Pure documentation for the OADO

1. Pulling data from Pure

All data used in the OADO is pulled from SDU's Pure system. Pure is a RIMS developed by Elsevier and is used by all Danish universities. The reports from Pure do not use complex data and it should be possible to copy this setup to other institutions with or even without Pure as long as open access is also managed or registered in the local RIMS. Depending on the data available, some modifications might be needed.

1.1 Publication report in Pure

Publications are pulled from the last five years not counting current year, starting from 2018 as it is the year the SDU Open Science Policy was established.

Values are added for title, publication type, submission year, publication category, public access to file, name, and current active organization,

Filters are added for submission year (2018-2022), current publication status (published), publication category is either 'research' or 'communication', person is currently active at SDU, current active organizations are filtered down to the organization in question (i.e., The library of the University of Southern Denmark).

1.2 Dataset report in Pure

Datasets are pulled from Pure. Datasets are not uploaded in Pure, so what is found in Pure is metadata entries that describe datasets.

Datasets are filtered by active persons at the department in question and latest five full years is set in 'Date made available'.

2. Open Access choices and limitations

2.1 Open Access data in Pure

There are several ways to report on open access in Pure. Public Access to file counts all electronic versions and returns a value for each version. The version can be either a link, DOI or a file. The advantage of using this way to indicate open access is that it can also count preprints as open.

For instance, this publication has three electronic versions labelled as 'closed', 'embargoed' and 'open' in the extracted data.

Title		Type		on yea	category	Peer review		Open Access status		
4 Fragmented publishing		Journal article		2019	Research	Peer-reviewed		Closed // Embargoed // Open		Eriks
9 PubMed coverage varied across		Journal article		2019	Research	Peer-reviewed		Closed // Embargoed // Open		Eriks

In Pure, we can see the three electronic versions are three different versions of which two are actually open access since the embargo on the accepted manuscript is already expired.

Electronic version(s) of this work
SUBMITTED MANUSCRIPT

Fragmented publishing_preprint
Submitted manuscript
Open
307 KB   3. Jun 2019 12:55

ACCEPTED MANUSCRIPT

Fragmented publishing
Accepted manuscript
Open   Embargo ended: 25/04/2020
307 KB | 23. May 2019 10:51

FINAL PUBLISHED VERSION

10.1007/s11192-019-03109-9
Final published version
Closed
Add electronic version (file, DOI, or link)...

Other links
Add other link...

Other files
Add other file...

Publications will be counted as open if one of the added electronic versions are 'open' or 'under embargo'. Keep in mind that the embargo may or may not be expired yet at the time of the calculation of the OADO – it will still be counted as open access since the file (often an accepted manuscript) has already been uploaded and will be open access in the future.

The OADO does not differentiate between which version of the publication is open access. A submitted manuscript, accepted manuscript and final, published version is counted equally.

Files and links may also be added as 'Other link' or 'Other file' in Pure. These will not be counted in the OADO since these fields are used for supplementary files and links that are not any version of the actual publication.

2.3 Possible sources of error in open access data

Open access in Pure is a manual process and differences in data quality and practices occur.

Data quality is expected to be better in recent times and going back, more inconsistencies will be present in data. We expect data to be significantly improved in 2017 and onwards since this was the time we started to implement standardization and documentation for how to handle open access in Pure. However, as a common practice takes time to implement and adjust, data quality in 2022 is expected to still be significantly better than in 2018.

Open access in Pure is handled both by authors of the publications and by the Library Team. Any author may add links or files to a registration. The Library Team will validate the links and files during the validation process.

Files uploaded and made openly available are monitored and have been monitored for the last three years by the Open Access Team in the Pure office.

Open Access practices for non-research publications are handled much more inconsistently than research publications. While files are monitored and checked the same is not done systematically for links.

The practice for adding links to research publications is that a publication must be open access with a license or at least made open access in an open access journal where all content is open access. 'Free' Open access is not counted as open access for research publications.

For communication links are allowed to be set to open much more if access is granted from the links. This is due to open access licenses being uncommon for these types of publications. We expect a greater inconsistency in the open access status of these publications.

3. Research and communication

All publications in Pure are divided into subcategories of either *Research*, *Communication*, *Commissioned* or *Education*.

Not all categories are available for all publication types. For instance, journal articles can be divided into *Research*, *Communication* or *Commissioned* but not *Education*. Book chapters can instead be divided into *Research*, *Communcation* or *Education*.

Other publication types will always fall into one subcategory and cannot be changed. An example of this is a PhD thesis which will always be labelled Research.

The definition for the *Research* category is as follows:
A product of a research activity that complies with the academic quality within the field and contributes to the development of the research field.

The definition of the *Communication* category is as follows: A product that seeks to share research findings with a wider audience or stakeholders.

The definitions are visible to researchers and other personnel when they enter data into the system.

When the categories are available on publication types, researchers will be able to choose the template they find most fitting as well as correct categories later if they have not themselves filled out the information from the beginning.

Categories are selected or changed by The Library Team when there is evidence that the correct category has not been selected.

Furthermore, the categories are sent to heads of departments monthly with the emphasis on that they should either approve or act when the categories are not correct.

4. Data processing in Excel

Data is processed in excel. See "OADO metric template.xlsx".

4.1 Sorting publications into 'OA' and 'Not OA'

Publications are sorted into two simple open access categories based on the column E 'Open Access Status'.

Open Access is determined based on the *most open* access to a publication since a publication can have more than one electronic version added.

Column G looks in column E and searches for any value that matches 'Open' in the text. If 'Open' is not found, it will look for 'embargoed' and return this value. If none of these values are found, the column will display an error.

```
=IFERROR(MID([@[Open Access status]];SEARCH("Open";[@[Open Access status]])+0;4);(MID([@[Open Access status]];SEARCH("Embargoed";[@[Open Access status]])+0;9)))
```

	B	C	D	E	F	G	H
	Type	Submission year	Publication category	Open Access status	Name	Open Access category	Final open acces category
as Working paper		2022	Research	Closed		#VALUE!	Not OA
re Net publication - Internet pu		2018	Communication	Open		Open	OA
ef Journal article		2019	Research	Open		Open	OA
ef Journal article		2019	Research	Open		Open	OA

Column H adjust names for Colum G and displays only 'OA' for the two values 'Open' and 'Embargoed'. If column G does not have one of these two values, column H will show 'Not OA'

```
=IFERROR(IFS([@[Open Access category]]="Open";"OA";[@[Open Access category]]="Embargoed";"OA");"Not OA")
```

	B	C	D	E	F	G	H
	Type	Submission year	Publication category	Open Access status	Name	Open Access category	Final open acces category
as Working paper		2022	Research	Closed		#VALUE!	Not OA
re Net publication - Internet pu		2018	Communication	Open		Open	OA
ef Journal article		2019	Research	Open		Open	OA

4.2 Counting publications

Publications are counted for both department and for each researcher. For further calculations we need to count the number of open access publications as well as the total number of publications of each of the types 'Research' and 'Communication'.

The counting prior to the calculation of the OADO metric is done in the excel sheet tab 'Count of publications'. The counting is done in a pivot table based on the data in the 'Publications' tab.

There is also a small table where total number of people are counted based on column A in the pivot table. Only people with at least one publication labelled either 'Research' or 'Communication' is included in the OADO metric which means that a person with only one 'Communication' publication and no 'Research' publication will be included in the calculations.

5. Calculating the OADO
All calculations are done in the tab 'OADO metric' in the Excel sheet of "OADO metric template.xlsx".

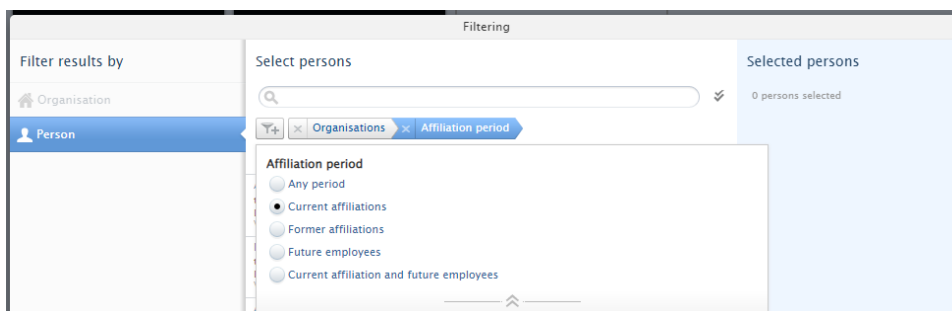## 6. Step-by-step guide for repeating the processes

6.1 How to make a new publication report in Pure:
1. Import the report template (called "OADO metric template.json") into the Pure reporting module (new reporting module)
2. Change the submission year to the last complete five years
3. Change the organizational unit to the one you need to filter on
4. Export the data in the Excel format – this may take a few minutes

This part takes 5-10 minutes (if template is used)

6.2 How to make a new dataset report in Pure:
1. Use the original reporting module and import the template "Datasets_for_OADO_metric.xml".
2. Open the filter called 'Organisational unit' and delete all selected persons to the right. Add a filter for 'Organisations' and seach for the relevant organization. Tick the box called 'Selected and all underlying organisations'. Choose 'Select all' and 'apply.

3. Change dates to the latest full five years.
4. Pull data out in XLS.

This part takes 5-10 minutes (if template is used)

6.3 How to process new data in Excel

1. Open the report pulled from Pure
2. Copy data from cell A2-F2 and all data below
3. Paste data into cell A2 in the excel template, sheet 'Publications'
4. Update the pivot table in tab 'Count of publications.
5. Copy all names in 'Count of publications' column A starting with cell A7.
6. Paste them into the tab called 'OADO metric' in cell A2.
7. Open your dataset report and copy data from column A-C starting with row 2.
8. Paste this data into the Excel template in the tab called 'Dataset' starting in cell A2

This part takes approx. 15 minutes.