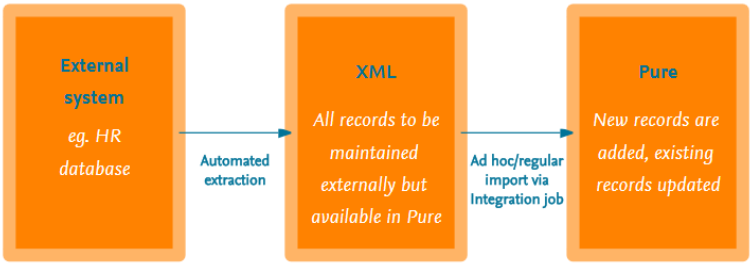


Figure 9. Data formats during an integration.

Note

Integration jobs match entities from your external data source with entities in Pure based on their ID. If:

- an unseen ID is found, a new entity in Pure is created from the external data source.
- an ID corresponding to an existing entity in Pure is found, the entity is updated from the external data source in accordance with the job's configuration settings. See Section 4.3. Configuring an integration job.
- an ID is missing for an entity in Pure which was originally created from the external data source, the end date for the entity is set to the current date. It is marked as historical and not synchronized, and can now be edited in the Pure interface.

The relevant ID field is defined in the XSD file for the content type. For example, the ID field for an organization is **organisationID**. See Section 2.4. Pure's XML import format and the XSD.

Before you set up an integration job, consider:

- Before you perform an integration job for the first time, ensure that you have defined all the necessary classifications used in your data.
- Integrated content will only be meaningful if you have accurately mapped the types in your external data source to the types in the Pure data model. See Section 2.6. Mapping types and fields to the Pure data model.

4.1. Creating an integration job

To create an integration job:

1. Click the Administrator tab in the top navigation.
2. Click the Jobs tab in the navigation.
3. Click the Create new Job dropdown menu.

4. Depending on the content type you want to integrate, choose the relevant job:

Content type	Job name
Organization	Configurable organisation synchronisation
User	Configurable user synchronisation
Person	Person synchronisation
Funding opportunity	Funding synchronisation
Project	Project synchronisation
Supervisor	Configurable supervisor synchronisation
Activity	Configurable activity synchronisation
External organization	Configurable external organisation synchronisation
External person	Configurable external person synchronisation
Journal	Configurable journal synchronisation
Publisher	Configurable publisher synchronisation
Event	Configurable event synchronisation
Press/Media item	Unified press/media synchronisation
Dataset	Configurable dataset synchronisation
Application	Application synchronisation
Award	Award synchronisation
Prize	Configurable prize synchronisation
Student thesis	Configurable student thesis synchronisation
Facilities/Equipment	Configurable equipment synchronisation

The Job Name page is opened for the job you selected.

Note

The Unsaved badge is shown under the job name when this page displays a new job.

5. Consult the example XSD and XML files for instructions about how to structure the XML file with data from your external system for this content type. See Section 2. Preparing data sources.
6. In the Schedule area, click Change schedule to edit when the job is performed. See Section 4.4. Scheduling an integration job.
7. In the Configuration area, click Edit configuration to edit the default configuration values for the job. See Section 4.3. Configuring an integration job.

Note

Take time to revise the specific instructions for the content type you are integrating. See Section 6. Importing data into Pure.

8. Click Save.

   The integration job is created and log information is displayed for the job. See Section 4.6. Viewing the status of an integration job.

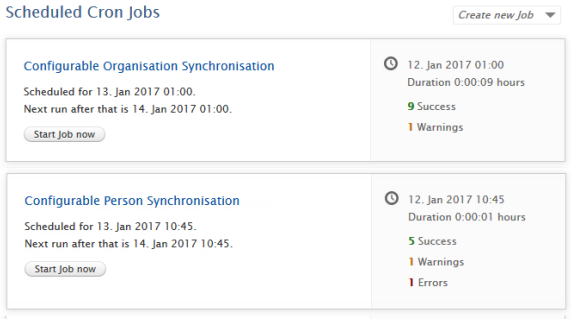
   You can either wait for the job to run at its scheduled time or you can run the job immediately. See Section 4.5. Running an integration job on an ad hoc basis.

4.2. Viewing and editing an integration job

To view an integration job:

1. Click the Administrator tab in the top navigation.
2. Click the Jobs tab in the navigation.

   The Scheduled Cron Jobs page is displayed with a list of scheduled jobs.



3. Click the name of the integration job you want to view or edit to open the Job Name page.

   Note

   Click *Success*, *Warnings* or *Errors* to automatically expand the respective log entries.

4. Use the Schedule area to view or edit the job schedule. See Section 4.4. Scheduling an integration job.
5. Use the Configuration area to view or edit the job settings. See Section 4.3. Configuring an integration job.
6. Use the Log entries area to view information about the status of the job when it was run. See Section 4.6. Viewing the status of an integration job.

4.3. Configuring an integration job

When you create or edit an integration job, you can change details about the job (including how fields in Pure are affected by the job) using the Configuration area.

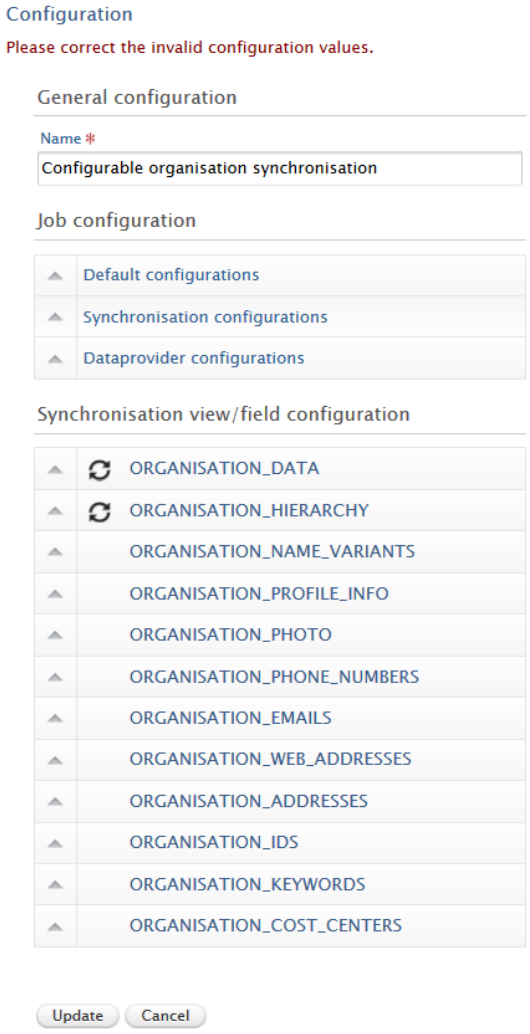
4.3.1. Opening the Configuration area

To open the Configuration area:

1. Open the Job Name page, either by creating a new integration job or viewing a scheduled job. See Section 4.1. Creating an integration job and Section 4.2. Viewing and editing an integration job.

2. If the Configuration area is not expanded, click Edit configuration.

   The Configuration area is displayed.



Note

The items in the *Synchronisation view/field configuration* area may be different depending on the content type of the integration job.

Tip

There is always an error message displayed when creating a new job until the mandatory fields have been completed.

4.3.2. Configuration area fields and buttons

General configuration area

Use the General configuration area to set basic values for the job.

Field/Button	Description
Name	The name of the job. The job name is used to identify the job on the: • Scheduled Cron Jobs page. • Job management page. • Job log page.

Job configuration area

Use the expanding menus in the Job configuration area to configure technical settings relating to the execution of the implementation job.

> **Note**
> Some job-specific configuration settings are not described here. See the descriptions in the Pure interface.

Field/Button	Suggested value
Synchronisation safe size	A sensible size based on the number of records available. **Note** When the integration covers your institution's complete records, enter a number slightly lower than the expected number of records in the external system. If the number of records in the external system is below this safe size the job will report an error and will not make any modifications in Pure. This avoids starting an integration where the source data was possibly corrupted or lost.
Development/Debug mode	Off. If On, verbose logging is enabled.
Number of job logs to keep	If 0 then all logs will be kept. Set this value as a sensible number considering the schedule of this job and internal policies.
Synchronisation administrator e-mail	Your email address, so that you will receive emails about errors as configured in the *E-mail when log level exceeds* field.
E-mail when log level exceeds	The minimum error level that triggers an email to be sent to the email address in the *Synchronisation administrator e-mail* field. Choose to receive emails for jobs with: • All jobs, even if there are no warnings or errors (INFO). • Warning messages (WARN). • Error messages (ERROR). • Fatal errors (FATAL). **Recommendation** We recommend that you continue to improve the data and transformation process for this integration job until there are no errors (and ideally no warnings) when running the job. You can then choose to be alerted when errors or warnings occur.
Continue even if duplicate IDs exist	Off. Set to On to allow multiple records with the same ID to be integrated, which may result in errors in the data. **Recommendation** Once you have gone live with Pure, ensure this is set to Off.
Try to save content with synchronisation errors	Off. If Off, records that raise an error during synchronization are not saved. Set to On to allow content that raises an error during synchronization but is still valid to be saved.

Note
Refer to the detailed explanations within the area in the Pure interface for a description of the field behavior.

Synchronisation view/field configuration area

Use the expanding menus in the Synchronisation view/field configuration area to manage which fields are updated during an integration.

The expanding menus available are different for each content type, though there are some common symbols and options:

Field/button	Description
	This icon indicates that at least one field described in this menu is set to *Sync* or *Sync once*.
Mandatory synchronisation fields Field name (capital letters)	For the given field name, choose how the field is updated. Choose either: • *Sync* to update the field in Pure to the value from the external data source each time the job is run. Fields populated in this manner cannot be edited in the Pure interface. • *Sync once* to set the field in Pure to the value from the external data source the first time this field has content, but do not change the field's value on consequent runs. Once the field is populated it is maintained in the Pure interface, so it is no longer updated from the external data source.
Synchronisation fields Field name (capital letters)	For the given field name, choose how the field is updated. Choose either: • *Sync* to update the field in Pure to the value from the external data source each time the job is run. Fields populated in this manner cannot be edited in the Pure interface. • *Sync once* to set the field in Pure to the value from the external data source the first time this field has content, but do not change the field's value on consequent runs. Once the field is populated it is maintained in the Pure interface, so it is no longer updated from the external data source. • *Do not sync* to leave the field unchanged by the job. The field is editable in the Pure interface and only maintained there, regardless of the values in the external data source. • *Lock* to leave the field unchanged by the job and make the field read-only in the Pure interface. Only set a field to *Lock* when there is no current or correct data available for the field in the external data source, and the field should not be edited in Pure. The field will appear blank and disabled in the Pure interface. **Warning** Beware of changing values from *Do not sync* to *Sync* as the job will replace all data in these fields with the values from the external data source.
Enable synchronization	Choose either: • On to allow synchronization from an optional item. • Off to disallow synchronization from an optional item.
Lock relation list	Choose either: • *Lock* if the external data source is the complete and authoritative source of data for this content type, and no edits that add additional relations are allowed for any records synchronized by this job. • *Do not lock* to allow both synchronizations with an external data source and users working in the Pure interface to add relations from records synchronized by this job, so that the record in Pure may have additional changes that are not reflected in the external data source.
Toggle common / specific configurations	Choose either: • On to show the *Specific synchronisation fields* area. • Off to show the *Common synchronisation fields* area.
Common synchronisation	Configure how the field is updated for multiple fields.

fields	
Specific synchronisation fields	Configure how the field is updated for each field specifically.

For more information about configuring these fields for a specific integration job, see Section 6. Importing data into Pure.

4.4. Scheduling an integration job

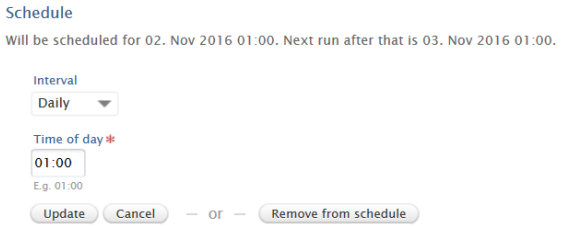
You can change how regularly and at what time an integration takes place using the Schedule area.

4.4.1. Opening the Schedule area

To open the Schedule area:

1. Click the Change schedule button on the Cron Job Schedulingpage when you are creating a new integration job. See Section 4.1. Creating an integration job.

   The Schedule area is displayed.



4.4.2. Schedule area fields and buttons

Field/Button	Description
Interval	How often this job is run. Either: • Daily • Weekly • Monthly • Quarterly • Yearly • Manual • Custom Some options will prompt you to choose the particular day, week or month that the integration is performed within this interval. **Note** If you choose *Custom*, you can define a job interval using regular expressions and click *Help With Cron Expressions* for an explanation of the possible values. If you choose *Manual*, the integration is only performed when you run the job on an ad hoc basis. See Section 4.5. Running an integration job on an ad hoc basis.
Time of day	The hour of the day when the integration is performed. Enter the hour in 24–hour in HH:MM format. For example, to perform an integration at eleven o'clock at night (11pm), enter **23:00**.
Update	Update the integration job with any changes made to the schedule and hide the Schedule area.
Cancel	Cancel any changes made to the schedule of the integration job and hide the Schedule area.
Remove from schedule	Remove any integration jobs of this type from the Scheduled Cron Jobs list. To add the job again, see Section 4.1. Creating an integration job. **Note** Clicking this button deletes the current integration job.

4.5. Running an integration job on an ad hoc basis

You can always start an integration job manually (ie. run the job on an ad hoc basis) regardless of whether it has been scheduled or not. If the job is scheduled to run at a certain time, running a job on an ad hoc basis does not affect the execution of the job as scheduled, rather runs the job once in addition to the scheduled runs.

To run an integration job manually, either:

- **Run the job from the Scheduled Cron Jobs page:**

  1. Click the Administrator tab in the top navigation.
  2. Click the Jobs tab in the navigation.
     The Scheduled Cron Jobs page is displayed with a list of scheduled jobs.
  3. Click Start Job now under the job you want to run.

- **Run the job from the Job Name page of the Cron Job Scheduling tab:**

  1. Click the Administrator tab in the top navigation.
  2. Click the Jobs tab in the navigation.

     The Scheduled Cron Jobs page is displayed with a list of scheduled jobs.
  3. Click the name of the job you want to run.

     The Job Name page of the Cron Job Scheduling tab is displayed.
  4. Click Start Job now in the Schedule area.

     The job is run and you remain on the Job Name page of the Cron Job Scheduling tab.

After the job has been run you can view the status of the job. See Section 4.6. Viewing the status of an integration job.

4.6. Viewing the status of an integration job

You can view the status, such as the success or failure of a job, in the log entry details area of the Job log area.

4.6.1. Opening the log entry details area

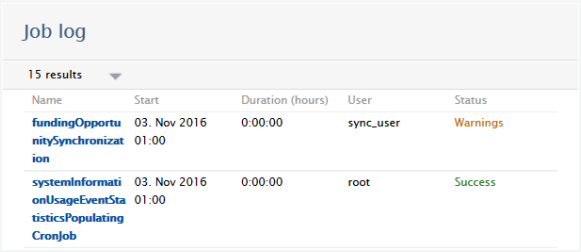
To open the log entry details area for a job, either:

- **View all jobs' logs from the Job log page.**

  1. Click the Administrator tab in the top navigation.
  2. Click the Jobs tab in the navigation.

     The Scheduled Cron Jobs page is displayed with a list of scheduled jobs.
  3. Click the Job log tab in the navigation.

     The Job log page is displayed.

  4. Click the name of the job you want to view log entries for.

     A detailed list is displayed with the results of each run.

  5. Click the *Start* column for the run you want to view details for.

     The log entry detail area is expanded.

     Tip

     You can also click in the *Success*, *Warnings*, *Errors* or *Fatal errors* columns for the run to show only the log details with that status.

- **View a specific job's logs from the Cron Job Scheduling page.**

  1. Click the Administrator tab in the top navigation.
  2. Click the Jobs tab in the navigation.

     The Scheduled Cron Jobs page is displayed with a list of scheduled jobs.
  3. Click the name of the job you want to run.

     The Job Name page of the Cron Job Scheduling tab is displayed.
  4. Navigate to the Log entries area.
  5. Click the name of the job you want to view log entries for.

     A detailed list is displayed with the results of each run.

  6. Click the *Start* column for the run you want to view details for.

     The log entry detail area is expanded.

     Tip

     You can also click in the *Success*, *Warnings*, *Errors* or *Fatal errors* columns for the run to show only the log details with that status.

- **View the log for the most recent run of each job from the Cron Job Scheduling page:**

  1. Click the Administrator tab in the top navigation.
  2. Click the Jobs tab in the navigation.

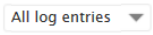
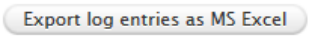
     The Scheduled Cron Jobs page is displayed with a list of scheduled jobs.
  3. In the last run status area, click *Success* or *Warnings*.

     The Job Name page is displayed with the log entry details area expanded.

Note

If the log you are looking for is not visible, it may have been deleted. You can configure the number of logs retained using the *Number of job logs to keep* job configuration setting. See Section 4.3. Configuring an integration job.

4.6.2. Using the log entry details area

Field/Button	Description
	Filter the log entries shown to: • All log entries. • Entries with a status of success. • Entries with a status of warning. • Entries with a status of error. • Entries with a status of fatal error.
	Enter a keyword and press **Enter** to show only log entries with one or more lines that exactly contain the search query. **Note** You cannot search for multiple keywords unless they appear exactly in the order given in one or more log entry lines. Search for an empty string to show all entries. **Tip** Search for part of the error message that does not relate to a specific record to see how commonly this error occurred during the job's execution.
	Generate and download an *XLS* file with the content of the *Title and description* field for the current filter and search selections.
Status	The status of the log entry.
Title and description	The title of the log entry, followed by warnings and information about the log entry.

4.6.3. Integration job error messages

Below are a few sample error messages relating to integration jobs.

Error message	Description
WARN: No classification found (Scheme: /dk/atira/pure/person/gender, mapKey: somevalue, mapping defined: yes, and map returned: null, default mapping defined: no)	A value "somevalue" was given. The value was expected to match one of the values of the classification scheme /dk/atira/pure/person/gender.
ERROR: Validation error - Gender is required - Exception details	No gender (or an invalid gender value) was specified in the XML. Gender is mandatory and therefore when the import tries to validate the content (in this case the Person content type) it fails and gives an error. **Note** You can expand the *Exception details* area, though this information may only be meaningful to support staff.
WARN: The organisation organisation1 ends on 2050-01-01 12:00:00.0, but the relation with id staffAssoc1 does not end before this date	If an Organization has an end date, a Person relation to this Organization must have an end date which is earlier than or equal to that date.

4.7. Using the Bulk Import Wizard to import external data

You can use the Bulk Import Wizard to populate Pure with data from an external data source before configuring and running the integration job that synchronizes these records on a regular basis. You may choose to import:

- All of the records in the external data source. When you run the integration job for the first time, records are updated as determined in the job's configuration.
- Some of the records in the external data source. The remaining records will be imported the first time you run the integration job.

To import external data into Pure using the Bulk Import Wizard:

1. Click the Administrator tab in the top navigation.
2. Click the Bulk import tab in the navigation.
3. Click the name of the content type you want to import or migrate in the navigation below the Bulk import tab.

   The Import wizard page is opened for the wizard you selected.
4. Review the following documents available on the Import wizard page:
   - the example data XML files for the content type.
   - the XSD files for the content type.
   - the User Guide PDF for the wizard for information about importing or migrating data.
5. Generate a well-formed and valid XML file from your external system with the records you want to populate.

   Note

   This is the same XML format as used for an integration job for this content type. When using the Bulk Import Wizard it is possible to import a subset of the records at one time, whereas when performing an integration job you should not exclude records that have already been integrated.
6. Click Open import wizard.

   The Bulk Import Wizard is opened and guides you through the import process.

4.8. Integrating data sources with Pure - detailed steps

These steps describe in detail how to integrate a content type from one external data source into Pure using the additive strategy described in Additive import strategy in brief (sec. 1.2.2).

In contrast to other tutorials, these detailed guides:

> describe strategies for approaching the amount and type of data to import when testing the import procedure.
> guide you through the appropriate steps to correct errors based on what kinds of errors are visible in the Pure interface.

You can see less-detailed instructions specific to each content type in Importing data into Pure (sec. 6).

1. Begin the integration process by reviewing the extraction and transformation processes that convert records from your data source into the XML format that can be imported into Pure. The best way to do this is to import some records using the Bulk Import Wizard, as described in the migration tutorial. See Section 2. Preparing data sources and Section 3. Migrating legacy data into Pure.

2. Review which records of this content type have already been imported into Pure.

   Note

   Do not re-run integrations where you are missing records that have previously been integrated. If those records were current, they will be marked as marked as historical and given an End date, though this can be corrected by including them in a future integration job. This does not apply to records that were only previously imported with the Bulk Import Wizard.

3. Plan how each field from your data source will be managed in Pure. For example, should fields be read-only and editable only in the external data source or writable from within Pure? Decide whether users should be able to create new records with this content type from within Pure.
4. Configure an integration job with your chosen sync settings. Do not yet set strict requirements for the job to proceed such as a minimum number of records. See Section 4.3. Configuring an integration job.
5. Run the integration job once. See Section 4.5. Running an integration job on an ad hoc basis.
6. Check the status of the job. See Section 4.6. Viewing the status of an integration job.
   - If the job succeeded, proceed to the following steps.
   - If there are errors or warnings for the job, address these by changing the job configuration or by cleaning the input data. If necessary, review your extraction, transformation or mapping, extract and transform your data again, and proceed to step 4 or 5.
7. Check some existing records in Pure that you expect to have been updated by the integration job.
   - If the data did not change when they should have changed:
     - Check the job log for errors.
     - Check that the data is delivered in the XML data feed.
     - Check that the sync settings are set as expected for the field.
     - Check that the IDs match and have not changed.

       IDs must be unique and persistent for all integrated records.

     Resolve these errors and proceed to step 3.
   - If the data changed when it should not have changed, edit the sync settings to resolve the problem. See Section 4.3. Configuring an integration job. Proceed to step 5.
8. Check whether fields can be edited in the Editor or Master data tabs.
   - If this behavior is expected, proceed to the following steps.
   - If this behavior is unexpected, edit the sync settings to resolve the problem. Proceed to step 5.
9. Check whether all of the records of this content type have now been imported.
   - If all records are present, proceed to the following steps.
   - If you are missing records, consider any errors reported for the job. See Section 4.6. Viewing the status of an integration job. Review the processes that extract and transform the data from your external system. Proceed to step 5.
10. Change the following integration job settings to strict values so that any problems with the input cause the job to fail and send an email notification:
    - *Synchronisation safe size*
    - *E-mail when log level exceeds*
    - *Continue even if duplicate IDs exist*
    - *Try to save content with synchronisation errors*
    - *Continue job even if validation fails*
11. Configure your external data source to regularly export the data to be integrated and transform this to XML.
12. Schedule the regular execution of the integration job. See Section 4.4. Scheduling an integration job.

    The integration setup is complete.

5. Evaluating the quality of imported data

Evaluate the quality of the data imported with the Bulk Import Wizard or with an integration job by:

1. **Inspecting the logs for the import job**
   Addressing errors and warnings in the logs is critical to ensuring high data quality and should be the first step after importing data. See Section 5.1. Using the job logs to evaluate import quality.
2. **Inspecting certain complex or problematic records manually**
   For each content type there are certain areas that most commonly cause problems either in the mapping or transformation. See below for some tips about which fields to pay particular attention to.

5.1. Using the job logs to evaluate import quality

Review the job logs for imports with the Bulk Import Wizard and for integration jobs to evaluate the success of the import.

See Section 3.2. Viewing the status of an import from the Bulk Import Wizard and Section 4.6. Viewing the status of an integration job.

When troubleshooting imports from the logs, it may help to:

- Filter the log for warnings and errors.
- Read the warning and error messages, and review the record that corresponds to the problematic value in the source data set and XML.
- Export the log to Microsoft Excel to examine it further.
- Search for part of the error message that does not relate to a specific record to see how commonly this error occurred during the job's execution.
- Bulk delete problematic content so that the job runs from a clean state when it is run the next time. See Section 5.4. Bulk deleting records.

> **Tip**
> You can configure an integration job to send an email when errors occur in the *E-mail when log level exceeds* field. See Section 4.3. Configuring an integration job.

5.2. Fixing incorrectly integrated records

If you have run an integration job but found that there are errors in the data due to a problem with the mapping or transformation to XML, you can fix the affected records by correcting the processes that caused the error and re-running the integration job.

> **Note**
> This process will leave an audit trail related to the Pure ID of the record, though it is preferable to bulk-deleting and re-importing all records, which can be a cumbersome process.

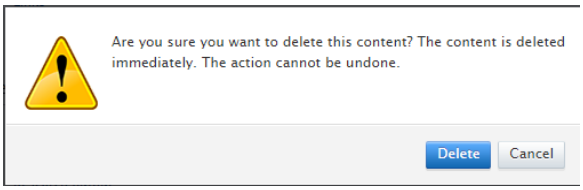
5.3. Deleting individual records

To delete a record from Pure:

1. In the top navigation click:
   - the Editor tab if the content type is a piece of research output, activity, course, prize, press/-media, application, award, project, ethical review, impact, dataset, student thesis, or CV.
   - the Master data tab if the content type is user, (external) person, (external) organization, journal, publisher, event, funding opportunity, or equipment/facility.
2. Click the tab for the content type of the record in the navigation.

3. Search for the record or records you want to delete:
    - Type keywords in the search bar and click Search to search by keywords.
    - Click in the search bar to open a dialog with advanced search options.
    - Click , select a variable to filter on, and make the relevant selections in the context-sensitive area to add a filter.
    - Click next to a filter to remove it.
    - Click to save the current filter selections as a filter set accessible from the navigation.
4. Click on the record to open the editor window.
5. Click .

   A dialog appears, prompting you to confirm your decision.



> **Note**
> If the record has relations to other content (such as being linked to a piece of Research output or an Organization), you must first either remove the affiliations or delete the affiliated content.

6. Click Delete.

   The record is deleted.

5.4. Bulk deleting records

You can bulk delete records that all share the same content type. If you want to bulk delete records of various different content types, perform a bulk delete once for each distinct content type.

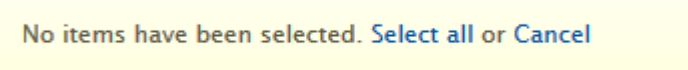
To bulk delete records:

1. In the top navigation click:
    - the Editor tab if the content type is a piece of research output, activity, course, prize, press/-media, application, award, project, ethical review, impact, dataset, student thesis, or CV.
    - the Master data tab if the content type is user, (external) person, (external) organization, journal, publisher, event, funding opportunity, or equipment/facility.
2. Click the tab for the content type of the records in the navigation.
3. Search for the record or records you want to delete:
    - Type keywords in the search bar and click Search to search by keywords.
    - Click in the search bar to open a dialog with advanced search options.
    - Click , select a variable to filter on, and make the relevant selections in the context-sensitive area to add a filter.
    - Click next to a filter to remove it.
    - Click to save the current filter selections as a filter set accessible from the navigation.

> **Note**
> Use search terms and filters so that only the records you want to delete are shown.

4. Click ✓✓ at the top of the search results area.

   A banner with the number of records selected is displayed.

5. Click *Select all* on the banner to select all the records.

   Selected records show ☑.

6. Deselect the checkbox (☐) next to individual records that should not be deleted.

7. Click Bulk edit on the banner.

   The Bulk Edit window is displayed.

8. Select *Delete content*.

9. Click Next.

10. Click Apply the changes to delete all of the selected records.

    The records are deleted.

Note
Deleted records may still be visible in some areas of Pure until the Preserved Content Update job is run.

6. Importing data into Pure

The following topics describe how to import data into Pure, with instructions specific to each content type.

6.1. Populating Organization data

> **Tip**
> Before you begin, consider reading the introductory section of this manual. See Section 1. Introduction.

About

In Pure, Organizations represent sections of your institution that have a hierarchical relationship with one top-level organization, for example a university. Typical organization types include university, faculty, or department, though these can be customized for your needs.

Requirements

Organizations must be populated in Pure before all other content types because all other content must be linked to at least one organization.

See the following related topics:
> Populating content in the correct order (sec. 1.2.1)

> **Tip**
> It is vital that organization hierarchy is of a high quality before you proceed with other imports.

Data preparation

When preparing data for an import, consider the following tips:

See the following related topics:
> Preparing data sources (sec. 2)
> Pure's XML import format and the XSD (sec. 2.4)
> Mapping types and fields to the Pure data model (sec. 2.6)
> Mapping external values to classifications (sec. 2.7)

- If your institution will use the Pure Portal, include the information that should be visible there. For example, a photo, contact information, address.
- Fill all available fields relating to contact information for each organization.
- Be aware of duplicate or similar organization names. If similar names represent name changes of an organization, be sure to check that this is appropriately reflected.
- Pay special attention to the start and end date of an organization. These determine whether an organization is considered **active** or not.
- Consider any uncommon types of organizations that do not produce research output that might exist in your external data source, such as cleaning offices or external partners. Ensure they are excluded from the transformation.
- If an organization has been taken over by another organization, ensure that this relationship is included.
- All organizations must be placed in a hierarchy with a single organization acting as the root of all others. This hierarchy is constructed based on the parent organization given for each entry. Ensure that a parent organization is always specified, and that this parent organization is also included in the import.

> **Tip**
> For some institutions the placement of organizations in a hierarchy may not be immediately clear. If no clear hierarchy can be agreed upon, it is possible to begin with a very simple hierarchy with one root organization that has all other organizations as its children. However this is not recommended if it does not represent the official hierarchy at your institution as the organizational hierarchy is also partially used in the publicly-visible Pure Portal.

- The Organization content type includes the fields *Profile information* and *Research information*, which allow for longer descriptions about the record that can be formatted with HTML. If HTML formatting is causing problems, consider restricting the content of this field to plain text.

- Ensure all of the mandatory fields are complete. You can see which fields are mandatory by adding a new record in the Pure interface and looking for fields with a * symbol next to the field name.

Data validation

When evaluating an import, pay attention to these areas where problems occur most commonly:

See the following related topics:

> Evaluating the quality of imported data (sec. 5)
> Bulk deleting records (sec. 5.4)

- Check that the appropriate organization is at the top of the hierarchy and that there is only one organization in this position.
- Check that the number of organizational levels is as intended.
- Check that the different levels of the hierarchy represent a functional hierarchy of organization types. For example, a university at the top level, followed by faculty, department etc.
- If HTML-formatted text fields are included in the import, check that these are formatted correctly.
- If an organization you are importing has previously been taken over by another organization, check it has an *End date* and that the relationship to the new parent organization exists.

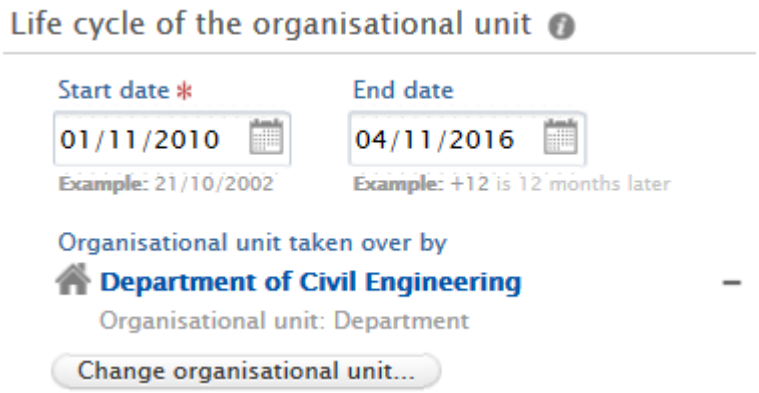


Figure 10. Life cycle of the organisatonal unit area on the Metadata tab of the Organisational unit editor window.

- Check that each field has a single, clearly-defined use, ideally reflected in the name of the field. There should be no fields that contain miscellaneous information from the data source.
- Check for fields that are consistently empty. Though it is not mandatory to populate all fields, it could be a warning that you have forgotten to map certain values or mapped them incorrectly.

Integration with an external data source

To integrate Pure with your institution's external data source for OrganizationFacilities/Equipment data:

1. Create an integration job using the job named *Configurable organisation synchronisation*.

> Creating an integration job (sec. 4.1)

2. Configure the integration job based on the dataset you are integrating.

> Configuring an integration job (sec. 4.3)

Note

If you have not previously imported records with this content type, we recommend that you start by integrating a small number of records so that you can manually check the accuracy of your data mapping and transformation processes.

3. Schedule a time for the job to run or run the job immediately.

> Scheduling an integration job (sec. 4.4)
> Running an integration job on an ad hoc basis (sec. 4.5)

4. After the job has run, use the logs to check whether it succeeded and which errors occurred. If no records were added to Pure, address the listed errors and try to run the integration job again.

> Viewing the status of an integration job (sec. 4.6)
> Using the job logs to evaluate import quality (sec. 5.1)

5. Check the status of the data you integrated and delete incorrect records.

   Depending on the configuration of the integration job, some records may be unable to be deleted in Pure. In this case, change the configuration settings and run the integration job again to update incorrect records.

> Evaluating the quality of imported data (sec. 5)
> Configuring an integration job (sec. 4.3)

6. If you deleted records, run the integration job again and check that both deleted and preserved records exist as expected. If the records have *Start date* and *End date* fields, check that these are correct.

7. Once you are satisfied that all records are in Pure and all fields are correctly configured, configure the integration job as you want it to run regularly, and set up your external data source to provide the source XML at the appropriate time.

> Configuring an integration job (sec. 4.3)
> Scheduling an integration job (sec. 4.4)

Tip

For general information about integrations, see Section 4. Integrating external data into Pure.

For detailed steps about how to select records and verify the correctness of your data when performing an integration, see Section 4.8. Integrating data sources with Pure - detailed steps.

6.2. Populating User data

Tip

Before you begin, consider reading the introductory section of this manual. See Section 1. Introduction.

About

Importing users creates user profiles, based on a username and email address.

Users can only log into Pure once their user profile has been linked to a Person record, or additional roles have been assigned.

Person records in Pure capture information about people affiliated with your institution, such as academic staff members. Records include information such as their name, employment information and qualifications. Person records are linked to other records, such as research output, and can be shown as a profile on the Pure Portal.

A **user** record represents a user account for someone at your institution who receives credentials for Pure, and logs in to manage data. Users receive roles which determine how they interact with content in Pure. If a user is associated with a person record, they can add data relating to their own person record such as research output or activities.

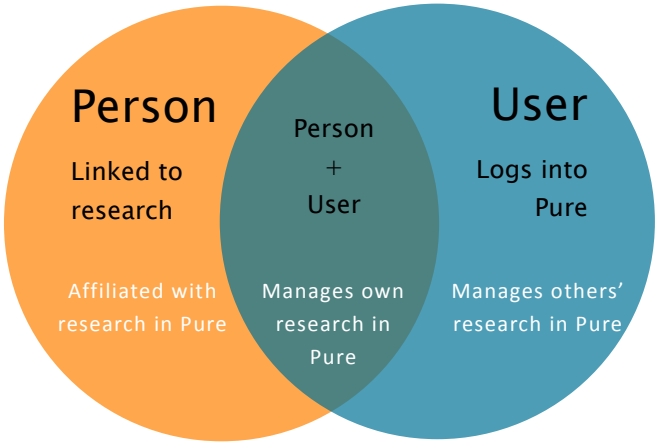


*Figure 11. Not all users are associated with **Person** records. For example, research managers or administrative staff may use Pure without having a corresponding **Person** record.*

In addition to **person**, Pure also has the content type **External person**. These are records of individuals that are not affiliated with your institution. For example, co-authors on articles and project collaborators.

Note

External persons are created automatically when a relationship is added that cannot be matched to a profile in Pure.

Requirements

We recommend that you import Users into Pure directly after the Organization content type. Users must be created before the Person content type is populated so that users and persons can be linked.

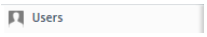
See the following related topics:

> Populating content in the correct order (sec. 1.2.1)

If you are using an external authentication system, you can import user information from this system to create the appropriate user credentials in Pure. You can configure aspects of the authentication mechanisms related to usernames on the Security tab of the Administrator tab.

Tip

If you do this using the *Configurable User Synchronization* job, then the field **Authenticated by external system** is set to true for all users synchronized by that job.

You can also create users manually in Pure using the icon next to the 

tab of the 

tab.

Note

Although email addresses are required when creating users, the default behavior is that no emails are sent from Pure until after the first time user has logged in successfully.

Tip

Users that are associated with a Person record automatically receive the role *Personal user*. Other roles need to be assigned manually in the Pure interface on the *Roles* tab of the User editor window.

Data preparation

When preparing data for an import, consider the following tips:

See the following related topics:

> Preparing data sources (sec. 2)
> Pure's XML import format and the XSD (sec. 2.4)
> Mapping types and fields to the Pure data model (sec. 2.6)
> Mapping external values to classifications (sec. 2.7)

- Include a complete list of current active users only. Users that once appeared in the data feed that no longer appear will either no longer be able to log in (locking), or will have their credentials deleted from Pure (deletion). The retirement strategy used depends on the *User retire strategy* setting on the *Configurable User Synchronisation* page.
- Depending on the external authentication system your institution is using, the username in Pure may need to be exactly the username in any external authentication system being used. If you are not using external authentication, you can use any static, unique username.
- If you are using Shibboleth, the username must be the correct SAML attribute that holds the user identifier **eduPersonPrincipalName**.
- Ensure all of the mandatory fields are complete. You can see which fields are mandatory by adding a new record in the Pure interface and looking for fields with a  symbol next to the field name.

Data validation

When evaluating an import, pay attention to these areas where problems occur most commonly:

See the following related topics:

> Evaluating the quality of imported data (sec. 5)
> Bulk deleting records (sec. 5.4)

- When importing persons, check that usernames corresponding to persons are correctly linked to the person record.
- If you are using external authentication, set up and test your authentication system at this stage. Test whether users you imported can access the system correctly.
- Check whether the imported users are visible on the *Master data* tab.

Note

The roles available to users are managed solely from inside Pure and cannot be imported through the Bulk Import Wizard or an integration job. If you delete then re-import users, be aware that you will have to manually make changes to any roles.

- Check that each field has a single, clearly-defined use, ideally reflected in the name of the field. There should be no fields that contain miscellaneous information from the data source.
- Check for fields that are consistently empty. Though it is not mandatory to populate all fields, it could be a warning that you have forgotten to map certain values or mapped them incorrectly.

Integration with an external data source

1. Create an integration job using the job named *Configurable user synchronisation*.

   > Creating an integration job (sec. 4.1)

2. Configure the integration job based on the dataset you are integrating.

   > Configuring an integration job (sec. 4.3)

   Note
   If you have not previously imported records with this content type, we recommend that you start by integrating a small number of records so that you can manually check the accuracy of your data mapping and transformation processes.

3. Schedule a time for the job to run or run the job immediately.

   > Scheduling an integration job (sec. 4.4)
   > Running an integration job on an ad hoc basis (sec. 4.5)

4. After the job has run, use the logs to check whether it succeeded and which errors occurred. If no records were added to Pure, address the listed errors and try to run the integration job again.

   > Viewing the status of an integration job (sec. 4.6)
   > Using the job logs to evaluate import quality (sec. 5.1)

5. Check the status of the data you integrated and delete incorrect records.

   Depending on the configuration of the integration job, some records may be unable to be deleted in Pure. In this case, change the configuration settings and run the integration job again to update incorrect records.

   > Evaluating the quality of imported data (sec. 5)
   > Configuring an integration job (sec. 4.3)

6. If you deleted records, run the integration job again and check that both deleted and preserved records exist as expected. If the records have *Start date* and *End date* fields, check that these are correct.

7. Once you are satisfied that all records are in Pure and all fields are correctly configured, configure the integration job as you want it to run regularly, and set up your external data source to provide the source XML at the appropriate time.

   > Configuring an integration job (sec. 4.3)
   > Scheduling an integration job (sec. 4.4)

   Tip
   For general information about integrations, see Section 4. Integrating external data into Pure.
   For detailed steps about how to select records and verify the correctness of your data when performing an integration, see Section 4.8. Integrating data sources with Pure - detailed steps.

6.3. Populating Person data

Tip

Before you begin, consider reading the introductory section of this manual. See Section 1. Introduction.

About

Person records in Pure capture information about people affiliated with your institution, such as academic staff members. Records include information such as their name, employment information and qualifications. Person records are linked to other records, such as research output, and can be shown as a profile on the Pure Portal.

A **user** record represents a user account for someone at your institution who receives credentials for Pure, and logs in to manage data. Users receive roles which determine how they interact with content in Pure. If a user is associated with a person record, they can add data relating to their own person record such as research output or activities.

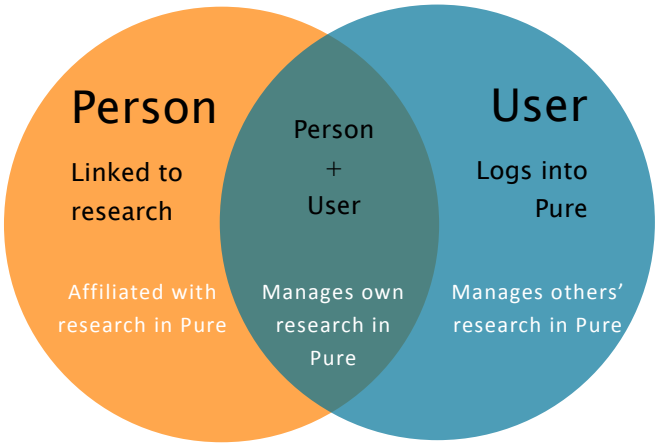


*Figure 12. Not all users are associated with **Person** records. For example, research managers or administrative staff may use Pure without having a corresponding **Person** record.*

In addition to **person**, Pure also has the content type **External person**. These are records of individuals that are not affiliated with your institution. For example, co-authors on articles and project collaborators.

Note

External persons are created automatically when a relationship is added that cannot be matched to a profile in Pure.

Requirements

The Person content type must be imported after Organizational units but before any other content type because every other content type is linked to a person as well as an Organization.

Importing Persons does not create user accounts for Pure. See Section 6.2. Populating User data.

See the following related topics:

> Populating content in the correct order (sec. 1.2.1)

Data preparation

When preparing data for an import, consider the following tips:

See the following related topics:
> Preparing data sources (sec. 2)
> Pure's XML import format and the XSD (sec. 2.4)
> Mapping types and fields to the Pure data model (sec. 2.6)
> Mapping external values to classifications (sec. 2.7)

- Pay special attention to ID fields, such as ORCID, Scopus Author ID etc. These are used later when matching research output to persons.
- If your institution will use the Pure Portal, include the information that should be visible there. For example, a photo, contact information, address.
- If you are using email addresses as a unique ID field, ensure that the capitalization of the addresses does not change. This could result in duplicate person entries being added.
- Consider which of the following affiliations your institution is using as the mandatory relation ID:
  - StaffOrganisationRelationID.
  - StudentOrganisationRelationID.
  - StudentFundingRelationID.
  - SupervisorRelationID.

  Ensure that the IDs are unique.
- If you want the ID of the record to be visible in the Pure interface, consider additionally including it as part of the **externalIds** field in the XML.
- Ensure all of the mandatory fields are complete. You can see which fields are mandatory by adding a new record in the Pure interface and looking for fields with a symbol next to the field name.

Data validation

When evaluating an import, pay attention to these areas where problems occur most commonly:

See the following related topics:
> Evaluating the quality of imported data (sec. 5)
> Bulk deleting records (sec. 5.4)

- Check the number of Persons visible on the Persons tab of the Master data tab.
- If Users have also been imported, make sure that the appropriate User and Person records have been linked (these are referred to as **Personal users**). Check the count of *Personal user (staff)* on the Master data overview tab of the Master data tab.
- Test on employees who have been associated with your institution for a long time and who you have a lot of information for, especially former roles.
  - Are the organizational affiliations correct?
  - Do their former roles/affiliations have correct start/end dates?
  - Does their FTE add up to 100%?
  - Check that the supervisor relations are correct.
- Test on an employee who also had a Pure username created. Make sure that the person record and the user record are correctly linked.
- Ensure that IDs are correct and present, including any additional legacy ID codes or unique values such as email addresses that might be referenced in other records. Relations between records in Pure are created based on various matching criteria, but the most accurate matches rely on unique codes such as IDs.
- Look at the *Master data overview* tab of the *Master data* tab to see an overview of the number of Person records that were imported, paying attention to the number and whether they are listed as **Former** or **Current**.
- Click on the *Person* tab of the *Master data* and view the options that appear under the *Person* tab in the side banner that list duplicate records. Consider why duplicates may be appearing — is there a problem with the import process? If there duplicates in your data source, consider whether you want to perform de-duplication within Pure or whether you can easily remove certain duplicate records before importing the data.
- Check that each field has a single, clearly-defined use, ideally reflected in the name of the field. There should be no fields that contain miscellaneous information from the data source.
- Check for fields that are consistently empty. Though it is not mandatory to populate all fields, it could be a warning that you have forgotten to map certain values or mapped them incorrectly.

Integration with an external data source

1. Create an integration job using the job named *Configurable person synchronisation*.

   > Creating an integration job (sec. 4.1)

2. Configure the integration job based on the dataset you are integrating.

   > Configuring an integration job (sec. 4.3)

   Note

   If you have not previously imported records with this content type, we recommend that you start by integrating a small number of records so that you can manually check the accuracy of your data mapping and transformation processes.

3. Schedule a time for the job to run or run the job immediately.

   > Scheduling an integration job (sec. 4.4)
   > Running an integration job on an ad hoc basis (sec. 4.5)

4. After the job has run, use the logs to check whether it succeeded and which errors occurred. If no records were added to Pure, address the listed errors and try to run the integration job again.

   > Viewing the status of an integration job (sec. 4.6)
   > Using the job logs to evaluate import quality (sec. 5.1)

5. Check the status of the data you integrated and delete incorrect records.

   Depending on the configuration of the integration job, some records may be unable to be deleted in Pure. In this case, change the configuration settings and run the integration job again to update incorrect records.

   > Evaluating the quality of imported data (sec. 5)
   > Configuring an integration job (sec. 4.3)

6. If you deleted records, run the integration job again and check that both deleted and preserved records exist as expected. If the records have *Start date* and *End date* fields, check that these are correct.

7. Once you are satisfied that all records are in Pure and all fields are correctly configured, configure the integration job as you want it to run regularly, and set up your external data source to provide the source XML at the appropriate time.

   > Configuring an integration job (sec. 4.3)
   > Scheduling an integration job (sec. 4.4)

   Tip

   For general information about integrations, see Section 4. Integrating external data into Pure.

   For detailed steps about how to select records and verify the correctness of your data when performing an integration, see Section 4.8. Integrating data sources with Pure - detailed steps.

6.4. Populating Supervisor data

Tip

Before you begin, consider reading the introductory section of this manual. See Section 1. Introduction.

About

A Supervisor is not a content type in Pure, rather a relation between two Persons where one supervises the other.

Supervisor information is imported into Pure when the Person content type is populated and should not be populated independently.

Nonetheless, you can use an integration job to update Supervisor information from an external system.

Requirements

Supervisor records can be populated as soon as the content types Organization and Person have been populated.

When scheduling the integration of Supervisor data from an external system, the job should be run after the regular integration of Person data.

See the following related topics:

> Populating content in the correct order (sec. 1.2.1)

Data preparation

When preparing data for an import, consider the following tips:

- Ensure all of the mandatory fields are complete. You can see which fields are mandatory by adding a new record in the Pure interface and looking for fields with a symbol next to the field name.

See the following related topics:

> Preparing data sources (sec. 2)
> Pure's XML import format and the XSD (sec. 2.4)
> Mapping types and fields to the Pure data model (sec. 2.6)
> Mapping external values to classifications (sec. 2.7)

Data validation

When evaluating an import, pay attention to these areas where problems occur most commonly:

- Check that each field has a single, clearly-defined use, ideally reflected in the name of the field. There should be no fields that contain miscellaneous information from the data source.
- Check for fields that are consistently empty. Though it is not mandatory to populate all fields, it could be a warning that you have forgotten to map certain values or mapped them incorrectly.

See the following related topics:

> Evaluating the quality of imported data (sec. 5)
> Bulk deleting records (sec. 5.4)

Integration with an external data source

1. Create an integration job using the job named *Configurable supervisor synchronisation*.

> Creating an integration job (sec. 4.1)

2. Configure the integration job based on the dataset you are integrating.

> Configuring an integration job (sec. 4.3)

Note

If you have not previously imported records with this content type, we recommend that you start by integrating a small number of records so that you can manually check the accuracy of your data mapping and transformation processes.

3. Schedule a time for the job to run or run the job immediately.

> Scheduling an integration job (sec. 4.4)
> Running an integration job on an ad hoc basis (sec. 4.5)

4. After the job has run, use the logs to check whether it succeeded and which errors occurred. If no records were added to Pure, address the listed errors and try to run the integration job again.

> Viewing the status of an integration job (sec. 4.6)
> Using the job logs to evaluate import quality (sec. 5.1)

5. Check the status of the data you integrated and delete incorrect records.

   Depending on the configuration of the integration job, some records may be unable to be deleted in Pure. In this case, change the configuration settings and run the integration job again to update incorrect records.

> Evaluating the quality of imported data (sec. 5)
> Configuring an integration job (sec. 4.3)

6. If you deleted records, run the integration job again and check that both deleted and preserved records exist as expected. If the records have *Start date* and *End date* fields, check that these are correct.

7. Once you are satisfied that all records are in Pure and all fields are correctly configured, configure the integration job as you want it to run regularly, and set up your external data source to provide the source XML at the appropriate time.

> Configuring an integration job (sec. 4.3)
> Scheduling an integration job (sec. 4.4)

Tip

For general information about integrations, see Section 4. Integrating external data into Pure.

For detailed steps about how to select records and verify the correctness of your data when performing an integration, see Section 4.8. Integrating data sources with Pure - detailed steps.

6.5. Populating Publisher data

Tip

Before you begin, consider reading the introductory section of this manual. See Section 1. Introduction.

About

A record with the content type of Publisher represents a publisher.

Requirements

Publishers are generally populated at the same time as Journals. They may be created when importing Research output, or from online sources.

See the following related topics:

> Populating content in the correct order (sec. 1.2.1)

Data preparation

When preparing data for an import, consider the following tips:

- Ensure all of the mandatory fields are complete. You can see which fields are mandatory by adding a new record in the Pure interface and looking for fields with a symbol next to the field name.

See the following related topics:

> Preparing data sources (sec. 2)
> Pure's XML import format and the XSD (sec. 2.4)
> Mapping types and fields to the Pure data model (sec. 2.6)
> Mapping external values to classifications (sec. 2.7)

Data validation

When evaluating an import, pay attention to these areas where problems occur most commonly:

- Check that each field has a single, clearly-defined use, ideally reflected in the name of the field. There should be no fields that contain miscellaneous information from the data source.
- Check for fields that are consistently empty. Though it is not mandatory to populate all fields, it could be a warning that you have forgotten to map certain values or mapped them incorrectly.

See the following related topics:

> Evaluating the quality of imported data (sec. 5)
> Bulk deleting records (sec. 5.4)

Initial population

Before you (optionally) configure an integration job for the Publisher content type, you can use the Bulk Import Wizard to import a subset of the data in an ad hoc fashion to test the data mapping.

To use the Bulk Import Wizard to import a Publisher:

1. Click the Administrator tab in the top navigation.
2. Click the Bulk import tab in the navigation.
3. Click the *Publishers* tab in the navigation.

   The Import wizard page is opened.

   Tip

   You can access documentation about the import process on this page, including XML schema definitions and example XML files.

4. Click Open import wizard.

   The Bulk Import Wizard is opened and guides you through the data import process.

See the following related topics:

> Migrating legacy data into Pure (sec. 3)
> Viewing the status of an import from the Bulk Import Wizard (sec. 3.2)
> Bulk populating content in Pure - detailed steps (sec. 3.3)

Tip

For general information about populating records with the Bulk Import Wizard, see Migrating legacy data into Pure or Using the Bulk Import Wizard to import external data.

For detailed steps about how to select records and verify the correctness of your data when importing records, see Bulk populating content in Pure - detailed steps.

Integration with an external data source

1. Create an integration job using the job named *Configurable publisher synchronisation*.

   > Creating an integration job (sec. 4.1)

2. Configure the integration job based on the dataset you are integrating.

   > Configuring an integration job (sec. 4.3)

Note

If you have not previously imported records with this content type, we recommend that you start by integrating a small number of records so that you can manually check the accuracy of your data mapping and transformation processes.

3. Schedule a time for the job to run or run the job immediately.

   > Scheduling an integration job (sec. 4.4)
   > Running an integration job on an ad hoc basis (sec. 4.5)

4. After the job has run, use the logs to check whether it succeeded and which errors occurred. If no records were added to Pure, address the listed errors and try to run the integration job again.

   > Viewing the status of an integration job (sec. 4.6)
   > Using the job logs to evaluate import quality (sec. 5.1)

5. Check the status of the data you integrated and delete incorrect records.

   Depending on the configuration of the integration job, some records may be unable to be deleted in Pure. In this case, change the configuration settings and run the integration job again to update incorrect records.

   > Evaluating the quality of imported data (sec. 5)
   > Configuring an integration job (sec. 4.3)

6. If you deleted records, run the integration job again and check that both deleted and preserved records exist as expected. If the records have *Start date* and *End date* fields, check that these are correct.

7. Once you are satisfied that all records are in Pure and all fields are correctly configured, configure the integration job as you want it to run regularly, and set up your external data source to provide the source XML at the appropriate time.

   > Configuring an integration job (sec. 4.3)
   > Scheduling an integration job (sec. 4.4)

Tip

For general information about integrations, see Section 4. Integrating external data into Pure.

For detailed steps about how to select records and verify the correctness of your data when performing an integration, see Section 4.8. Integrating data sources with Pure - detailed steps.

6.6. Populating Journal data

> **Tip**
> Before you begin, consider reading the introductory section of this manual. See Section 1. Introduction.

About

A record with the content type Journal represents a scientific journal.

Requirements

Journals are generally populated at the same time as Publishers. They may be created when importing Research output, or from online sources.

See the following related topics:

> Populating content in the correct order (sec. 1.2.1)

Data preparation

When preparing data for an import, consider the following tips:

- Ensure all of the mandatory fields are complete. You can see which fields are mandatory by adding a new record in the Pure interface and looking for fields with a symbol next to the field name.

See the following related topics:

> Preparing data sources (sec. 2)
> Pure's XML import format and the XSD (sec. 2.4)
> Mapping types and fields to the Pure data model (sec. 2.6)
> Mapping external values to classifications (sec. 2.7)

Data validation

When evaluating an import, pay attention to these areas where problems occur most commonly:

- Check that each field has a single, clearly-defined use, ideally reflected in the name of the field. There should be no fields that contain miscellaneous information from the data source.
- Check for fields that are consistently empty. Though it is not mandatory to populate all fields, it could be a warning that you have forgotten to map certain values or mapped them incorrectly.

See the following related topics:

> Evaluating the quality of imported data (sec. 5)
> Bulk deleting records (sec. 5.4)

Initial population

Before you (optionally) configure an integration job for the Journal content type, you can use the Bulk Import Wizard to import a subset of the data in an ad hoc fashion to test the data mapping.

To use the Bulk Import Wizard to import a Journal:

1. Click the Administrator tab in the top navigation.
2. Click the Bulk import tab in the navigation.
3. Click the *Journals* tab in the navigation.

   The Import wizard page is opened.

   > **Tip**
   > You can access documentation about the import process on this page, including XML schema definitions and example XML files.

4. Click Open import wizard.

   The Bulk Import Wizard is opened and guides you through the data import process.

See the following related topics:

> Migrating legacy data into Pure (sec. 3)
> Viewing the status of an import from the Bulk Import Wizard (sec. 3.2)
> Bulk populating content in Pure - detailed steps (sec. 3.3)

Tip

For general information about populating records with the Bulk Import Wizard, see Migrating legacy data into Pure or Using the Bulk Import Wizard to import external data.

For detailed steps about how to select records and verify the correctness of your data when importing records, see Bulk populating content in Pure - detailed steps.

Integration with an external data source

1. Create an integration job using the job named *Configurable journal synchronisation*.

   > Creating an integration job (sec. 4.1)

2. Configure the integration job based on the dataset you are integrating.

   > Configuring an integration job (sec. 4.3)

Note

If you have not previously imported records with this content type, we recommend that you start by integrating a small number of records so that you can manually check the accuracy of your data mapping and transformation processes.

3. Schedule a time for the job to run or run the job immediately.

   > Scheduling an integration job (sec. 4.4)
   > Running an integration job on an ad hoc basis (sec. 4.5)

4. After the job has run, use the logs to check whether it succeeded and which errors occurred. If no records were added to Pure, address the listed errors and try to run the integration job again.

   > Viewing the status of an integration job (sec. 4.6)
   > Using the job logs to evaluate import quality (sec. 5.1)

5. Check the status of the data you integrated and delete incorrect records.

   Depending on the configuration of the integration job, some records may be unable to be deleted in Pure. In this case, change the configuration settings and run the integration job again to update incorrect records.

   > Evaluating the quality of imported data (sec. 5)
   > Configuring an integration job (sec. 4.3)

6. If you deleted records, run the integration job again and check that both deleted and preserved records exist as expected. If the records have *Start date* and *End date* fields, check that these are correct.

7. Once you are satisfied that all records are in Pure and all fields are correctly configured, configure the integration job as you want it to run regularly, and set up your external data source to provide the source XML at the appropriate time.

   > Configuring an integration job (sec. 4.3)
   > Scheduling an integration job (sec. 4.4)

Tip

For general information about integrations, see Section 4. Integrating external data into Pure.

For detailed steps about how to select records and verify the correctness of your data when performing an integration, see Section 4.8. Integrating data sources with Pure - detailed steps.

6.7. Populating Event data

Tip

Before you begin, consider reading the introductory section of this manual. See Section 1. Introduction.

About

An Event record in Pure describes specific information about any kind of event which academic or administrative staff may attend or organize.

Note

Academic staff are not linked directly to events, rather they are linked to an Activity record which is linked to this event. This is due to the different activities through which attendees can participate in events, such as giving a keynote or presenting a poster.

Requirements

Event records can be populated as soon as the content types Organization and Person have been populated.

See the following related topics:

> Populating content in the correct order (sec. 1.2.1)

Data preparation

When preparing data for an import, consider the following tips:

- Ensure all of the mandatory fields are complete. You can see which fields are mandatory by adding a new record in the Pure interface and looking for fields with a symbol next to the field name.

See the following related topics:

> Preparing data sources (sec. 2)
> Pure's XML import format and the XSD (sec. 2.4)
> Mapping types and fields to the Pure data model (sec. 2.6)
> Mapping external values to classifications (sec. 2.7)

Data validation

When evaluating an import, pay attention to these areas where problems occur most commonly:

See the following related topics:

> Evaluating the quality of imported data (sec. 5)
> Bulk deleting records (sec. 5.4)

- Check that each field has a single, clearly-defined use, ideally reflected in the name of the field. There should be no fields that contain miscellaneous information from the data source.
- Check for fields that are consistently empty. Though it is not mandatory to populate all fields, it could be a warning that you have forgotten to map certain values or mapped them incorrectly.

Initial population

Before you (optionally) configure an integration job for the Event content type, you can use the Bulk Import Wizard to import a subset of the data in an ad hoc fashion to test the data mapping.

To use the Bulk Import Wizard to import a Event:

See the following related topics:

> Migrating legacy data into Pure (sec. 3)
> Viewing the status of an import from the Bulk Import Wizard (sec. 3.2)
> Bulk populating content in Pure - detailed steps (sec. 3.3)

1. Click the Administrator tab in the top navigation.
2. Click the Bulk import tab in the navigation.

3. Click the *Event* tab in the navigation.

   The Import wizard page is opened.

   Tip

   You can access documentation about the import process on this page, including XML schema definitions and example XML files.

4. Click Open import wizard.

   The Bulk Import Wizard is opened and guides you through the data import process.

   Tip

   For general information about populating records with the Bulk Import Wizard, see Migrating legacy data into Pure or Using the Bulk Import Wizard to import external data.

   For detailed steps about how to select records and verify the correctness of your data when importing records, see Bulk populating content in Pure - detailed steps.

Integration with an external data source

1. Create an integration job using the job named *Configurable event synchronisation*.

   > Creating an integration job (sec. 4.1)

2. Configure the integration job based on the dataset you are integrating.

   > Configuring an integration job (sec. 4.3)

   Note

   If you have not previously imported records with this content type, we recommend that you start by integrating a small number of records so that you can manually check the accuracy of your data mapping and transformation processes.

3. Schedule a time for the job to run or run the job immediately.

   > Scheduling an integration job (sec. 4.4)
   > Running an integration job on an ad hoc basis (sec. 4.5)

4. After the job has run, use the logs to check whether it succeeded and which errors occurred. If no records were added to Pure, address the listed errors and try to run the integration job again.

   > Viewing the status of an integration job (sec. 4.6)
   > Using the job logs to evaluate import quality (sec. 5.1)

5. Check the status of the data you integrated and delete incorrect records.

   Depending on the configuration of the integration job, some records may be unable to be deleted in Pure. In this case, change the configuration settings and run the integration job again to update incorrect records.

   > Evaluating the quality of imported data (sec. 5)
   > Configuring an integration job (sec. 4.3)

6. If you deleted records, run the integration job again and check that both deleted and preserved records exist as expected. If the records have *Start date* and *End date* fields, check that these are correct.

7. Once you are satisfied that all records are in Pure and all fields are correctly configured, configure the integration job as you want it to run regularly, and set up your external data source to provide the source XML at the appropriate time.

   > Configuring an integration job (sec. 4.3)
   > Scheduling an integration job (sec. 4.4)

Tip

For general information about integrations, see Section 4. Integrating external data into Pure.

For detailed steps about how to select records and verify the correctness of your data when performing an integration, see Section 4.8. Integrating data sources with Pure - detailed steps.

6.8. Populating External organization data

Tip

Before you begin, consider reading the introductory section of this manual. See Section 1. Introduction.

About

An External organization record in Pure generally represents a workplace linked to an External person record. They may also represent funding councils or other stakeholders - in short, any organization that is not a part of your institution.

Requirements

External organizations are generally created while importing Master data or Research output records , though they can also be created via a dedicated import.

See the following related topics:

> Populating content in the correct order (sec. 1.2.1)

Data preparation

When preparing data for an import, consider the following tips:

- Ensure all of the mandatory fields are complete. You can see which fields are mandatory by adding a new record in the Pure interface and looking for fields with a ✱ symbol next to the field name.

See the following related topics:

> Preparing data sources (sec. 2)
> Pure's XML import format and the XSD (sec. 2.4)
> Mapping types and fields to the Pure data model (sec. 2.6)
> Mapping external values to classifications (sec. 2.7)

Data validation

When evaluating an import, pay attention to these areas where problems occur most commonly:

See the following related topics:

> Evaluating the quality of imported data (sec. 5)
> Bulk deleting records (sec. 5.4)

- Check that each field has a single, clearly-defined use, ideally reflected in the name of the field. There should be no fields that contain miscellaneous information from the data source.
- Check for fields that are consistently empty. Though it is not mandatory to populate all fields, it could be a warning that you have forgotten to map certain values or mapped them incorrectly.

Initial population

Before you (optionally) configure an integration job for the External organization content type, you can use the Bulk Import Wizard to import a subset of the data in an ad hoc fashion to test the data mapping.

To use the Bulk Import Wizard to import a External organization:

See the following related topics:

> Migrating legacy data into Pure (sec. 3)
> Viewing the status of an import from the Bulk Import Wizard (sec. 3.2)
> Bulk populating content in Pure - detailed steps (sec. 3.3)

1. Click the Administrator tab in the top navigation.
2. Click the Bulk import tab in the navigation.
3. Click the *External organisation* tab in the navigation.

   The Import wizard page is opened.

Tip

You can access documentation about the import process on this page, including XML schema definitions and example XML files.

4. Click Open import wizard.

   The Bulk Import Wizard is opened and guides you through the data import process.

> **Tip**
>
> For general information about populating records with the Bulk Import Wizard, see Migrating legacy data into Pure or Using the Bulk Import Wizard to import external data.
>
> For detailed steps about how to select records and verify the correctness of your data when importing records, see Bulk populating content in Pure - detailed steps.

Integration with an external data source

1. Create an integration job using the job named *Configurable external organisation synchronisation*.

   > Creating an integration job (sec. 4.1)

2. Configure the integration job based on the dataset you are integrating.

   > Configuring an integration job (sec. 4.3)

> **Note**
>
> If you have not previously imported records with this content type, we recommend that you start by integrating a small number of records so that you can manually check the accuracy of your data mapping and transformation processes.

3. Schedule a time for the job to run or run the job immediately.

   > Scheduling an integration job (sec. 4.4)
   > Running an integration job on an ad hoc basis (sec. 4.5)

4. After the job has run, use the logs to check whether it succeeded and which errors occurred. If no records were added to Pure, address the listed errors and try to run the integration job again.

   > Viewing the status of an integration job (sec. 4.6)
   > Using the job logs to evaluate import quality (sec. 5.1)

5. Check the status of the data you integrated and delete incorrect records.

   Depending on the configuration of the integration job, some records may be unable to be deleted in Pure. In this case, change the configuration settings and run the integration job again to update incorrect records.

   > Evaluating the quality of imported data (sec. 5)
   > Configuring an integration job (sec. 4.3)

6. If you deleted records, run the integration job again and check that both deleted and preserved records exist as expected. If the records have *Start date* and *End date* fields, check that these are correct.

7. Once you are satisfied that all records are in Pure and all fields are correctly configured, configure the integration job as you want it to run regularly, and set up your external data source to provide the source XML at the appropriate time.

   > Configuring an integration job (sec. 4.3)
   > Scheduling an integration job (sec. 4.4)

> **Tip**
>
> For general information about integrations, see Section 4. Integrating external data into Pure.
>
> For detailed steps about how to select records and verify the correctness of your data when performing an integration, see Section 4.8. Integrating data sources with Pure - detailed steps.

6.9. Populating External person data

> **Tip**
> Before you begin, consider reading the introductory section of this manual. See Section 1. Introduction.

About

An External person in Pure is a record with information to describe people outside your institution who are linked to records from within your institution, such as an external collaborator on a research paper.

> **Note**
> External persons are created automatically when a relationship is added that cannot be matched to a profile in Pure.

Requirements

External persons are generally created while importing Research output records, though they can also be created via a dedicated import.

See the following related topics:

> Populating content in the correct order (sec. 1.2.1)

Data preparation

When preparing data for an import, consider the following tips:

- Ensure all of the mandatory fields are complete. You can see which fields are mandatory by adding a new record in the Pure interface and looking for fields with a * symbol next to the field name.

See the following related topics:

> Preparing data sources (sec. 2)
> Pure's XML import format and the XSD (sec. 2.4)
> Mapping types and fields to the Pure data model (sec. 2.6)
> Mapping external values to classifications (sec. 2.7)

Data validation

When evaluating an import, pay attention to these areas where problems occur most commonly:

- Check that each field has a single, clearly-defined use, ideally reflected in the name of the field. There should be no fields that contain miscellaneous information from the data source.
- Check for fields that are consistently empty. Though it is not mandatory to populate all fields, it could be a warning that you have forgotten to map certain values or mapped them incorrectly.

See the following related topics:

> Evaluating the quality of imported data (sec. 5)
> Bulk deleting records (sec. 5.4)

Initial population

Before you (optionally) configure an integration job for the External person content type, you can use the Bulk Import Wizard to import a subset of the data in an ad hoc fashion to test the data mapping.

To use the Bulk Import Wizard to import a External person:

1. Click the Administrator tab in the top navigation.
2. Click the Bulk import tab in the navigation.

See the following related topics:

> Migrating legacy data into Pure (sec. 3)
> Viewing the status of an import from the Bulk Import Wizard (sec. 3.2)
> Bulk populating content in Pure - detailed steps (sec. 3.3)

3. Click the *External person* tab in the navigation.

   The Import wizard page is opened.

   Tip

   You can access documentation about the import process on this page, including XML schema definitions and example XML files.

4. Click Open import wizard.

   The Bulk Import Wizard is opened and guides you through the data import process.

   Tip

   For general information about populating records with the Bulk Import Wizard, see Migrating legacy data into Pure or Using the Bulk Import Wizard to import external data.

   For detailed steps about how to select records and verify the correctness of your data when importing records, see Bulk populating content in Pure - detailed steps.

Integration with an external data source

1. Create an integration job using the job named *Configurable external person synchronisation*.

   > Creating an integration job (sec. 4.1)

2. Configure the integration job based on the dataset you are integrating.

   > Configuring an integration job (sec. 4.3)

   Note

   If you have not previously imported records with this content type, we recommend that you start by integrating a small number of records so that you can manually check the accuracy of your data mapping and transformation processes.

3. Schedule a time for the job to run or run the job immediately.

   > Scheduling an integration job (sec. 4.4)
   > Running an integration job on an ad hoc basis (sec. 4.5)

4. After the job has run, use the logs to check whether it succeeded and which errors occurred. If no records were added to Pure, address the listed errors and try to run the integration job again.

   > Viewing the status of an integration job (sec. 4.6)
   > Using the job logs to evaluate import quality (sec. 5.1)

5. Check the status of the data you integrated and delete incorrect records.

   Depending on the configuration of the integration job, some records may be unable to be deleted in Pure. In this case, change the configuration settings and run the integration job again to update incorrect records.

   > Evaluating the quality of imported data (sec. 5)
   > Configuring an integration job (sec. 4.3)

6. If you deleted records, run the integration job again and check that both deleted and preserved records exist as expected. If the records have *Start date* and *End date* fields, check that these are correct.

7. Once you are satisfied that all records are in Pure and all fields are correctly configured, configure the integration job as you want it to run regularly, and set up your external data source to provide the source XML at the appropriate time.

   > Configuring an integration job (sec. 4.3)
   > Scheduling an integration job (sec. 4.4)

Tip

For general information about integrations, see Section 4. Integrating external data into Pure.

For detailed steps about how to select records and verify the correctness of your data when performing an integration, see Section 4.8. Integrating data sources with Pure - detailed steps.

6.10. Populating Research output

Tip

Before you begin, consider reading the introductory section of this manual. See Section 1. Introduction.

About

The content type Research output describes a wide range of academic publications and productions, ranging from classical formats such as peer-reviewed articles or books to software, patents or performances.

In the context of populating Pure, the source of Research output data is data maintained in a legacy system at your institution detailing publications and research contributions made by past and present staff.

As the purpose of moving your data to Pure is to maintain complete, comprehensive and accurate records about your research output, it is important that legacy research data is imported in a way that does not introduce errors.

If your legacy data is noisy, you can make use of various data-cleaning tools to clean your research output data once it has been imported into Pure, such as de-duplication and bulk editing after applying filters. Keep in mind that if you follow an iterative import strategy this noise will frequently be re-imported, so any data-cleaning within Pure should only start once you have finished migrating your Research output in full. You may consider investing in data cleaning before importing it into Pure, though this is not recommended.

Requirements

Research output can only be imported once Organization and Person records have been imported and are complete and correct. If your research output includes publications by past staff members, it is important that these staff members have also been imported.

See the following related topics:

> Populating content in the correct order (sec. 1.2.1)

Recommendation

It is possible to begin importing Research output in parallel with Organizational units and Persons, however once you begin importing Research output no changes can be made to existing ID values of these content types, or to the hierarchical structure of Organizational units. Therefore it is recommended to wait until at least all mandatory fields for Organizational units and Persons are complete.

If your institution is using keyword groups to categorize Research output, you need to configure these before importing Research output records.

Note

Keyword groups allow for additional fields that can be used to group or filter records in some custom way, such as by subject area or theme. They can be accessed on the Keywords tab of the Administrator tab. Keywords should *not* be used as a way to directly apply complex values from your institution's legacy system to Pure. Instead, you should concentrate on appropriately mapping these values to the Pure data model.

Data preparation

When preparing data for an import, consider the following tips:

See the following related topics:

> Preparing data sources (sec. 2)
> Pure's XML import format and the XSD (sec. 2.4)
> Mapping types and fields to the Pure data model (sec. 2.6)
> Mapping external values to classifications (sec. 2.7)

- If Scopus IDs are available for the research data, include these in the data feed. Later you will be able to import citations based on these IDs.
- Inspect your dataset for uncommon types of research output. If some types have only a very small number of records and creating and testing a mapping for these types would be a time-consuming task, you may consider entering these records manually in the Pure interface.

- Ensure all of the mandatory fields are complete. You can see which fields are mandatory by adding a new record in the Pure interface and looking for fields with a * symbol next to the field name.

Data validation

When evaluating an import, pay attention to these areas where problems occur most commonly:

See the following related topics:

> Evaluating the quality of imported data (sec. 5)
> Bulk deleting records (sec. 5.4)

- Begin by checking *Contribution to journal* records, as these are likely the most numerous type of record.
- Check that unique ID fields such as DOI, ISBN and ISSN have been populated and have the correct format.
- Check that the *Peer-reviewed* value is set correctly.
- Check that there is a managing organizational unit present in the *Research output managed by* section of the editor window. If not, the section is not visible.
- Test on a set of publications, with at least one publication for each research output template. For example, check publications for your institution's top researchers with a broad range of media, such as books, journal articles, conference submissions etc.
    - Is the author order correct?
    - Do authors match to the correct person entities? Are internal persons misidentified as external persons?
    - Are similar-sounding terms confused? For example, the role of **Editor** should be distinct from being the author of an **Editorial book**.
    - Is the field mapping correct? For example are titles and subtitles correctly assigned?
    - Are full text files available where expected? Are the mime type and file names correct?
- Test on several conference publications:
    - Were the corresponding events created?
- Test on several journals and books.
    - Were the appropriate journal and publisher entities were created?
    - Were journals and publishers appropriately named?
    - Were duplicate journals and publishers properly linked to existing entities?
- Create a report and check that this matches expectations for:
    - the number of Research output items per Organization.
    - the number of Research output items per year.
- Click on the *Research output* tab of the *Editor tab* and view the options that appear under the *Research output* tab in the side banner that list duplicate records. Consider why duplicates may be appearing — is there a problem with the import process? If there duplicates in your data source, consider whether you want to perform de-duplication within Pure or whether you can easily remove certain duplicate records before importing the data.
- Check that each field has a single, clearly-defined use, ideally reflected in the name of the field. There should be no fields that contain miscellaneous information from the data source.
- Check for fields that are consistently empty. Though it is not mandatory to populate all fields, it could be a warning that you have forgotten to map certain values or mapped them incorrectly.

Initial population

Before you (optionally) configure an integration job for the Research output content type, you can use the Bulk Import Wizard to import a subset of the data in an ad hoc fashion to test the data mapping.

See the following related topics:

> Migrating legacy data into Pure (sec. 3)
> Viewing the status of an import from the Bulk Import Wizard (sec. 3.2)
> Bulk populating content in Pure - detailed steps (sec. 3.3)

To use the Bulk Import Wizard to import a Research output:

1. Click the Administrator tab in the top navigation.
2. Click the Bulk import tab in the navigation.

3. Click the *Research Output* tab in the navigation.

   The Import wizard page is opened.

   Tip
   You can access documentation about the import process on this page, including XML schema definitions and example XML files.

4. Click Open import wizard.

   The Bulk Import Wizard is opened and guides you through the data import process.

   Tip
   For general information about populating records with the Bulk Import Wizard, see Migrating legacy data into Pure or Using the Bulk Import Wizard to import external data.

   For detailed steps about how to select records and verify the correctness of your data when importing records, see Bulk populating content in Pure - detailed steps.

Integration with an external data source

1. Create an integration job using the job named *!NOCRONJOB?*.

   > Creating an integration job (sec. 4.1)

2. Configure the integration job based on the dataset you are integrating.

   > Configuring an integration job (sec. 4.3)

   Note
   If you have not previously imported records with this content type, we recommend that you start by integrating a small number of records so that you can manually check the accuracy of your data mapping and transformation processes.

3. Schedule a time for the job to run or run the job immediately.

   > Scheduling an integration job (sec. 4.4)
   > Running an integration job on an ad hoc basis (sec. 4.5)

4. After the job has run, use the logs to check whether it succeeded and which errors occurred. If no records were added to Pure, address the listed errors and try to run the integration job again.

   > Viewing the status of an integration job (sec. 4.6)
   > Using the job logs to evaluate import quality (sec. 5.1)

5. Check the status of the data you integrated and delete incorrect records.

   Depending on the configuration of the integration job, some records may be unable to be deleted in Pure. In this case, change the configuration settings and run the integration job again to update incorrect records.

   > Evaluating the quality of imported data (sec. 5)
   > Configuring an integration job (sec. 4.3)

6. If you deleted records, run the integration job again and check that both deleted and preserved records exist as expected. If the records have *Start date* and *End date* fields, check that these are correct.

7. Once you are satisfied that all records are in Pure and all fields are correctly configured, configure the integration job as you want it to run regularly, and set up your external data source to provide the source XML at the appropriate time.

   > Configuring an integration job (sec. 4.3)
   > Scheduling an integration job (sec. 4.4)

Tip

For general information about integrations, see Section 4. Integrating external data into Pure.

For detailed steps about how to select records and verify the correctness of your data when performing an integration, see Section 4.8. Integrating data sources with Pure - detailed steps.

6.11. Populating Activity data

Tip

Before you begin, consider reading the introductory section of this manual. See Section 1. Introduction.

About

A record with the content type Activity can represent a range of academic activities, such as conference participation, external academic engagement, editorial work or peer reviewing, public engagement etc. These activities are often collected to exhibit with researcher profiles online, or for reporting purposes.

Requirements

Activity records can be populated as soon as the content types Organization and Person have been populated.

See the following related topics:

> Populating content in the correct order (sec. 1.2.1)

Data preparation

When preparing data for an import, consider the following tips:

- Ensure all of the mandatory fields are complete. You can see which fields are mandatory by adding a new record in the Pure interface and looking for fields with a symbol next to the field name.

See the following related topics:

> Preparing data sources (sec. 2)
> Pure's XML import format and the XSD (sec. 2.4)
> Mapping types and fields to the Pure data model (sec. 2.6)
> Mapping external values to classifications (sec. 2.7)

Data validation

When evaluating an import, pay attention to these areas where problems occur most commonly:

- Check that each field has a single, clearly-defined use, ideally reflected in the name of the field. There should be no fields that contain miscellaneous information from the data source.
- Check for fields that are consistently empty. Though it is not mandatory to populate all fields, it could be a warning that you have forgotten to map certain values or mapped them incorrectly.

See the following related topics:

> Evaluating the quality of imported data (sec. 5)
> Bulk deleting records (sec. 5.4)

Initial population

Before you (optionally) configure an integration job for the Activity content type, you can use the Bulk Import Wizard to import a subset of the data in an ad hoc fashion to test the data mapping.

To use the Bulk Import Wizard to import a Activity:

1. Click the Administrator tab in the top navigation.
2. Click the Bulk import tab in the navigation.
3. Click the *Activities* tab in the navigation.

   The Import wizard page is opened.

See the following related topics:

> Migrating legacy data into Pure (sec. 3)
> Viewing the status of an import from the Bulk Import Wizard (sec. 3.2)
> Bulk populating content in Pure - detailed steps (sec. 3.3)

Tip

You can access documentation about the import process on this page, including XML schema definitions and example XML files.

4. Click Open import wizard.

   The Bulk Import Wizard is opened and guides you through the data import process.

> **Tip**
>
> For general information about populating records with the Bulk Import Wizard, see Migrating legacy data into Pure or Using the Bulk Import Wizard to import external data.
>
> For detailed steps about how to select records and verify the correctness of your data when importing records, see Bulk populating content in Pure - detailed steps.

Integration with an external data source

1. Create an integration job using the job named *Configurable activity synchronisation*.

   > Creating an integration job (sec. 4.1)

2. Configure the integration job based on the dataset you are integrating.

   > Configuring an integration job (sec. 4.3)

> **Note**
>
> If you have not previously imported records with this content type, we recommend that you start by integrating a small number of records so that you can manually check the accuracy of your data mapping and transformation processes.

3. Schedule a time for the job to run or run the job immediately.

   > Scheduling an integration job (sec. 4.4)
   > Running an integration job on an ad hoc basis (sec. 4.5)

4. After the job has run, use the logs to check whether it succeeded and which errors occurred. If no records were added to Pure, address the listed errors and try to run the integration job again.

   > Viewing the status of an integration job (sec. 4.6)
   > Using the job logs to evaluate import quality (sec. 5.1)

5. Check the status of the data you integrated and delete incorrect records.

   Depending on the configuration of the integration job, some records may be unable to be deleted in Pure. In this case, change the configuration settings and run the integration job again to update incorrect records.

   > Evaluating the quality of imported data (sec. 5)
   > Configuring an integration job (sec. 4.3)

6. If you deleted records, run the integration job again and check that both deleted and preserved records exist as expected. If the records have *Start date* and *End date* fields, check that these are correct.

7. Once you are satisfied that all records are in Pure and all fields are correctly configured, configure the integration job as you want it to run regularly, and set up your external data source to provide the source XML at the appropriate time.

   > Configuring an integration job (sec. 4.3)
   > Scheduling an integration job (sec. 4.4)

> **Tip**
>
> For general information about integrations, see Section 4. Integrating external data into Pure.
>
> For detailed steps about how to select records and verify the correctness of your data when performing an integration, see Section 4.8. Integrating data sources with Pure - detailed steps.

6.12. Populating Funding Opportunity data

Tip

Before you begin, consider reading the introductory section of this manual. See Section 1. Introduction.

About

A Funding opportunity in Pure is a record with detailed information about research funding being available from a funder. This record contains information about the funder, the funding program, eligibility criteria, deadlines, award ceilings etc.

Requirements

Funding opportunity records can be populated as soon as the content types Organization and Person have been populated.

See the following related topics:

> Populating content in the correct order (sec. 1.2.1)

Data preparation

When preparing data for an import, consider the following tips:

- Ensure all of the mandatory fields are complete. You can see which fields are mandatory by adding a new record in the Pure interface and looking for fields with a * symbol next to the field name.

See the following related topics:

> Preparing data sources (sec. 2)
> Pure's XML import format and the XSD (sec. 2.4)
> Mapping types and fields to the Pure data model (sec. 2.6)
> Mapping external values to classifications (sec. 2.7)

Data validation

When evaluating an import, pay attention to these areas where problems occur most commonly:

- Check that each field has a single, clearly-defined use, ideally reflected in the name of the field. There should be no fields that contain miscellaneous information from the data source.
- Check for fields that are consistently empty. Though it is not mandatory to populate all fields, it could be a warning that you have forgotten to map certain values or mapped them incorrectly.

See the following related topics:

> Evaluating the quality of imported data (sec. 5)
> Bulk deleting records (sec. 5.4)

Integration with an external data source

1. Create an integration job using the job named *Funding synchronisation*.

> Creating an integration job (sec. 4.1)

2. Configure the integration job based on the dataset you are integrating.

> Configuring an integration job (sec. 4.3)

Note

If you have not previously imported records with this content type, we recommend that you start by integrating a small number of records so that you can manually check the accuracy of your data mapping and transformation processes.

3. Schedule a time for the job to run or run the job immediately.

> Scheduling an integration job (sec. 4.4)
> Running an integration job on an ad hoc basis (sec. 4.5)

4. After the job has run, use the logs to check whether it succeeded and which errors occurred. If no records were added to Pure, address the listed errors and try to run the integration job again.

> Viewing the status of an integration job (sec. 4.6)
> Using the job logs to evaluate import quality (sec. 5.1)

5. Check the status of the data you integrated and delete incorrect records.

   Depending on the configuration of the integration job, some records may be unable to be deleted in Pure. In this case, change the configuration settings and run the integration job again to update incorrect records.

> Evaluating the quality of imported data (sec. 5)
> Configuring an integration job (sec. 4.3)

6. If you deleted records, run the integration job again and check that both deleted and preserved records exist as expected. If the records have *Start date* and *End date* fields, check that these are correct.

7. Once you are satisfied that all records are in Pure and all fields are correctly configured, configure the integration job as you want it to run regularly, and set up your external data source to provide the source XML at the appropriate time.

> Configuring an integration job (sec. 4.3)
> Scheduling an integration job (sec. 4.4)

Tip

For general information about integrations, see Section 4. Integrating external data into Pure.

For detailed steps about how to select records and verify the correctness of your data when performing an integration, see Section 4.8. Integrating data sources with Pure - detailed steps.

6.13. Populating Application data

Tip

Before you begin, consider reading the introductory section of this manual. See Section 1. Introduction.

About

An Application in Pure is a record with information about researchers' applications for research funding from a specific funder. The record holds basic information about the application, applicants and collaborators, as well as the amount and the type of funding applied for. It can also hold information about the budget for spending the funding.

Applications can be managed solely in Pure or are integrated with an external data source.

Requirements

Application records can be populated as soon as the content types Organization and Person have been populated.

See the following related topics:

> Populating content in the correct order (sec. 1.2.1)

Data preparation

When preparing data for an import, consider the following tips:

- Ensure all of the mandatory fields are complete. You can see which fields are mandatory by adding a new record in the Pure interface and looking for fields with a ✱ symbol next to the field name.

See the following related topics:

> Preparing data sources (sec. 2)
> Pure's XML import format and the XSD (sec. 2.4)
> Mapping types and fields to the Pure data model (sec. 2.6)
> Mapping external values to classifications (sec. 2.7)

Data validation

When evaluating an import, pay attention to these areas where problems occur most commonly:

See the following related topics:

> Evaluating the quality of imported data (sec. 5)
> Bulk deleting records (sec. 5.4)

- Check that each field has a single, clearly-defined use, ideally reflected in the name of the field. There should be no fields that contain miscellaneous information from the data source.
- Check for fields that are consistently empty. Though it is not mandatory to populate all fields, it could be a warning that you have forgotten to map certain values or mapped them incorrectly.

Integration with an external data source

1. Create an integration job using the job named *Application synchronisation*.

> Creating an integration job (sec. 4.1)

2. Configure the integration job based on the dataset you are integrating.

> Configuring an integration job (sec. 4.3)

Note

If you have not previously imported records with this content type, we recommend that you start by integrating a small number of records so that you can manually check the accuracy of your data mapping and transformation processes.

3. Schedule a time for the job to run or run the job immediately.

> Scheduling an integration job (sec. 4.4)
> Running an integration job on an ad hoc basis (sec. 4.5)

4. After the job has run, use the logs to check whether it succeeded and which errors occurred. If no records were added to Pure, address the listed errors and try to run the integration job again.

> Viewing the status of an integration job (sec. 4.6)
> Using the job logs to evaluate import quality (sec. 5.1)

5. Check the status of the data you integrated and delete incorrect records.

   Depending on the configuration of the integration job, some records may be unable to be deleted in Pure. In this case, change the configuration settings and run the integration job again to update incorrect records.

> Evaluating the quality of imported data (sec. 5)
> Configuring an integration job (sec. 4.3)

6. If you deleted records, run the integration job again and check that both deleted and preserved records exist as expected. If the records have *Start date* and *End date* fields, check that these are correct.

7. Once you are satisfied that all records are in Pure and all fields are correctly configured, configure the integration job as you want it to run regularly, and set up your external data source to provide the source XML at the appropriate time.

> Configuring an integration job (sec. 4.3)
> Scheduling an integration job (sec. 4.4)

Tip

For general information about integrations, see Section 4. Integrating external data into Pure.

For detailed steps about how to select records and verify the correctness of your data when performing an integration, see Section 4.8. Integrating data sources with Pure - detailed steps.

6.14. Populating Ethical review data

> **Tip**
> Before you begin, consider reading the introductory section of this manual. See Section 1. Introduction.

About

An Ethical review in Pure is a record that describes a formal review undertaken to evaluate whether a research proposal adheres to ethical standards. The record contains information about people, documents and Applications associated with the review, and maintains information about whether the review is pending, or has been approved or declined. Multiple Applications can be associated with one Ethical review.

Requirements

Ethical reviews can be populated after you have populated Applications. See Section 6.13. Populating Application data.

See the following related topics:

> Populating content in the correct order (sec. 1.2.1)

Data preparation

When preparing data for an import, consider the following tips:

- Ensure the *Description* is plain-text and does not include HTML formatting.
- Ensure all of the mandatory fields are complete. You can see which fields are mandatory by adding a new record in the Pure interface and looking for fields with a symbol next to the field name.

See the following related topics:

> Preparing data sources (sec. 2)
> Pure's XML import format and the XSD (sec. 2.4)
> Mapping types and fields to the Pure data model (sec. 2.6)
> Mapping external values to classifications (sec. 2.7)

Data validation

When evaluating an import, pay attention to these areas where problems occur most commonly:

See the following related topics:

> Evaluating the quality of imported data (sec. 5)
> Bulk deleting records (sec. 5.4)

- Check that each field has a single, clearly-defined use, ideally reflected in the name of the field. There should be no fields that contain miscellaneous information from the data source.
- Check for fields that are consistently empty. Though it is not mandatory to populate all fields, it could be a warning that you have forgotten to map certain values or mapped them incorrectly.

Initial population

Before you (optionally) configure an integration job for the Ethical review content type, you can use the Bulk Import Wizard to import a subset of the data in an ad hoc fashion to test the data mapping.

To use the Bulk Import Wizard to import a Ethical review:

1. Click the Administrator tab in the top navigation.
2. Click the Bulk import tab in the navigation.

See the following related topics:

> Migrating legacy data into Pure (sec. 3)
> Viewing the status of an import from the Bulk Import Wizard (sec. 3.2)
> Bulk populating content in Pure - detailed steps (sec. 3.3)

3. Click the *Ethical Reviews* tab in the navigation.

   The Import wizard page is opened.

   Tip

   You can access documentation about the import process on this page, including XML schema definitions and example XML files.

4. Click Open import wizard.

   The Bulk Import Wizard is opened and guides you through the data import process.

   Tip

   For general information about populating records with the Bulk Import Wizard, see Migrating legacy data into Pure or Using the Bulk Import Wizard to import external data.

   For detailed steps about how to select records and verify the correctness of your data when importing records, see Bulk populating content in Pure - detailed steps.

Integration with an external data source

1. Create an integration job using the job named *Configurable ethical review synchronisation*.

   > Creating an integration job (sec. 4.1)

2. Configure the integration job based on the dataset you are integrating.

   > Configuring an integration job (sec. 4.3)

   Note

   If you have not previously imported records with this content type, we recommend that you start by integrating a small number of records so that you can manually check the accuracy of your data mapping and transformation processes.

3. Schedule a time for the job to run or run the job immediately.

   > Scheduling an integration job (sec. 4.4)
   > Running an integration job on an ad hoc basis (sec. 4.5)

4. After the job has run, use the logs to check whether it succeeded and which errors occurred. If no records were added to Pure, address the listed errors and try to run the integration job again.

   > Viewing the status of an integration job (sec. 4.6)
   > Using the job logs to evaluate import quality (sec. 5.1)

5. Check the status of the data you integrated and delete incorrect records.

   Depending on the configuration of the integration job, some records may be unable to be deleted in Pure. In this case, change the configuration settings and run the integration job again to update incorrect records.

   > Evaluating the quality of imported data (sec. 5)
   > Configuring an integration job (sec. 4.3)

6. If you deleted records, run the integration job again and check that both deleted and preserved records exist as expected. If the records have *Start date* and *End date* fields, check that these are correct.

7. Once you are satisfied that all records are in Pure and all fields are correctly configured, configure the integration job as you want it to run regularly, and set up your external data source to provide the source XML at the appropriate time.

   > Configuring an integration job (sec. 4.3)
   > Scheduling an integration job (sec. 4.4)

Tip

For general information about integrations, see Section 4. Integrating external data into Pure.

For detailed steps about how to select records and verify the correctness of your data when performing an integration, see Section 4.8. Integrating data sources with Pure - detailed steps.

6.15. Populating Award data

Tip

Before you begin, consider reading the introductory section of this manual. See Section 1. Introduction.

About

An Award in Pure is a record with information about research funding granted by a funder. Many fields are shared with the Application content type, though award records can hold unique information about the amount awarded and non-financial contributions.

Requirements

Award records can be populated as soon as the content types Organization and Person have been populated.

See the following related topics:

> Populating content in the correct order (sec. 1.2.1)

Note

You may consider importing certain Awards as Projects if you want them to be visible on the Pure Portal.

Data preparation

When preparing data for an import, consider the following tips:

- Ensure all of the mandatory fields are complete. You can see which fields are mandatory by adding a new record in the Pure interface and looking for fields with a symbol next to the field name.

See the following related topics:

> Preparing data sources (sec. 2)
> Pure's XML import format and the XSD (sec. 2.4)
> Mapping types and fields to the Pure data model (sec. 2.6)
> Mapping external values to classifications (sec. 2.7)

Data validation

When evaluating an import, pay attention to these areas where problems occur most commonly:

- Check that each field has a single, clearly-defined use, ideally reflected in the name of the field. There should be no fields that contain miscellaneous information from the data source.
- Check for fields that are consistently empty. Though it is not mandatory to populate all fields, it could be a warning that you have forgotten to map certain values or mapped them incorrectly.

See the following related topics:

> Evaluating the quality of imported data (sec. 5)
> Bulk deleting records (sec. 5.4)

Initial population

Before you (optionally) configure an integration job for the Award content type, you can use the Bulk Import Wizard to import a subset of the data in an ad hoc fashion to test the data mapping.

To use the Bulk Import Wizard to import a Award:

1. Click the Administrator tab in the top navigation.
2. Click the Bulk import tab in the navigation.

See the following related topics:

> Migrating legacy data into Pure (sec. 3)
> Viewing the status of an import from the Bulk Import Wizard (sec. 3.2)
> Bulk populating content in Pure - detailed steps (sec. 3.3)

3. Click the *Awards and Projects* tab in the navigation.

   The Import wizard page is opened.

   Tip

   You can access documentation about the import process on this page, including XML schema definitions and example XML files.

4. Click Open import wizard.

   The Bulk Import Wizard is opened and guides you through the data import process.

   Tip

   For general information about populating records with the Bulk Import Wizard, see Migrating legacy data into Pure or Using the Bulk Import Wizard to import external data.

   For detailed steps about how to select records and verify the correctness of your data when importing records, see Bulk populating content in Pure - detailed steps.

Integration with an external data source

1. Create an integration job using the job named *Award synchronisation*.

   > Creating an integration job (sec. 4.1)

2. Configure the integration job based on the dataset you are integrating.

   > Configuring an integration job (sec. 4.3)

   Note

   If you have not previously imported records with this content type, we recommend that you start by integrating a small number of records so that you can manually check the accuracy of your data mapping and transformation processes.

3. Schedule a time for the job to run or run the job immediately.

   > Scheduling an integration job (sec. 4.4)
   > Running an integration job on an ad hoc basis (sec. 4.5)

4. After the job has run, use the logs to check whether it succeeded and which errors occurred. If no records were added to Pure, address the listed errors and try to run the integration job again.

   > Viewing the status of an integration job (sec. 4.6)
   > Using the job logs to evaluate import quality (sec. 5.1)

5. Check the status of the data you integrated and delete incorrect records.

   Depending on the configuration of the integration job, some records may be unable to be deleted in Pure. In this case, change the configuration settings and run the integration job again to update incorrect records.

   > Evaluating the quality of imported data (sec. 5)
   > Configuring an integration job (sec. 4.3)

6. If you deleted records, run the integration job again and check that both deleted and preserved records exist as expected. If the records have *Start date* and *End date* fields, check that these are correct.

7. Once you are satisfied that all records are in Pure and all fields are correctly configured, configure the integration job as you want it to run regularly, and set up your external data source to provide the source XML at the appropriate time.

   > Configuring an integration job (sec. 4.3)
   > Scheduling an integration job (sec. 4.4)

Tip

For general information about integrations, see Section 4. Integrating external data into Pure.

For detailed steps about how to select records and verify the correctness of your data when performing an integration, see Section 4.8. Integrating data sources with Pure - detailed steps.

6.16. Populating Project data

> **Tip**
> Before you begin, consider reading the introductory section of this manual. See Section 1. Introduction.

About

A Project in Pure is a record with information about a research project, such as the participants and collaborators, lifecycle, grant applications and awards, links to research output, activities, datasets, equipment and press clippings.

Requirements

Project records can be populated as soon as the content types Organization and Person have been populated.

See the following related topics:

> Populating content in the correct order (sec. 1.2.1)

Data preparation

When preparing data for an import, consider the following tips:

- Ensure all of the mandatory fields are complete. You can see which fields are mandatory by adding a new record in the Pure interface and looking for fields with a ✱ symbol next to the field name.

See the following related topics:

> Preparing data sources (sec. 2)
> Pure's XML import format and the XSD (sec. 2.4)
> Mapping types and fields to the Pure data model (sec. 2.6)
> Mapping external values to classifications (sec. 2.7)

Data validation

When evaluating an import, pay attention to these areas where problems occur most commonly:

- Check that each field has a single, clearly-defined use, ideally reflected in the name of the field. There should be no fields that contain miscellaneous information from the data source.
- Check for fields that are consistently empty. Though it is not mandatory to populate all fields, it could be a warning that you have forgotten to map certain values or mapped them incorrectly.

See the following related topics:

> Evaluating the quality of imported data (sec. 5)
> Bulk deleting records (sec. 5.4)

Initial population

Before you (optionally) configure an integration job for the Project content type, you can use the Bulk Import Wizard to import a subset of the data in an ad hoc fashion to test the data mapping.

To use the Bulk Import Wizard to import a Project:

1. Click the Administrator tab in the top navigation.
2. Click the Bulk import tab in the navigation.
3. Click the *Award and Project Import* tab in the navigation.

   The Import wizard page is opened.

> **Tip**
> You can access documentation about the import process on this page, including XML schema definitions and example XML files.

See the following related topics:

> Migrating legacy data into Pure (sec. 3)
> Viewing the status of an import from the Bulk Import Wizard (sec. 3.2)
> Bulk populating content in Pure - detailed steps (sec. 3.3)

4. Click Open import wizard.

   The Bulk Import Wizard is opened and guides you through the data import process.

> **Tip**
>
> For general information about populating records with the Bulk Import Wizard, see Migrating legacy data into Pure or Using the Bulk Import Wizard to import external data.
>
> For detailed steps about how to select records and verify the correctness of your data when importing records, see Bulk populating content in Pure - detailed steps.

Integration with an external data source

1. Create an integration job using the job named *Project synchronisation (new)*.

   > Creating an integration job (sec. 4.1)

2. Configure the integration job based on the dataset you are integrating.

   > Configuring an integration job (sec. 4.3)

> **Note**
>
> If you have not previously imported records with this content type, we recommend that you start by integrating a small number of records so that you can manually check the accuracy of your data mapping and transformation processes.

3. Schedule a time for the job to run or run the job immediately.

   > Scheduling an integration job (sec. 4.4)
   > Running an integration job on an ad hoc basis (sec. 4.5)

4. After the job has run, use the logs to check whether it succeeded and which errors occurred. If no records were added to Pure, address the listed errors and try to run the integration job again.

   > Viewing the status of an integration job (sec. 4.6)
   > Using the job logs to evaluate import quality (sec. 5.1)

5. Check the status of the data you integrated and delete incorrect records.

   Depending on the configuration of the integration job, some records may be unable to be deleted in Pure. In this case, change the configuration settings and run the integration job again to update incorrect records.

   > Evaluating the quality of imported data (sec. 5)
   > Configuring an integration job (sec. 4.3)

6. If you deleted records, run the integration job again and check that both deleted and preserved records exist as expected. If the records have *Start date* and *End date* fields, check that these are correct.

7. Once you are satisfied that all records are in Pure and all fields are correctly configured, configure the integration job as you want it to run regularly, and set up your external data source to provide the source XML at the appropriate time.

   > Configuring an integration job (sec. 4.3)
   > Scheduling an integration job (sec. 4.4)

> **Tip**
>
> For general information about integrations, see Section 4. Integrating external data into Pure.
>
> For detailed steps about how to select records and verify the correctness of your data when performing an integration, see Section 4.8. Integrating data sources with Pure - detailed steps.

6.17. Populating Dataset data

> **Tip**
> Before you begin, consider reading the introductory section of this manual. See Section 1. Introduction.

About

A Dataset record in Pure stores metadata about a dataset, as well as optionally the dataset itself.

Documents associated with Datasets can only be imported with the Bulk Import Wizard - only metadata can be synchronized with an integration job.

Requirements

Dataset records can be populated as soon as the content types Organization and Person have been populated.

However, they are commonly imported after research data.

> See the following related topics:
> > Populating content in the correct order (sec. 1.2.1)

Data preparation

When preparing data for an import, consider the following tips:

- Ensure all of the mandatory fields are complete. You can see which fields are mandatory by adding a new record in the Pure interface and looking for fields with a * symbol next to the field name.

> See the following related topics:
> > Preparing data sources (sec. 2)
> > Pure's XML import format and the XSD (sec. 2.4)
> > Mapping types and fields to the Pure data model (sec. 2.6)
> > Mapping external values to classifications (sec. 2.7)

Data validation

When evaluating an import, pay attention to these areas where problems occur most commonly:

- Check that each field has a single, clearly-defined use, ideally reflected in the name of the field. There should be no fields that contain miscellaneous information from the data source.
- Check for fields that are consistently empty. Though it is not mandatory to populate all fields, it could be a warning that you have forgotten to map certain values or mapped them incorrectly.

> See the following related topics:
> > Evaluating the quality of imported data (sec. 5)
> > Bulk deleting records (sec. 5.4)

Initial population

Before you (optionally) configure an integration job for the Dataset content type, you can use the Bulk Import Wizard to import a subset of the data in an ad hoc fashion to test the data mapping.

To use the Bulk Import Wizard to import a Dataset:

1. Click the Administrator tab in the top navigation.
2. Click the Bulk import tab in the navigation.
3. Click the *Dataset* tab in the navigation.

   The Import wizard page is opened.

> See the following related topics:
> > Migrating legacy data into Pure (sec. 3)
> > Viewing the status of an import from the Bulk Import Wizard (sec. 3.2)
> > Bulk populating content in Pure - detailed steps (sec. 3.3)

> **Tip**
> You can access documentation about the import process on this page, including XML schema definitions and example XML files.

4. Click Open import wizard.

   The Bulk Import Wizard is opened and guides you through the data import process.

> **Tip**
>
> For general information about populating records with the Bulk Import Wizard, see Migrating legacy data into Pure or Using the Bulk Import Wizard to import external data.
>
> For detailed steps about how to select records and verify the correctness of your data when importing records, see Bulk populating content in Pure - detailed steps.

Integration with an external data source

1. Create an integration job using the job named *Configurable dataset synchronisation*.

   > Creating an integration job (sec. 4.1)

2. Configure the integration job based on the dataset you are integrating.

   > Configuring an integration job (sec. 4.3)

> **Note**
>
> If you have not previously imported records with this content type, we recommend that you start by integrating a small number of records so that you can manually check the accuracy of your data mapping and transformation processes.

3. Schedule a time for the job to run or run the job immediately.

   > Scheduling an integration job (sec. 4.4)
   > Running an integration job on an ad hoc basis (sec. 4.5)

4. After the job has run, use the logs to check whether it succeeded and which errors occurred. If no records were added to Pure, address the listed errors and try to run the integration job again.

   > Viewing the status of an integration job (sec. 4.6)
   > Using the job logs to evaluate import quality (sec. 5.1)

5. Check the status of the data you integrated and delete incorrect records.

   Depending on the configuration of the integration job, some records may be unable to be deleted in Pure. In this case, change the configuration settings and run the integration job again to update incorrect records.

   > Evaluating the quality of imported data (sec. 5)
   > Configuring an integration job (sec. 4.3)

6. If you deleted records, run the integration job again and check that both deleted and preserved records exist as expected. If the records have *Start date* and *End date* fields, check that these are correct.

7. Once you are satisfied that all records are in Pure and all fields are correctly configured, configure the integration job as you want it to run regularly, and set up your external data source to provide the source XML at the appropriate time.

   > Configuring an integration job (sec. 4.3)
   > Scheduling an integration job (sec. 4.4)

> **Tip**
>
> For general information about integrations, see Section 4. Integrating external data into Pure.
>
> For detailed steps about how to select records and verify the correctness of your data when performing an integration, see Section 4.8. Integrating data sources with Pure - detailed steps.

6.18. Populating Press/Media data

> **Tip**
> Before you begin, consider reading the introductory section of this manual. See Section 1. Introduction.

About

Press/Media records describe press articles that were published in the media. These records may contain the article itself as text or an attachment in addition to the metadata.

Requirements

Press/Media item records can be populated as soon as the content types Organization and Person have been populated.

However, they are commonly imported after research data.

See the following related topics:

> Populating content in the correct order (sec. 1.2.1)

Data preparation

When preparing data for an import, consider the following tips:

- Ensure all of the mandatory fields are complete. You can see which fields are mandatory by adding a new record in the Pure interface and looking for fields with a * symbol next to the field name.

See the following related topics:

> Preparing data sources (sec. 2)
> Pure's XML import format and the XSD (sec. 2.4)
> Mapping types and fields to the Pure data model (sec. 2.6)
> Mapping external values to classifications (sec. 2.7)

Data validation

When evaluating an import, pay attention to these areas where problems occur most commonly:

- Check that each field has a single, clearly-defined use, ideally reflected in the name of the field. There should be no fields that contain miscellaneous information from the data source.
- Check for fields that are consistently empty. Though it is not mandatory to populate all fields, it could be a warning that you have forgotten to map certain values or mapped them incorrectly.

See the following related topics:

> Evaluating the quality of imported data (sec. 5)
> Bulk deleting records (sec. 5.4)

Initial population

Before you (optionally) configure an integration job for the Press/Media item content type, you can use the Bulk Import Wizard to import a subset of the data in an ad hoc fashion to test the data mapping.

To use the Bulk Import Wizard to import a Press/Media item:

1. Click the Administrator tab in the top navigation.
2. Click the Bulk import tab in the navigation.
3. Click the *Press / Media* tab in the navigation.

   The Import wizard page is opened.

> **Tip**
> You can access documentation about the import process on this page, including XML schema definitions and example XML files.

See the following related topics:

> Migrating legacy data into Pure (sec. 3)
> Viewing the status of an import from the Bulk Import Wizard (sec. 3.2)
> Bulk populating content in Pure - detailed steps (sec. 3.3)

4. Click Open import wizard.

   The Bulk Import Wizard is opened and guides you through the data import process.

> **Tip**
>
> For general information about populating records with the Bulk Import Wizard, see Migrating legacy data into Pure or Using the Bulk Import Wizard to import external data.
>
> For detailed steps about how to select records and verify the correctness of your data when importing records, see Bulk populating content in Pure - detailed steps.

Integration with an external data source

1. Create an integration job using the job named *Unified press/media synchronisation*.

   > Creating an integration job (sec. 4.1)

2. Configure the integration job based on the dataset you are integrating.

   > Configuring an integration job (sec. 4.3)

> **Note**
>
> If you have not previously imported records with this content type, we recommend that you start by integrating a small number of records so that you can manually check the accuracy of your data mapping and transformation processes.

3. Schedule a time for the job to run or run the job immediately.

   > Scheduling an integration job (sec. 4.4)
   > Running an integration job on an ad hoc basis (sec. 4.5)

4. After the job has run, use the logs to check whether it succeeded and which errors occurred. If no records were added to Pure, address the listed errors and try to run the integration job again.

   > Viewing the status of an integration job (sec. 4.6)
   > Using the job logs to evaluate import quality (sec. 5.1)

5. Check the status of the data you integrated and delete incorrect records.

   Depending on the configuration of the integration job, some records may be unable to be deleted in Pure. In this case, change the configuration settings and run the integration job again to update incorrect records.

   > Evaluating the quality of imported data (sec. 5)
   > Configuring an integration job (sec. 4.3)

6. If you deleted records, run the integration job again and check that both deleted and preserved records exist as expected. If the records have *Start date* and *End date* fields, check that these are correct.

7. Once you are satisfied that all records are in Pure and all fields are correctly configured, configure the integration job as you want it to run regularly, and set up your external data source to provide the source XML at the appropriate time.

   > Configuring an integration job (sec. 4.3)
   > Scheduling an integration job (sec. 4.4)

> **Tip**
>
> For general information about integrations, see Section 4. Integrating external data into Pure.
>
> For detailed steps about how to select records and verify the correctness of your data when performing an integration, see Section 4.8. Integrating data sources with Pure - detailed steps.

6.19. Populating Prize data

Tip

Before you begin, consider reading the introductory section of this manual. See Section 1. Introduction.

About

A Prize record in Pure describes prizes that have been conferred upon Persons affiliated with your institution, for example, a Nobel prize.

These records do not describe the ongoing activity that was the reason for or result of receiving the prize, such as research output published or stipends received. However, prizes can be linked to other content types, such as related (external) organizations, or research output etc.

Requirements

Prize records can be populated as soon as the content types Organization and Person have been populated.

However, they are commonly imported after research data.

See the following related topics:

> Populating content in the correct order (sec. 1.2.1)

Data preparation

When preparing data for an import, consider the following tips:

- Ensure all of the mandatory fields are complete. You can see which fields are mandatory by adding a new record in the Pure interface and looking for fields with a symbol next to the field name.

See the following related topics:

> Preparing data sources (sec. 2)
> Pure's XML import format and the XSD (sec. 2.4)
> Mapping types and fields to the Pure data model (sec. 2.6)
> Mapping external values to classifications (sec. 2.7)

Data validation

When evaluating an import, pay attention to these areas where problems occur most commonly:

See the following related topics:

> Evaluating the quality of imported data (sec. 5)
> Bulk deleting records (sec. 5.4)

- Check that each field has a single, clearly-defined use, ideally reflected in the name of the field. There should be no fields that contain miscellaneous information from the data source.
- Check for fields that are consistently empty. Though it is not mandatory to populate all fields, it could be a warning that you have forgotten to map certain values or mapped them incorrectly.

Initial population

Before you (optionally) configure an integration job for the Prize content type, you can use the Bulk Import Wizard to import a subset of the data in an ad hoc fashion to test the data mapping.

To use the Bulk Import Wizard to import a Prize:

See the following related topics:

> Migrating legacy data into Pure (sec. 3)
> Viewing the status of an import from the Bulk Import Wizard (sec. 3.2)
> Bulk populating content in Pure - detailed steps (sec. 3.3)

1. Click the Administrator tab in the top navigation.
2. Click the Bulk import tab in the navigation.

3. Click the *Prizes* tab in the navigation.

   The Import wizard page is opened.

   Tip

   You can access documentation about the import process on this page, including XML schema definitions and example XML files.

4. Click Open import wizard.

   The Bulk Import Wizard is opened and guides you through the data import process.

   Tip

   For general information about populating records with the Bulk Import Wizard, see Migrating legacy data into Pure or Using the Bulk Import Wizard to import external data.

   For detailed steps about how to select records and verify the correctness of your data when importing records, see Bulk populating content in Pure - detailed steps.

Integration with an external data source

1. Create an integration job using the job named *Configurable prize synchronisation*.

   > Creating an integration job (sec. 4.1)

2. Configure the integration job based on the dataset you are integrating.

   > Configuring an integration job (sec. 4.3)

   Note

   If you have not previously imported records with this content type, we recommend that you start by integrating a small number of records so that you can manually check the accuracy of your data mapping and transformation processes.

3. Schedule a time for the job to run or run the job immediately.

   > Scheduling an integration job (sec. 4.4)
   > Running an integration job on an ad hoc basis (sec. 4.5)

4. After the job has run, use the logs to check whether it succeeded and which errors occurred. If no records were added to Pure, address the listed errors and try to run the integration job again.

   > Viewing the status of an integration job (sec. 4.6)
   > Using the job logs to evaluate import quality (sec. 5.1)

5. Check the status of the data you integrated and delete incorrect records.

   Depending on the configuration of the integration job, some records may be unable to be deleted in Pure. In this case, change the configuration settings and run the integration job again to update incorrect records.

   > Evaluating the quality of imported data (sec. 5)
   > Configuring an integration job (sec. 4.3)

6. If you deleted records, run the integration job again and check that both deleted and preserved records exist as expected. If the records have *Start date* and *End date* fields, check that these are correct.

7. Once you are satisfied that all records are in Pure and all fields are correctly configured, configure the integration job as you want it to run regularly, and set up your external data source to provide the source XML at the appropriate time.

   > Configuring an integration job (sec. 4.3)
   > Scheduling an integration job (sec. 4.4)

Tip

For general information about integrations, see Section 4. Integrating external data into Pure.

For detailed steps about how to select records and verify the correctness of your data when performing an integration, see Section 4.8. Integrating data sources with Pure - detailed steps.

6.20. Populating Impact data

> **Tip**
> Before you begin, consider reading the introductory section of this manual. See Section 1. Introduction.

About

An Impact in Pure is a record that describes the impact that research has on society, and includes evidence of this impact.

Requirements

Impact records can be populated as soon as the content types Organization and Person have been populated.

See the following related topics:

> Populating content in the correct order (sec. 1.2.1)

Data preparation

When preparing data for an import, consider the following tips:

- Ensure the *Description* is plain-text and does not include HTML formatting.
- Ensure all of the mandatory fields are complete. You can see which fields are mandatory by adding a new record in the Pure interface and looking for fields with a symbol next to the field name.

See the following related topics:

> Preparing data sources (sec. 2)
> Pure's XML import format and the XSD (sec. 2.4)
> Mapping types and fields to the Pure data model (sec. 2.6)
> Mapping external values to classifications (sec. 2.7)

Data validation

When evaluating an import, pay attention to these areas where problems occur most commonly:

- Check that each field has a single, clearly-defined use, ideally reflected in the name of the field. There should be no fields that contain miscellaneous information from the data source.
- Check for fields that are consistently empty. Though it is not mandatory to populate all fields, it could be a warning that you have forgotten to map certain values or mapped them incorrectly.

See the following related topics:

> Evaluating the quality of imported data (sec. 5)
> Bulk deleting records (sec. 5.4)

Initial population

Before you (optionally) configure an integration job for the Impact content type, you can use the Bulk Import Wizard to import a subset of the data in an ad hoc fashion to test the data mapping.

To use the Bulk Import Wizard to import a Impact:

1. Click the Administrator tab in the top navigation.
2. Click the Bulk import tab in the navigation.
3. Click the *Impact* tab in the navigation.

   The Import wizard page is opened.

> **Tip**
> You can access documentation about the import process on this page, including XML schema definitions and example XML files.

See the following related topics:

> Migrating legacy data into Pure (sec. 3)
> Viewing the status of an import from the Bulk Import Wizard (sec. 3.2)
> Bulk populating content in Pure - detailed steps (sec. 3.3)

4. Click Open import wizard.

   The Bulk Import Wizard is opened and guides you through the data import process.

> **Tip**
>
> For general information about populating records with the Bulk Import Wizard, see Migrating legacy data into Pure or Using the Bulk Import Wizard to import external data.
>
> For detailed steps about how to select records and verify the correctness of your data when importing records, see Bulk populating content in Pure - detailed steps.

Integration with an external data source

1. Create an integration job using the job named *Configurable impact synchronisation*.

   > Creating an integration job (sec. 4.1)

2. Configure the integration job based on the dataset you are integrating.

   > Configuring an integration job (sec. 4.3)

> **Note**
>
> If you have not previously imported records with this content type, we recommend that you start by integrating a small number of records so that you can manually check the accuracy of your data mapping and transformation processes.

3. Schedule a time for the job to run or run the job immediately.

   > Scheduling an integration job (sec. 4.4)
   > Running an integration job on an ad hoc basis (sec. 4.5)

4. After the job has run, use the logs to check whether it succeeded and which errors occurred. If no records were added to Pure, address the listed errors and try to run the integration job again.

   > Viewing the status of an integration job (sec. 4.6)
   > Using the job logs to evaluate import quality (sec. 5.1)

5. Check the status of the data you integrated and delete incorrect records.

   Depending on the configuration of the integration job, some records may be unable to be deleted in Pure. In this case, change the configuration settings and run the integration job again to update incorrect records.

   > Evaluating the quality of imported data (sec. 5)
   > Configuring an integration job (sec. 4.3)

6. If you deleted records, run the integration job again and check that both deleted and preserved records exist as expected. If the records have *Start date* and *End date* fields, check that these are correct.

7. Once you are satisfied that all records are in Pure and all fields are correctly configured, configure the integration job as you want it to run regularly, and set up your external data source to provide the source XML at the appropriate time.

   > Configuring an integration job (sec. 4.3)
   > Scheduling an integration job (sec. 4.4)

> **Tip**
>
> For general information about integrations, see Section 4. Integrating external data into Pure.
>
> For detailed steps about how to select records and verify the correctness of your data when performing an integration, see Section 4.8. Integrating data sources with Pure - detailed steps.

6.21. Populating Course data

> **Tip**
> Before you begin, consider reading the introductory section of this manual. See Section 1. Introduction.

About

A Course record in Pure describes courses that have been developed and taught to students.

Requirements

Course records can be populated as soon as the content types Organization and Person have been populated.

See the following related topics:

> Populating content in the correct order (sec. 1.2.1)

Data preparation

When preparing data for an import, consider the following tips:

- Ensure the *Description* is plain-text and does not include HTML formatting.
- Ensure all of the mandatory fields are complete. You can see which fields are mandatory by adding a new record in the Pure interface and looking for fields with a * symbol next to the field name.

See the following related topics:

> Preparing data sources (sec. 2)
> Pure's XML import format and the XSD (sec. 2.4)
> Mapping types and fields to the Pure data model (sec. 2.6)
> Mapping external values to classifications (sec. 2.7)

Data validation

When evaluating an import, pay attention to these areas where problems occur most commonly:

- Check that each field has a single, clearly-defined use, ideally reflected in the name of the field. There should be no fields that contain miscellaneous information from the data source.
- Check for fields that are consistently empty. Though it is not mandatory to populate all fields, it could be a warning that you have forgotten to map certain values or mapped them incorrectly.

See the following related topics:

> Evaluating the quality of imported data (sec. 5)
> Bulk deleting records (sec. 5.4)

Initial population

Before you (optionally) configure an integration job for the Course content type, you can use the Bulk Import Wizard to import a subset of the data in an ad hoc fashion to test the data mapping.

To use the Bulk Import Wizard to import a Course:

1. Click the Administrator tab in the top navigation.
2. Click the Bulk import tab in the navigation.
3. Click the *Courses* tab in the navigation.
   The Import wizard page is opened.

   > **Tip**
   > You can access documentation about the import process on this page, including XML schema definitions and example XML files.

4. Click Open import wizard.
   The Bulk Import Wizard is opened and guides you through the data import process.

See the following related topics:

> Migrating legacy data into Pure (sec. 3)
> Viewing the status of an import from the Bulk Import Wizard (sec. 3.2)
> Bulk populating content in Pure - detailed steps (sec. 3.3)

Tip

For general information about populating records with the Bulk Import Wizard, see Migrating legacy data into Pure or Using the Bulk Import Wizard to import external data.

For detailed steps about how to select records and verify the correctness of your data when importing records, see Bulk populating content in Pure - detailed steps.

Integration with an external data source

1. Create an integration job using the job named *Configurable course synchronisation*.

   > Creating an integration job (sec. 4.1)

2. Configure the integration job based on the dataset you are integrating.

   > Configuring an integration job (sec. 4.3)

Note

If you have not previously imported records with this content type, we recommend that you start by integrating a small number of records so that you can manually check the accuracy of your data mapping and transformation processes.

3. Schedule a time for the job to run or run the job immediately.

   > Scheduling an integration job (sec. 4.4)
   > Running an integration job on an ad hoc basis (sec. 4.5)

4. After the job has run, use the logs to check whether it succeeded and which errors occurred. If no records were added to Pure, address the listed errors and try to run the integration job again.

   > Viewing the status of an integration job (sec. 4.6)
   > Using the job logs to evaluate import quality (sec. 5.1)

5. Check the status of the data you integrated and delete incorrect records.

   Depending on the configuration of the integration job, some records may be unable to be deleted in Pure. In this case, change the configuration settings and run the integration job again to update incorrect records.

   > Evaluating the quality of imported data (sec. 5)
   > Configuring an integration job (sec. 4.3)

6. If you deleted records, run the integration job again and check that both deleted and preserved records exist as expected. If the records have *Start date* and *End date* fields, check that these are correct.

7. Once you are satisfied that all records are in Pure and all fields are correctly configured, configure the integration job as you want it to run regularly, and set up your external data source to provide the source XML at the appropriate time.

   > Configuring an integration job (sec. 4.3)
   > Scheduling an integration job (sec. 4.4)

Tip

For general information about integrations, see Section 4. Integrating external data into Pure.

For detailed steps about how to select records and verify the correctness of your data when performing an integration, see Section 4.8. Integrating data sources with Pure - detailed steps.

6.22. Populating Student Thesis data

Tip

Before you begin, consider reading the introductory section of this manual. See Section 1. Introduction.

About

The student thesis content type collects data about various types of theses authored at your institution.

Data collected includes the qualification and award, author and supervisor information, and other information about the document.

Requirements

Student thesis records can be populated as soon as the content types Organization and Person have been populated.

It is also advisable to populate External organization records first, as these are often linked to thesis records.

See the following related topics:

> Populating content in the correct order (sec. 1.2.1)

Data preparation

When preparing data for an import, consider the following tips:

- Make sure that any full-text files or documents that you want to attach to the record are accessible from the Pure server. For example, ensure that the permissions are set appropriately.
- Ensure all of the mandatory fields are complete. You can see which fields are mandatory by adding a new record in the Pure interface and looking for fields with a * symbol next to the field name.

See the following related topics:

> Preparing data sources (sec. 2)
> Pure's XML import format and the XSD (sec. 2.4)
> Mapping types and fields to the Pure data model (sec. 2.6)
> Mapping external values to classifications (sec. 2.7)

Data validation

When evaluating an import, pay attention to these areas where problems occur most commonly:

- Check that each field has a single, clearly-defined use, ideally reflected in the name of the field. There should be no fields that contain miscellaneous information from the data source.
- Check for fields that are consistently empty. Though it is not mandatory to populate all fields, it could be a warning that you have forgotten to map certain values or mapped them incorrectly.

See the following related topics:

> Evaluating the quality of imported data (sec. 5)
> Bulk deleting records (sec. 5.4)

Initial population

Before you (optionally) configure an integration job for the Student thesis content type, you can use the Bulk Import Wizard to import a subset of the data in an ad hoc fashion to test the data mapping.

To use the Bulk Import Wizard to import a Student thesis:

1. Click the Administrator tab in the top navigation.
2. Click the Bulk import tab in the navigation.

See the following related topics:

> Migrating legacy data into Pure (sec. 3)
> Viewing the status of an import from the Bulk Import Wizard (sec. 3.2)
> Bulk populating content in Pure - detailed steps (sec. 3.3)

3. Click the *Student thesis* tab in the navigation.

   The Import wizard page is opened.

   Tip

   You can access documentation about the import process on this page, including XML schema definitions and example XML files.

4. Click Open import wizard.

   The Bulk Import Wizard is opened and guides you through the data import process.

   Tip

   For general information about populating records with the Bulk Import Wizard, see Migrating legacy data into Pure or Using the Bulk Import Wizard to import external data.

   For detailed steps about how to select records and verify the correctness of your data when importing records, see Bulk populating content in Pure - detailed steps.

Integration with an external data source

To integrate Pure with your institution's external data source for Student thesisFacilities/Equipment data:

1. Create an integration job using the job named *Configurable student thesis synchronisation*.

   > Creating an integration job (sec. 4.1)

2. Configure the integration job based on the dataset you are integrating.

   > Configuring an integration job (sec. 4.3)

   Note

   If you have not previously imported records with this content type, we recommend that you start by integrating a small number of records so that you can manually check the accuracy of your data mapping and transformation processes.

3. Schedule a time for the job to run or run the job immediately.

   > Scheduling an integration job (sec. 4.4)
   > Running an integration job on an ad hoc basis (sec. 4.5)

4. After the job has run, use the logs to check whether it succeeded and which errors occurred. If no records were added to Pure, address the listed errors and try to run the integration job again.

   > Viewing the status of an integration job (sec. 4.6)
   > Using the job logs to evaluate import quality (sec. 5.1)

5. Check the status of the data you integrated and delete incorrect records.

   Depending on the configuration of the integration job, some records may be unable to be deleted in Pure. In this case, change the configuration settings and run the integration job again to update incorrect records.

   > Evaluating the quality of imported data (sec. 5)
   > Configuring an integration job (sec. 4.3)

6. If you deleted records, run the integration job again and check that both deleted and preserved records exist as expected. If the records have *Start date* and *End date* fields, check that these are correct.

7. Once you are satisfied that all records are in Pure and all fields are correctly configured, configure the integration job as you want it to run regularly, and set up your external data source to provide the source XML at the appropriate time.

   > Configuring an integration job (sec. 4.3)
   > Scheduling an integration job (sec. 4.4)

Tip

For general information about integrations, see Section 4. Integrating external data into Pure.

For detailed steps about how to select records and verify the correctness of your data when performing an integration, see Section 4.8. Integrating data sources with Pure - detailed steps.

6.23. Populating Facilities/Equipment data

Tip

Before you begin, consider reading the introductory section of this manual. See Section 1. Introduction.

About

A Facilities/Equipment record represents an item of research infrastructure at your institution. This may be a facility, such as a building/room that may be booked, an item to be showcased or loaned etc.

Requirements

Facility/Equipment records can be populated as soon as the content types Organization and Person have been populated.

See the following related topics:

> Populating content in the correct order (sec. 1.2.1)

Data preparation

When preparing data for an import, consider the following tips:

- Ensure all of the mandatory fields are complete. You can see which fields are mandatory by adding a new record in the Pure interface and looking for fields with a symbol next to the field name.

See the following related topics:

> Preparing data sources (sec. 2)
> Pure's XML import format and the XSD (sec. 2.4)
> Mapping types and fields to the Pure data model (sec. 2.6)
> Mapping external values to classifications (sec. 2.7)

Data validation

When evaluating an import, pay attention to these areas where problems occur most commonly:

See the following related topics:

> Evaluating the quality of imported data (sec. 5)
> Bulk deleting records (sec. 5.4)

- Check that each field has a single, clearly-defined use, ideally reflected in the name of the field. There should be no fields that contain miscellaneous information from the data source.
- Check for fields that are consistently empty. Though it is not mandatory to populate all fields, it could be a warning that you have forgotten to map certain values or mapped them incorrectly.

Initial population

Before you (optionally) configure an integration job for the Facility/Equipment content type, you can use the Bulk Import Wizard to import a subset of the data in an ad hoc fashion to test the data mapping.

To use the Bulk Import Wizard to import a Facility/Equipment:

See the following related topics:

> Migrating legacy data into Pure (sec. 3)
> Viewing the status of an import from the Bulk Import Wizard (sec. 3.2)
> Bulk populating content in Pure - detailed steps (sec. 3.3)

1. Click the Administrator tab in the top navigation.
2. Click the Bulk import tab in the navigation.
3. Click the *Equipment* tab in the navigation.

   The Import wizard page is opened.

Tip

You can access documentation about the import process on this page, including XML schema definitions and example XML files.

4. Click Open import wizard.

   The Bulk Import Wizard is opened and guides you through the data import process.

> **Tip**
>
> For general information about populating records with the Bulk Import Wizard, see Migrating legacy data into Pure or Using the Bulk Import Wizard to import external data.
>
> For detailed steps about how to select records and verify the correctness of your data when importing records, see Bulk populating content in Pure - detailed steps.

Integration with an external data source

1. Create an integration job using the job named *Configurable equipment synchronisation*.

   > Creating an integration job (sec. 4.1)

2. Configure the integration job based on the dataset you are integrating.

   > Configuring an integration job (sec. 4.3)

> **Note**
>
> If you have not previously imported records with this content type, we recommend that you start by integrating a small number of records so that you can manually check the accuracy of your data mapping and transformation processes.

3. Schedule a time for the job to run or run the job immediately.

   > Scheduling an integration job (sec. 4.4)
   > Running an integration job on an ad hoc basis (sec. 4.5)

4. After the job has run, use the logs to check whether it succeeded and which errors occurred. If no records were added to Pure, address the listed errors and try to run the integration job again.

   > Viewing the status of an integration job (sec. 4.6)
   > Using the job logs to evaluate import quality (sec. 5.1)

5. Check the status of the data you integrated and delete incorrect records.

   Depending on the configuration of the integration job, some records may be unable to be deleted in Pure. In this case, change the configuration settings and run the integration job again to update incorrect records.

   > Evaluating the quality of imported data (sec. 5)
   > Configuring an integration job (sec. 4.3)

6. If you deleted records, run the integration job again and check that both deleted and preserved records exist as expected. If the records have *Start date* and *End date* fields, check that these are correct.

7. Once you are satisfied that all records are in Pure and all fields are correctly configured, configure the integration job as you want it to run regularly, and set up your external data source to provide the source XML at the appropriate time.

   > Configuring an integration job (sec. 4.3)
   > Scheduling an integration job (sec. 4.4)

> **Tip**
>
> For general information about integrations, see Section 4. Integrating external data into Pure.
>
> For detailed steps about how to select records and verify the correctness of your data when performing an integration, see Section 4.8. Integrating data sources with Pure - detailed steps.