Getting started with Pure

What is the purpose of Pure?

Pure's main function is that of the Current Research Information System (CRIS) for Eindhoven University. Pure is the source system for TU/e research output in all forms; from articles and books to software and datasets. This (meta)data is entered and stored so that Pure can give an up to date view on TU/e research output, which is used in various other workflows, products and reports:

- Data from Pure is used in several evaluations and assessments such as the Standard Evaluation Protocol (SEP) created by VSNU, NWO and KNAW
- Publication statistics for the TU/e year report are derived from Pure
- The TU/e repository is based on the data coming from Pure
- The National Academic Research and Collaborations Information System (Narcis) uses data from our repository (which in turn comes from Pure)
- The TU/e repository is seen as an institutional repository by the Dutch government and therefore qualifies as a means to publish open access. Any publication which is openly accessible from the TU/e repository can be counted towards open access requirements and thus helps us to meet the goals set. More on open access here.
- Data from Pure is used to populate the research part of the TU/e website; it supplies data about research groups, researchers and their projects and research output. From the research website it is possible to dive into more detail in the research portal.

Traditionally, a CRIS/Repository would serve as a storage system for research output (in the form of metadata as well as full text) with the purpose of reporting in situations as described above. It serves as the long term storage place for research output.

There is also increased demand from funding bodies as well as project managers to track progress and results during the course of a project via its output. This requires the link to be made from application to funding to project to result in the form of publications. Pure creates this chain by linking results back to the project. This is why we also import project data into Pure from the project administration. Pure can provide automated periodic reports on output related to specific projects.

Which data is stored in Pure, and what do I need to supply?

Since Pure is the TU/e source system for research output, this is the predominant type of data found in Pure. Alongside the publication data there is a number of related metadata found in the system, which can be linked to a publication to provide context; staff, projects, facilities, equipment, media coverage, courses and the internal TU/e organisational structure are some examples.

Most of these entities are not maintained in Pure; they are synchronized from
other software applications on campus, making Pure the center of a complex web of data. The researcher's job is to maintain research output in Pure that is the result of the research appointment at TU/e. It is also possible to enter research output from previous or external appointments into Pure to form a complete personal overview of research, but when doing so it is important to set your affiliation correctly. The publication data can be linked to all related entities to give a broad context of research.

Pure is the official TU/e source system for publications, activities, events, prizes, datasets and student theses. Facilities, equipment and press mentions are also optionally maintained in Pure.

Most of these data types can be linked to form a context for research: a publication is affiliated to organisations via its author, and has links to projects, equipment or facilities. Research presented at a convention can be linked to events, the activity and any awarded prizes (note that an award in Pure is project funding). Staff can have relations to courses, media mentions, organisations, facilities etc. An end user of Pure can create most of these relations in Pure as required, but please note that staff, organisations, appointments and projects cannot be maintained in Pure. They are read only since this data is sourced from other systems on campus.

A more detailed affiliation

Affiliation is usually regarded as the link between author, organisation and publication; in the scientific context it's important to know the organisation which the research (output) can be contributed to via the author. The publication title page usually shows the affiliation on university level; this information is stored by the publisher and shown in various online databases.

Within Pure, affiliation is an important part of the metadata that describes the publication. But since Pure contains the complete internal organisation structure of Eindhoven University, affiliations in Pure can be set to a more detailed level, which is mainly important for reporting purposes within the university. This is why Pure asks you to select your research group or institution when importing or manually adding publications.

Information about affiliations can be found here.

Where is the organisation list I use in Pure for affiliations derived from?

There are a number of places in Pure where you need to select a relevant TU/e organisation, for instance when affiliating a publication. The research organisation is a complex structure; there are three main structures on campus:
The research organisation is a mix of financial and virtual organisations. Also each faculty has their own labelling for different levels in the structure. In Pure this is simplified to a limited set of organisation types. More information on the organisation structure in Pure can be found [here](https://purefaq.tue.nl/pure/faq/index.php?action=artikel&cat=0&id=70&artlang=en).

Since the research organisation is becoming more flexible as well as virtual, it is a complex thing to visualize in Pure. Because of that this is an ongoing process of improvement.

**My data is already in the publisher database or in Google scholar, why do I need to enter in into Pure?**

Once a publication has been submitted to the publisher, the full text and metadata are copied into various other systems. Major research database vendors such as Elsevier and Clarivate aggregate publications from multiple sources to create large research databases that can be queried or examined. Elsevier has a repository called Scopus, and Thomson Reuters has Web of Science. These databases are mostly used for building commercial products that give insight in research trends; Elsevier offers Scival and Thomson Reuters offers Incites as research tools built on the respective databases. And of course Google uses its indexing and harvesting techniques to gather as many publications as possible in Google Scholar.

Most CRIS systems have functionality to import publication metadata from other systems (harvesting); Pure can harvest from Scopus, Web of Science, PubMed, ArXiv, etc. This leverages researchers in that they do not need to enter the complete metadata record by hand. Some systems have the option to harvest (at a price), others just flat out refuse it (Google Scholar). Most publications end up in several places depending on what can be harvested or linked from the original publisher, but not all data is available at all levels. We’ve seen that localized affiliations are not available at the publisher level, and not all research data is made public. This is where Pure plays an important role since there is a need to report on the complete set of TU/e research, and Pure is the only system where this can be performed.

We’re constantly trying to simplify the workload for researchers by implementing new import sources supplied by Pure and suggesting new sources to Elsevier to implement in the software.
So what do I need to do to get started?

Once you start your appointment at TU/e, you should have received a staff ID number and network login. Pure needs both, the staff number to identify you as a person/author in Pure and the network login will get you into Pure. It is however important that these two are linked together, this task is performed by your local HR representative. It might take a day or so for the login to start working after everything has been set up. In case of problems please read this explanation.

Once you've logged into Pure, the first thing you need to do is set up the automated import to help you get your retrospective research into Pure. This is explained here.

We recommend that each researcher maintains their own research output in Pure. In some cases administrative staff at the research group or unit will assist or even completely take over research output maintenance. Check with your local faculty to learn more about the local workflow. The automated import will suggest results to you which can be imported into Pure. There are multiple questions and answers in this FAQ to assist, more information here.

If you encounter any other issues please contact the helpdesk via mail.

Which import sources should I pick?

Pure has several import sources to choose from. The most relevant sources are Scopus and Web of Science (WoS). While you can set up multiple import sources, it is important to note that there are consequences to the order in which the import sources are used. When a publication is found in multiple import sources, it is imported from the first source you select/use. For any subsequent source, Pure will inform you that a duplicate title exists (since you already imported it from another source). Most users now choose to reject (i.e. remove) the duplicate matches from other sources from the list to stop it from reappearing. In theory you could download the title from both sources and merge it, preserving all metadata from both sources. Most people do not do this since it is a tedious activity. So what happens then?

The Scopus route: a title is imported from Scopus and rejected from other sources (such as WoS). The title is loaded into Pure including Scopus specific fields and ID's.

The WoS route: a title is imported from Web of Science and rejected from other sources (such as Scopus). The title is loaded into Pure including any WoS specific data and ID's.

Choosing either route means that information (mainly ID's) from the other is missing. Merging is however a complex option, and Pure runs a process that will try
to find Scopus ID's for titles from the WoS route (not the other way around unfortunately).

Scopus and/or WoS ID's are important in maintaining data quality, and more importantly the data links between systems work best when they are based on the ID's. Importing from Scopus into Pure works best when Pure contains the Scopus Author ID, and analysis in Scival (see next chapter) works best when it can use the Scopus Article ID's in Pure.

Scopus and WoS each contain about 30,000 publications affiliated to TU/e over the period 2000 - early 2017. There are minor differences in publication types; Scopus slightly more conference proceedings. So we recommend using Scopus as the main source of import at all times, and only when titles cannot be found there, the WoS import should be used. If you think that important import sources are missing, please mention this to the Pure helpdesk.

**Why the focus on Elsevier products?**

Currently there are only two major vendors on the market providing research discovery tools; Elsevier and Clarivate (formerly Thomson Reuters). Both provide a toolset where three systems are linked together:

- An international publication database (Scopus/WoS)
- A current research information system (Pure/Converis)
- An analysis tool to compare the university with other institutions (Scival/Incites)

The publication database is filled using publication input from publishers, the CRIS is (mostly) filled using this data, and the analysis tool uses both the database and CRIS to create indexes and overviews.
Since each publisher built or bought their own set of software tools, these tools work in tandem and are best used within the same vendor ecosystem. TU/e chose Pure as the successor to Metis (Not long after that Thomson Reuters decided not to offer Converis in the Netherlands anymore. Almost all Dutch universities currently use Pure). Pure works best when importing from Scopus, Scival is based on Scopus data; all three systems benefit from each other. Scival is used to analyze performance based on indexes built on Scopus, and any fixes to data in Scopus (detected from Scival analysis) will improve matching of data in Pure.

Ideally each part of both triangles should be interchangable, something that all Elsevier and Thomson Reuters customers demand. But in reality this will never be a viable option. For instance, both publishers offer abstracts to be downloaded into the CRIS, but Elsevier is not allowed to show WoS abstracts on any public Pure interface for copyright reasons.

**How much effort do I need to put in Pure in the long term?**

Once your publication backlog is cleared and an automated search has been set up, you will be notified by email of any new matches. Pure searches for import candidates on an hourly schedule, and will inform you by email whenever matches are found. After importing new candidates or rejecting duplicates, the worklog will be cleared. Depending on the volume of research output you create we estimate that Pure maintenance should not require more than an hour per month on average. Maintaining output in Pure subsequenly saves you and other staff time further down the road, for instance when creating reports for assessments.

**Who do I talk to when I have issues or questions?**

Eindhoven university has set up a multi tier support structure when using Pure:

- If you have any specific questions on day to day use of Pure (including validation), consult this FAQ or contact the helpdesk via mail.
- The faculty policy advisor is the primary contact for Pure when it comes to reporting or assessments on faculty level.
- The faculty HR advisor can assist in any issues dealing with inconsistencies in your appointment(s) and workplace(s).
- The information Expertise Center maintains Pure, hosts the helpdesk and FAQ, and has a backoffice team that validates all research output entered in Pure.