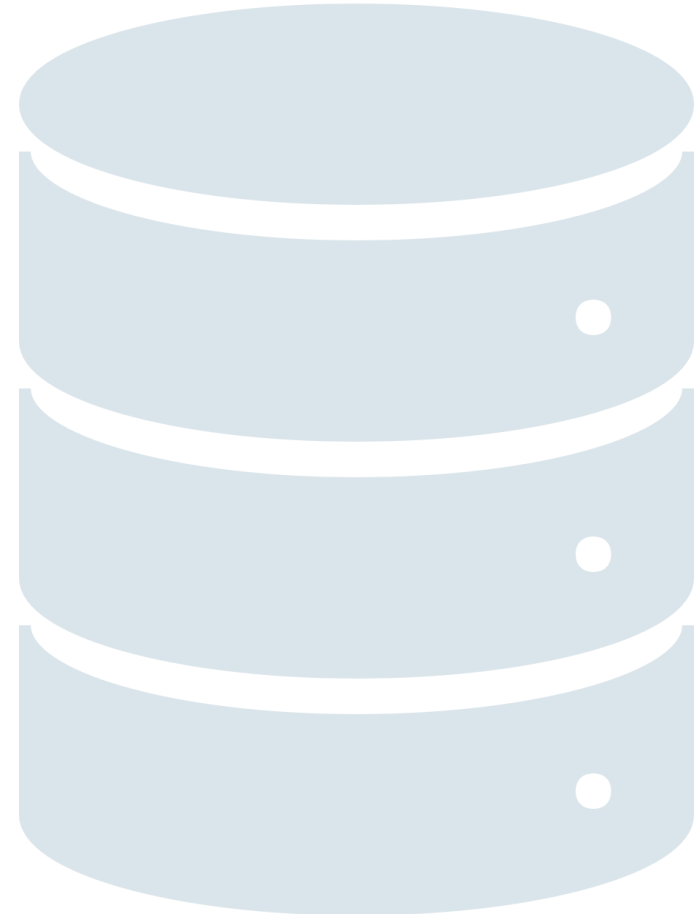




SoleCRIS Datasets

Navigating multiple import sources and
ways to identify duplicate datasets



Niilo Lahti

Product owner, SoleCRIS

niilo.lahti@solenovo.fi

Short introduction to Datasets feature



Importing

Metax

DataCite



SoleCRIS

Data enrichment

- process
- validations

Displaying datasets

Searching

Reporting

Commenting



Exporting non-Metax datasets to Metax



Identifying duplicate metadata entities

- Persistent identifier
 - Single primary identifier
 - Read-only
- Other identifier(s)
 - Secondary identifier(s)
 - Automatically generated & user can add
- Previous version identifier
 - Automatically generated & user can add
- (Title)



Persistent identifier & Other identifier

 Publications

 **Research datasets**

 Research activities

 Conferences

 Units

 Persons

 Reports

 Administration

Edit

Persistent identifier and location of research dataset

PERSISTENT IDENTIFIER

Persistent identifier *

nbn:fi:fsd:T-FSD3806

Persistent identifier type *

URN



OTHER IDENTIFIERS

Other identifier *

10.60686/t-fsd3806

Other identifier type *

DOI



Link to dataset

Previous version identifier

Home

Publications

Research datasets

Research activities

Conferences

Units

Persons

Reports

Version

4

PREVIOUS VERSIONS

Previous version

3

Previous version identifier

10.23729/dca63af1-3bad-4664-9eb4-b5124c5b1a11

Previous version

2

Previous version identifier

10.23729/d563917d-f347-4c59-9668-e72cdec9689f

Previous version

1

Previous version identifier

10.23729/6a2021ad-672c-4738-ac29-12f41b403ea3

+ Add new



Import/export scenarios

Matching with Persistent identifier

1. Metax + DataCite (other than Metax)
 2. Metax + Metax
 3. DataCite + Metax
 4. DataCite + DataCite
- Metax entities is prioritized over non-Metax (DataCite) entities.
 - Results of import scenarios are logged in integration report.

Scenario 1: Metax + DataCite

1

A Metax catalog

Metax

New import

New import
reported

SoleCRIS

No export

2

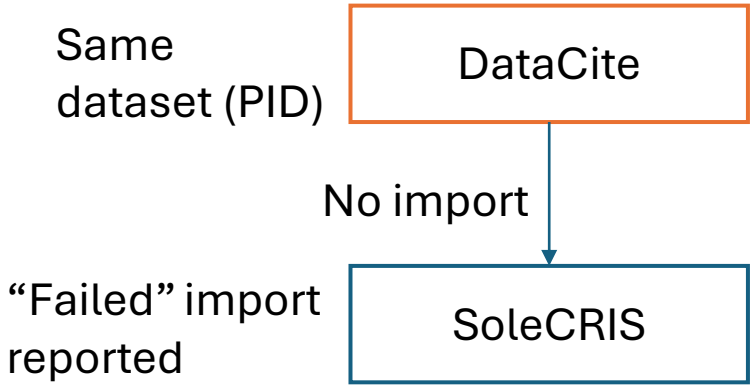
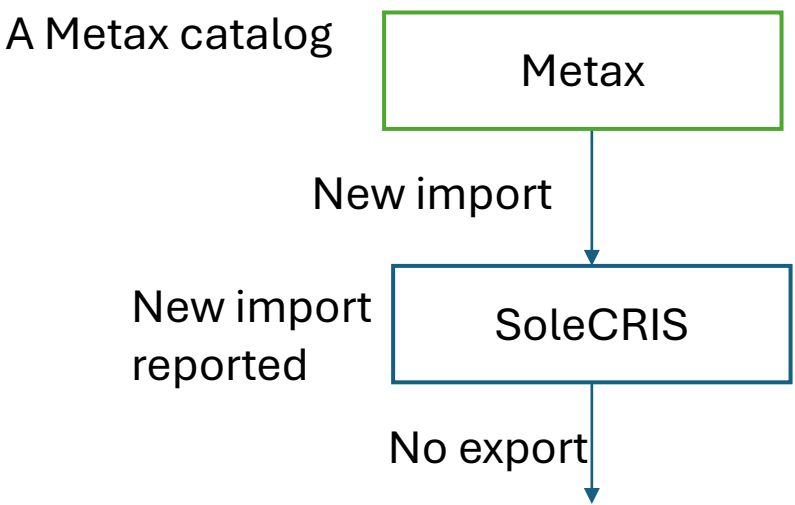
Same
dataset (PID)

DataCite

No import

“Failed” import
reported

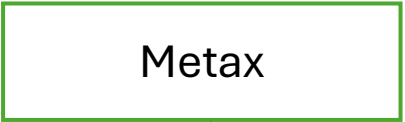
SoleCRIS



Scenario 2: Metax + Metax

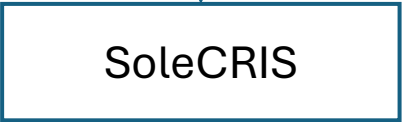
1

A Metax catalog



New import

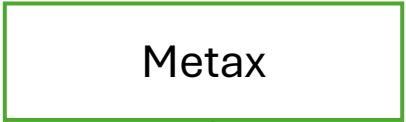
New import reported



No export

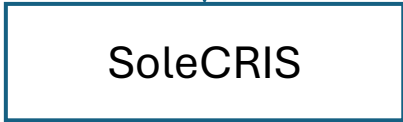
2

Same or another catalog
Same dataset (PID)



Import

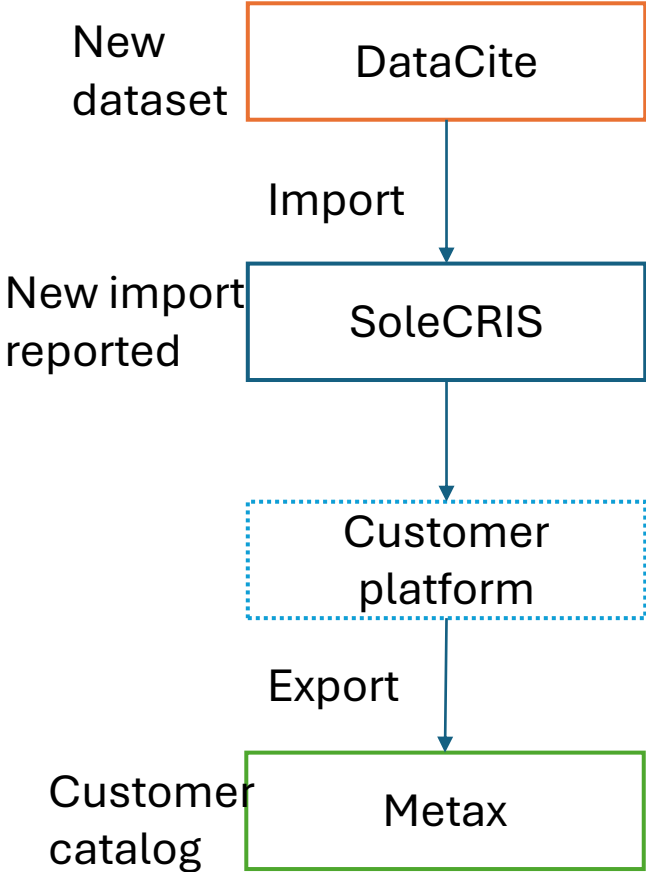
Specified fields are overwritten
Update reported



No export

Scenario 3: DataCite + Metax

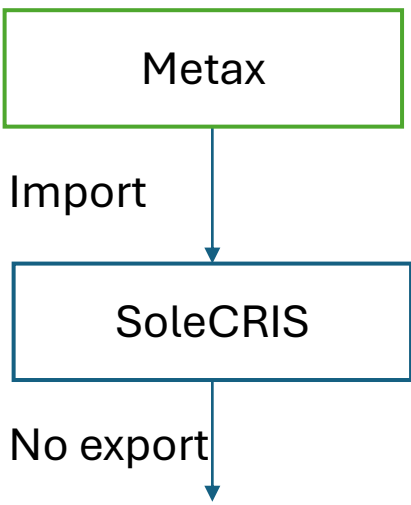
1



2

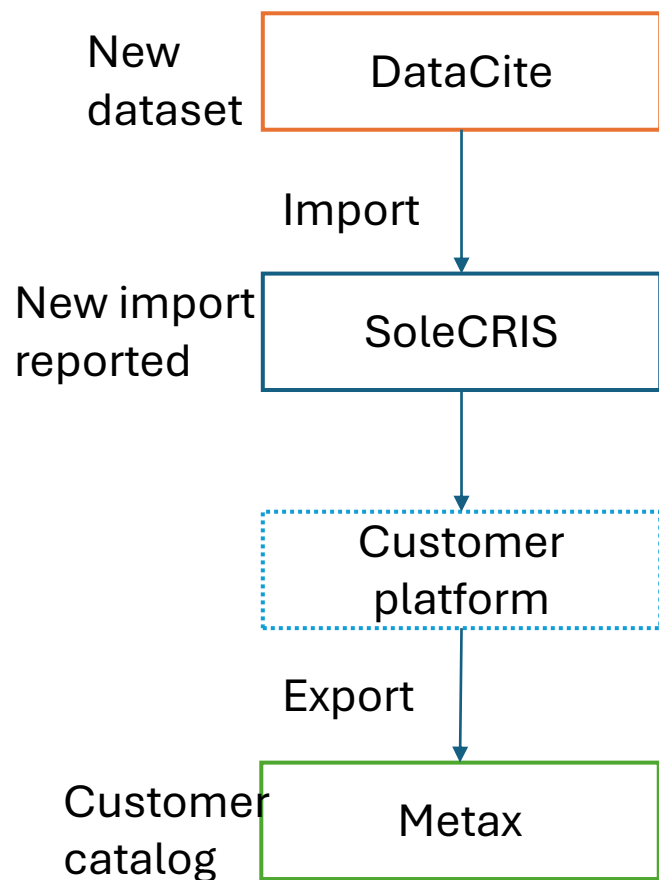
A Metax catalog
Same dataset (PID)

Specified fields
are overwritten
Update reported

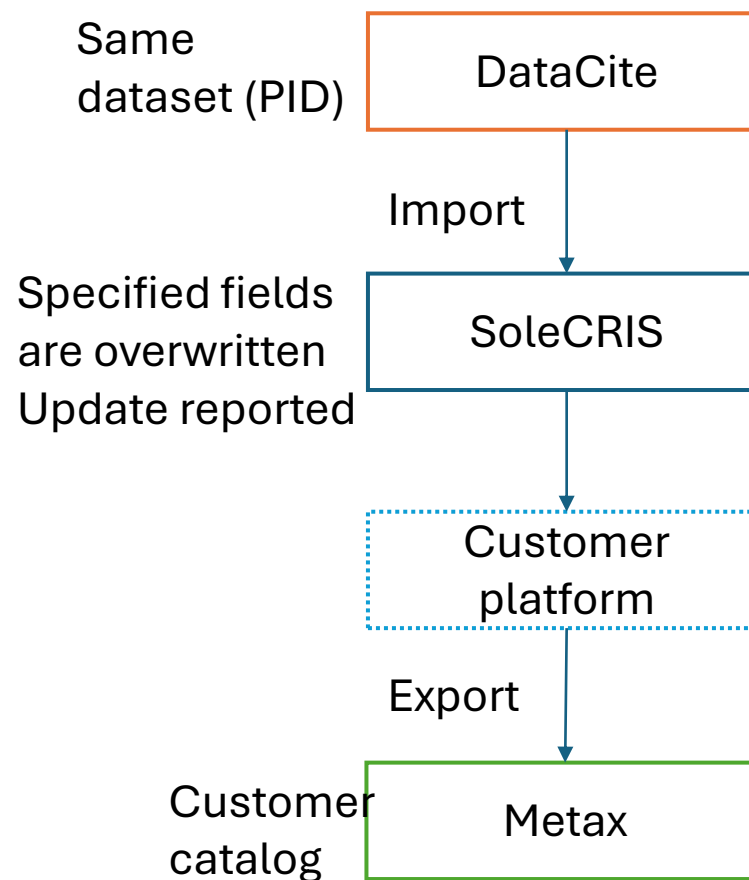


Scenario 4: DataCite + DataCite

1



2



Summary of development phase 1



- We have basic import scenarios which prioritize Metax entities and limit double datasets to a degree
- Admin can monitor these scenarios from integration report (new, updated, failed)
- In some scenarios, data is updated without explicit admin consent
- Limited production use will inform next development



Next developments & vision of “final” feature

- Double datasets are better identified with
 - Other identifiers
 - Previous version identifiers
 - Title
- → Expanded integration report
- User can compare and has the final verdict what is updated and by what source



Thank you!

Niilo Lahti

Product owner, SoleCRIS

niilo.lahti@solenovo.fi

