

Test af Data Monitor-integration i Pure

Detlevsen, Mette; Grip, Lone

DOI:
10.21996/2n6t-vt61

Publication date:
2020

Document version:
Forlagets udgivne version

Citation for published version (APA):
Detlevsen, M., & Grip, L. (2020, aug. 21). Test af Data Monitor-integration i Pure. Syddansk Universitet.
<https://doi.org/10.21996/2n6t-vt61>

Go to publication entry in University of Southern Denmark's Research Portal

Terms of use

This work is brought to you by the University of Southern Denmark.
Unless otherwise specified it has been shared according to the terms for self-archiving.
If no other license is stated, these terms apply:

- You may download this work for personal use only.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying this open access version

If you believe that this document breaches copyright please contact us providing details and we will investigate your claim.
Please direct all enquiries to puresupport@bib.sdu.dk

Test af Data Monitor-integration i Pure

Forfattere: Mette Detlevsen & Lone Grip, Syddansk Universitetsbibliotek

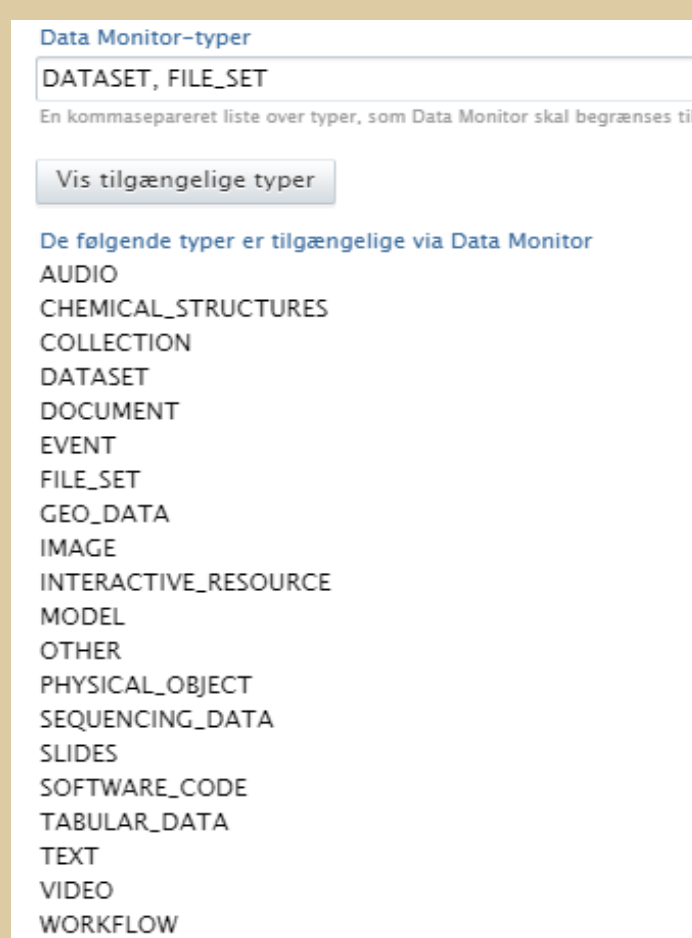
På baggrund af præsentationen *Automatically track datasets in Pure, wherever they were published** givet på den internationale Pure konference 2019, deltog Syddansk Universitetsbibliotek i en test af den nye Data Monitor-integration.

Baggrund

Data Monitor er en Elsevier udviklet tjeneste, der indekserer datasæt fra forskellige kilder til import i CRIS-systemer. I tjenestens berigede version sammenholdes datasæt med publikations-, forfatter- og affilieringsdata for at øge andelen af datasæt, der kan attribueres til en institution. I den berigede version opdateres datasættene desuden, hvis der foretages ændringer i kildedata. Data monitor-integrationen i Pure har funktionalitet, så import af datasæt kan ske enten automatisk eller manuelt afhængig af opsætning.

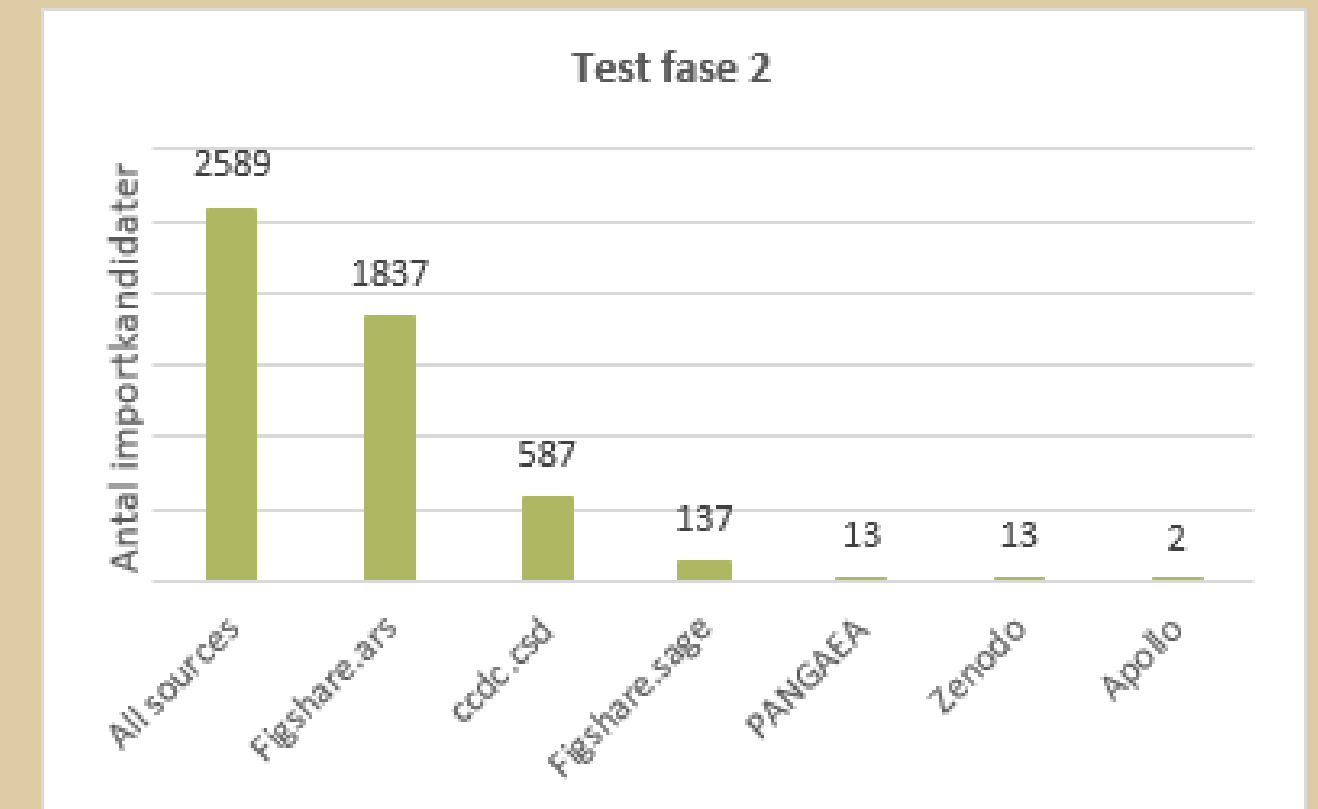
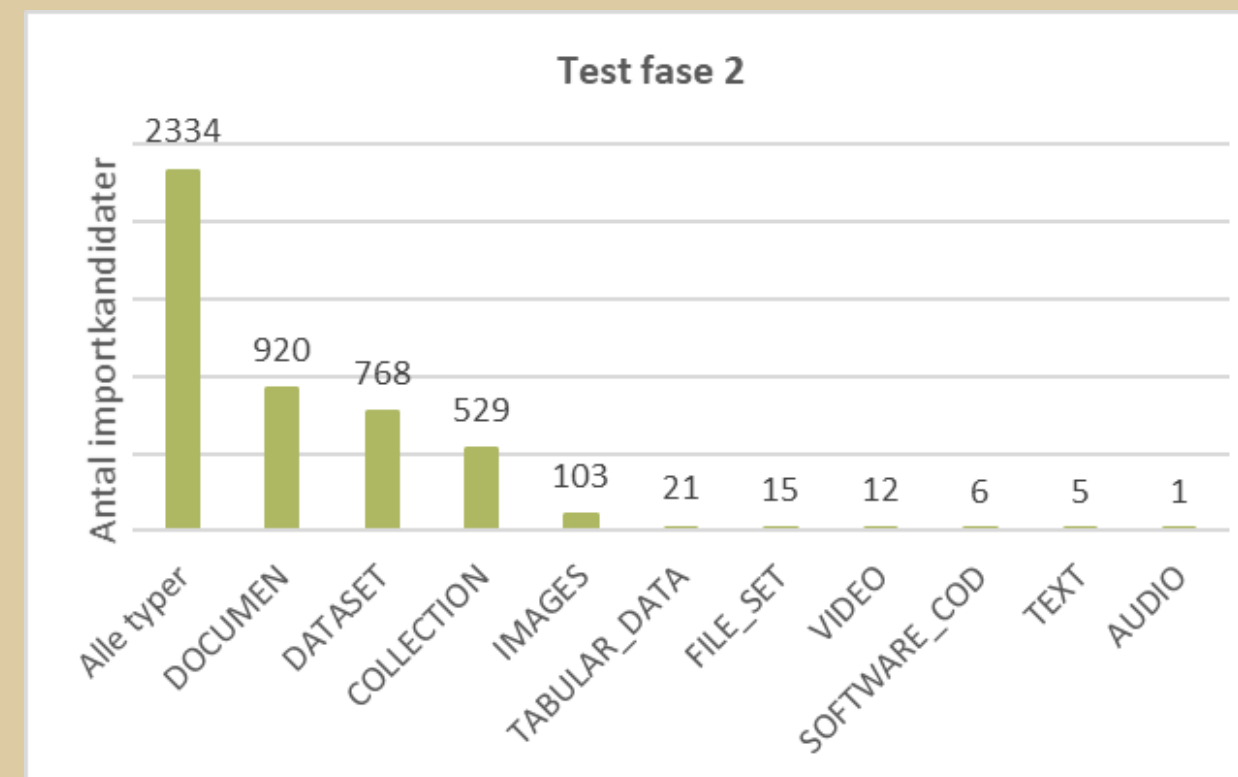
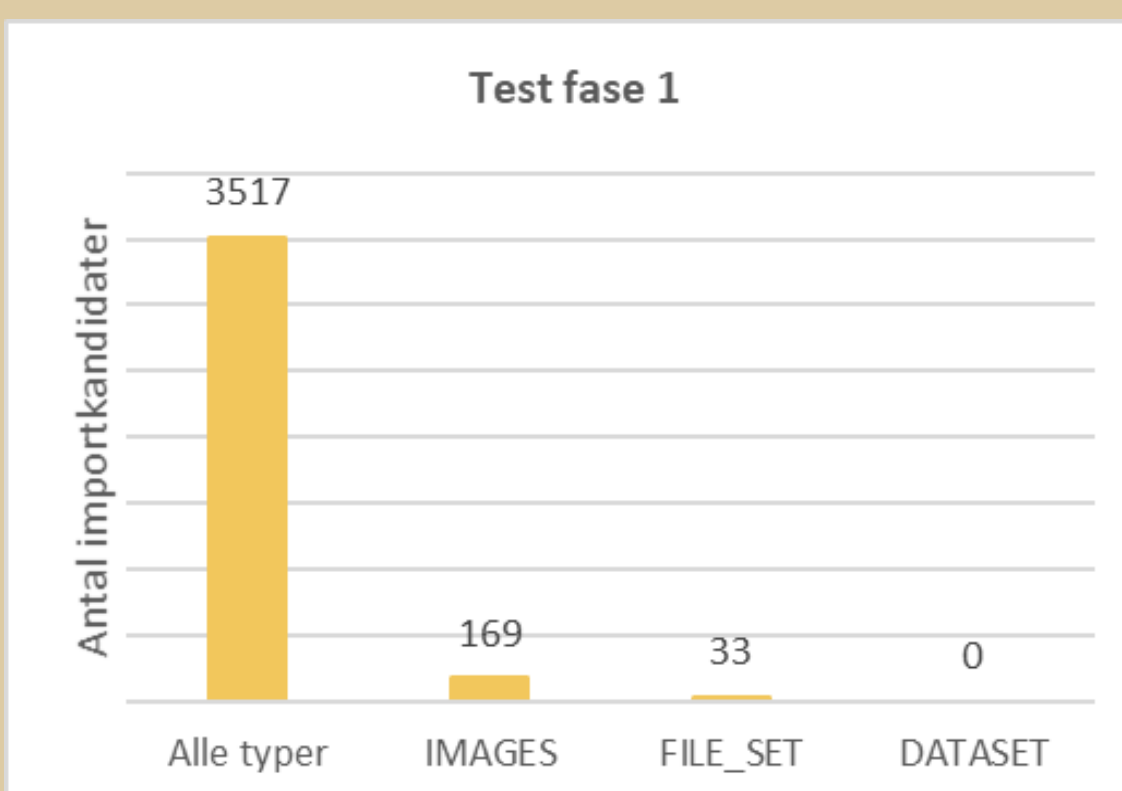
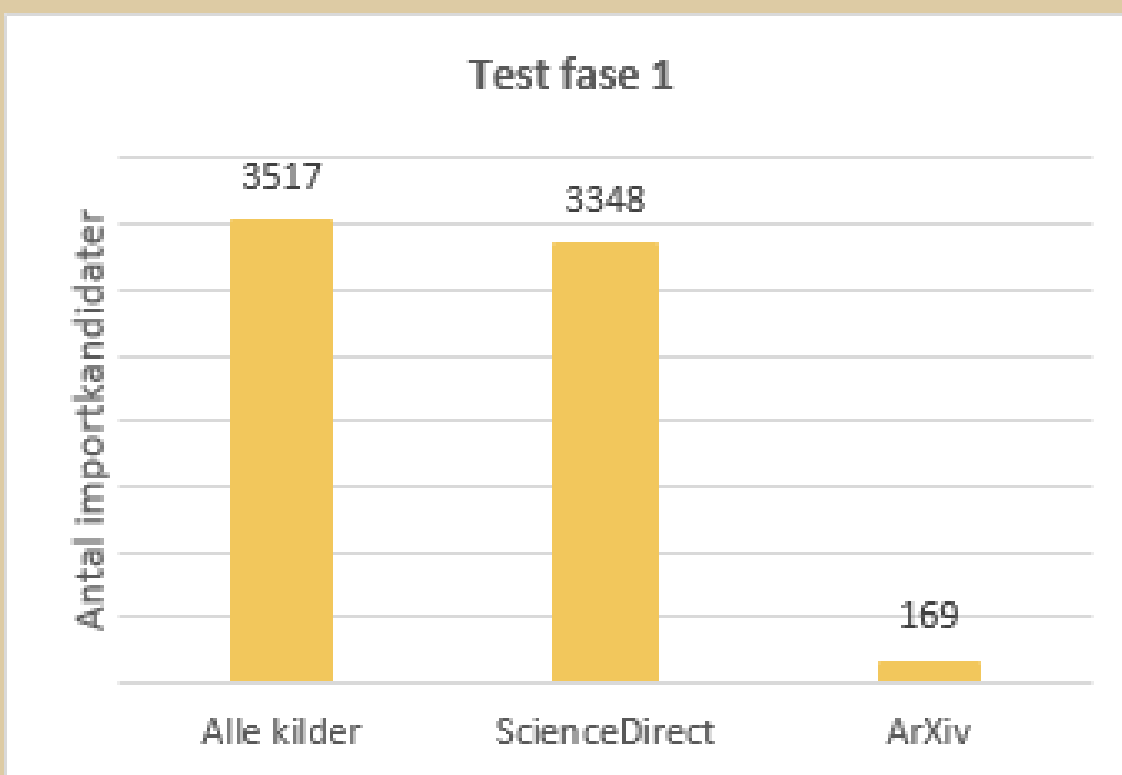
Formål

- Vurdere om tjenesten kan benyttes til at identificere og importere datasæt til SDU's Pure
- Vurdere om det skaber merværdi at benytte den berigede version
- Afklare om der skabes relation mellem importerede datasæt og relaterede publikationer
- Vurdere datakvaliteten af datasættene



Metode

Testen af Data Monitor-integrationen blev styret af Elsevier med deltagelse af fem forskellige europæiske universiteter. Testen bestod af to faser. I fase 1 blev den ikke-berigede version testet og i fase 2 den berigede version. Til de to faser blev der stillet en kort række spørgsmål til deltagernes erfaringer og resultater samt afholdt løbende onlinemøder til diskussion af disse. Testen løb fra februar 2020 til slut juni 2020.



Resultater

Identifikation af datasæt:

- ✗ Den ikke-berigede version finder kun datasæt fra to kilder
- ✗ Den ikke-berigede version finder primært datasæt, der er tæt relateret til publikationerne, dvs. billeder, grafer og tabeller, der allerede forekommer i publikationerne
- ✓ Den berigede version finder datasæt fra betydeligt flere forskellige kilder
- ✓ Den berigede version finder betydeligt flere relevante datasæt. Resultaterne inkluderer supplerende data, interviewguides og kemiske formler

Automatisk import af datasæt:

- ✗ Ingen mulighed for at redigere i datasættets felter, ej heller bidragsydernes organisationstilknytninger
- ✗ Ingen mulighed for at fjerne en tilknyttet person fra datasæt efter import

Relationer mellem datasæt og publikationer:

- ✓ Relationen mellem datasæt og publikation i Pure skabes

Diskussion

Det er nødvendigt at klarlægge, hvad man betragter som et datasæt, da der er stor bredde i importkandidaterne. Integrationens værdi afhænger således af, om afgrænsningsmulighederne på kilde og type kan imødekomme din institutions definition af et datasæt. En restriktiv definition af datasæt vil potentielt begrænse antallet af importkandidater.

Muligheden for at vælge mellem automatisk eller manuel import af datasæt er positiv, men de indbyggede redigeringsrestriktioner er problematiske uanset importmetode.

Konklusion

Den berigede version af tjenesten identificerer flere datasæt fra flere kilder og af mere varierende typer end den ikke-berigede og skaber dermed mere værdi. Data Monitor-integrationens svageste led er dog af teknisk art. Løsningen giver ikke fri mulighed for at redigere i datasættets felter efter import og det faktum, at det ikke er muligt at fjerne en tilknyttet intern organisation fra et importeret datasæt er yderst problematisk ift. datakvalitet.

* <https://www.brighttalk.com/webcast/13819/376860>

Test of the Data Monitor-integration in Pure

Authors: Mette Detlevsen & Lone Grip, University Library of Southern Denmark

The University Library of Southern Denmark decided to participate in a test of the new Data Monitor-integration based on the presentation *Automatically track datasets in Pure, wherever they were published** given at the 2019 International Pure Conference.

Background

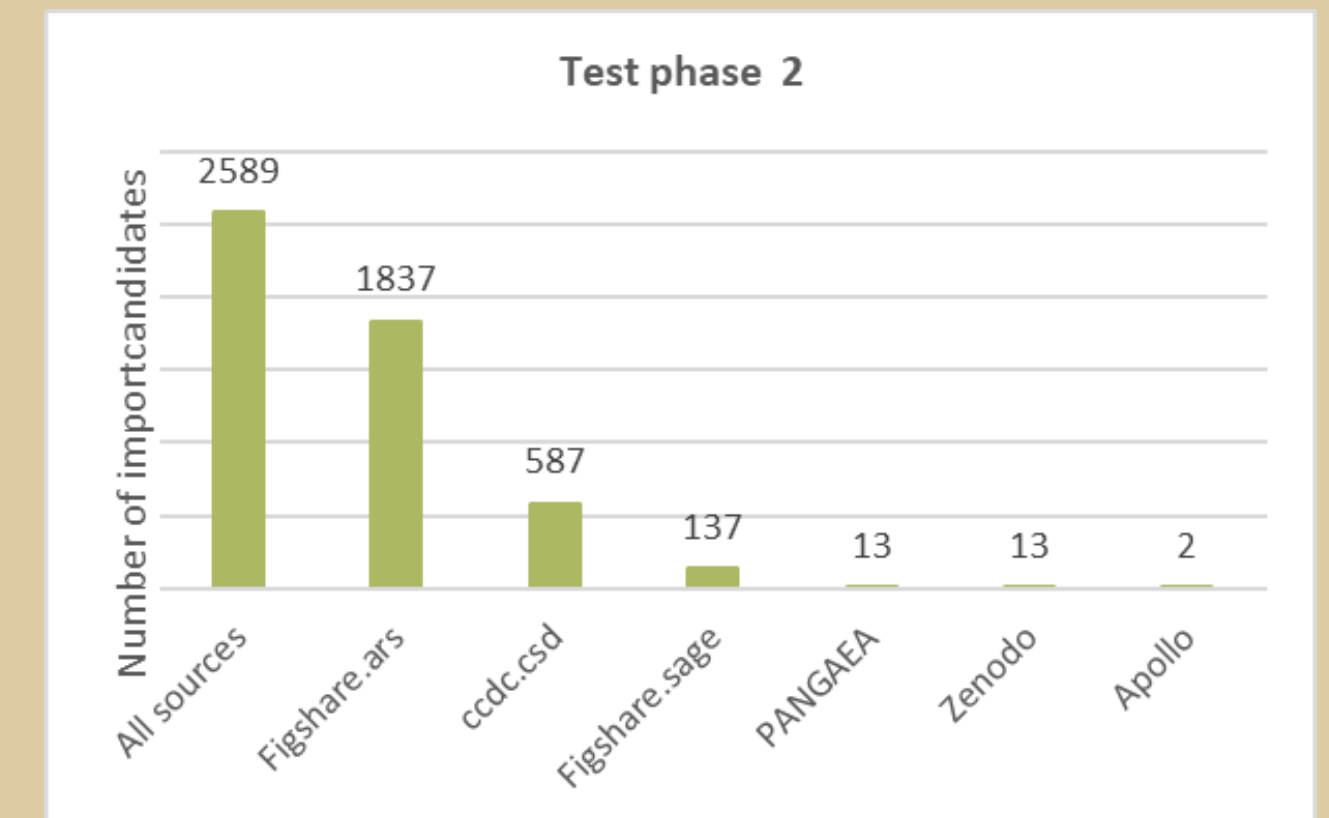
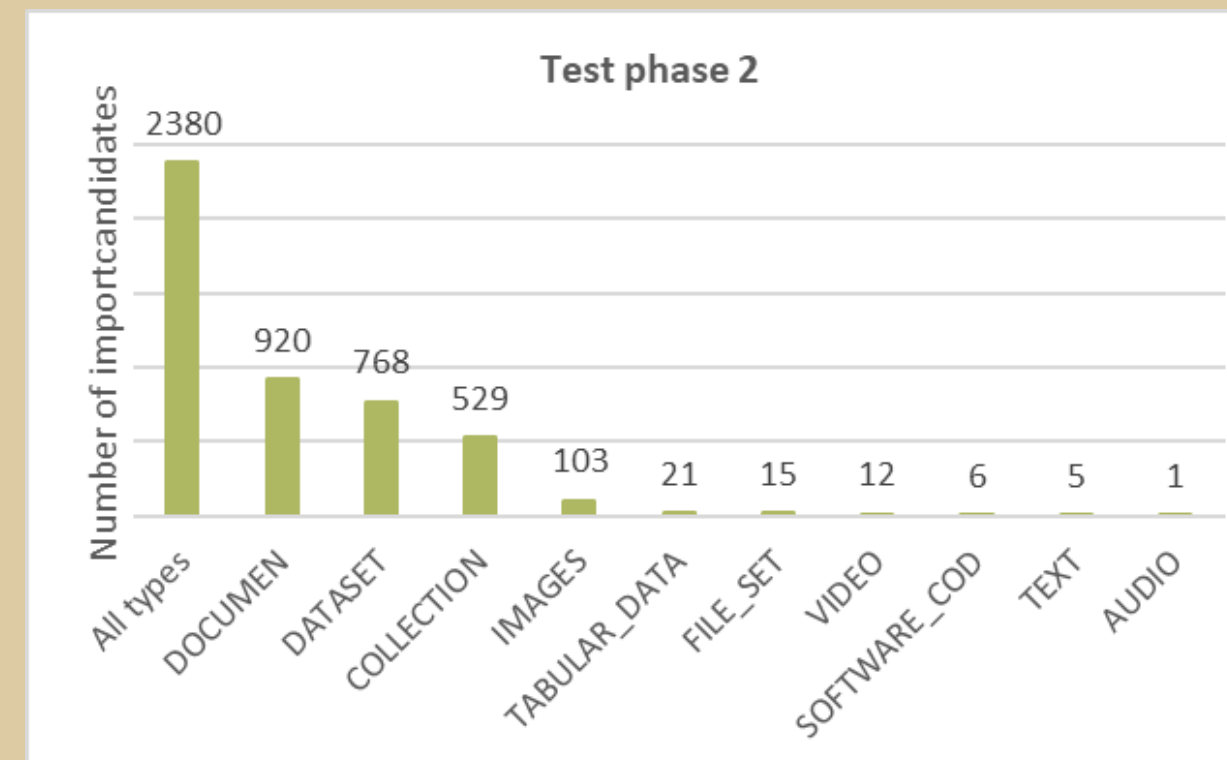
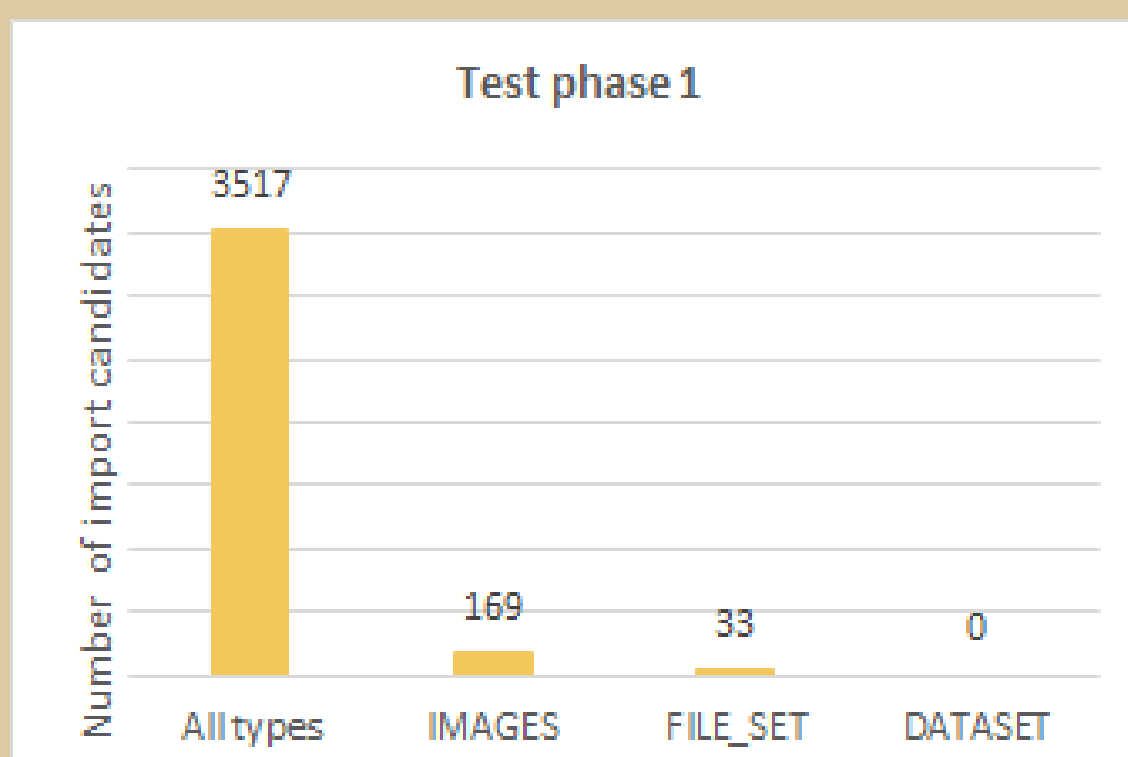
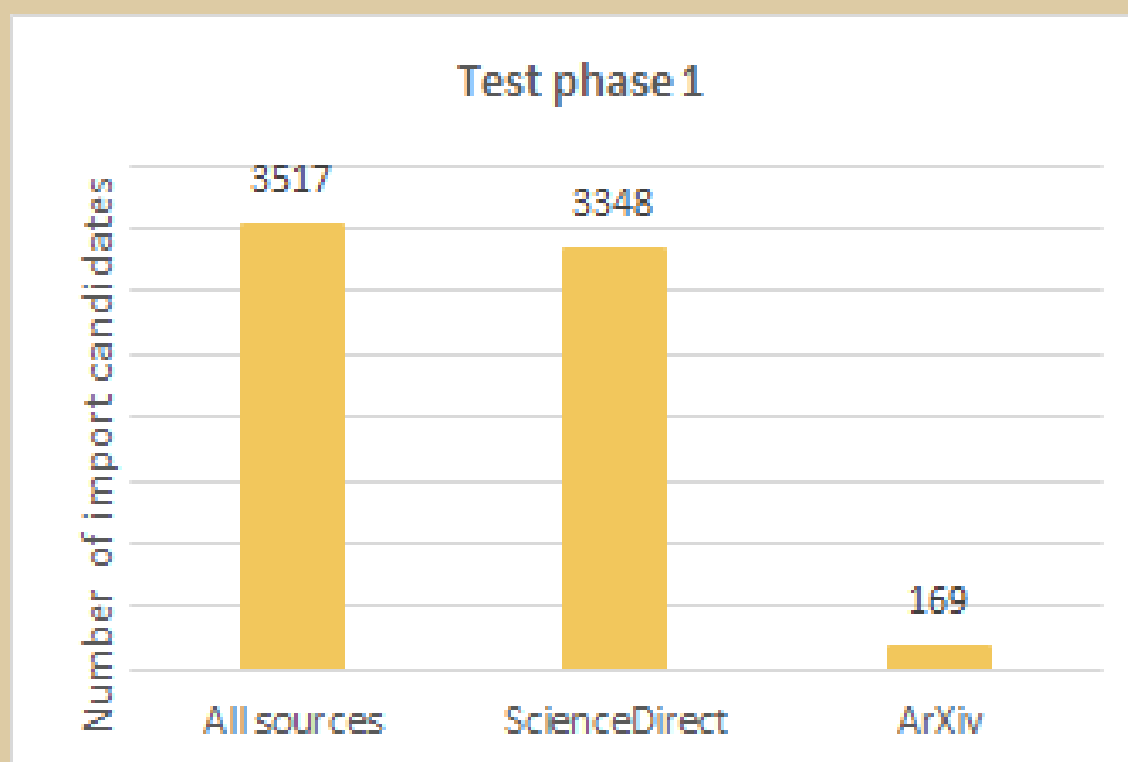
Data Monitor is a service developed by Elsevier which indexes datasets from different sources for import into CRIS systems. In the enriched version of the service datasets are compared with publication, author and affiliation data to increase the proportion of datasets attributed to an institution. In the enriched version of the service datasets are furthermore updated if changes are made in the source data. The Data Monitor-integration in Pure offers functionality to have datasets imported either manually or automatically.

Purpose

- Assess if the service can be used to identify and import datasets into Pure.
- Assess if it generates added value to use the enriched version.
- Clarify whether relationships are made between imported datasets and related publications.
- Assess the data quality of the datasets.

Method

The Data Monitor-integration test was directed by Elsevier with participants from 5 different European Universities. The test was conducted in two phases. In phase 1 the non-enriched version was tested and in phase 2 the enriched version. A series of short questions regarding the participants' experiences and results were asked during both phases of the test and online meetings were held regularly to discuss these. The test was conducted from February 2020 till the end of June 2020.



Results

Dataset identification:

- ✗ The unenriched version finds datasets from only two sources.
- ✗ The non-enriched version primarily finds datasets which are closely related to publications, i.e. pictures, graphs and tables that already appear as part of the publications.
- ✓ The enriched version finds datasets from significantly more sources.
- ✓ The enriched version finds significantly more relevant datasets including supplementary data, interview guides, and chemical formulas.

Automatic import of datasets:

- ✗ No option to edit data on datasets after import, not even contributors' organizational affiliation.
- ✗ No option to remove an associated person from a dataset after import.

Relationships between datasets and publications:

- ✓ The relationships are created in Pure.

Discussion

It is necessary to have an operational definition of datasets, since the import candidates comes in a wide range and the value of the integration depends on whether the type and source qualification options are adequate to accommodate your institutions definition of a dataset. A restrictive definition of datasets will potentially limit an institutions number of import candidates. The option to choose between automatic or manual import of datasets is positive, but the build in editing restrictions are problematic regardless of import method.

Conclusion

The enriched version of the service identifies more datasets from more sources and of more varied types than the non-enriched version and is therefore of higher value.

The weakest link in the Data Monitor-integration is of a technical nature. The integration does not allow one to freely edit fields on datasets after import, and the fact that it's not possible to remove a connected internal organization from an imported dataset is problematic in relation to data quality.

