



Vlaamse
overheid

Archiving SharePoint data

**Dealing with challenges and
avoiding them where possible**

Koen Dobbelaere and Quincy Oeyen
Digital Archives Flanders

About us

Employer

Flemish government – The Agency for Facility Operations –
Digital Archives Flanders

Quincy Oeyen

Product manager Digital Archives Flanders



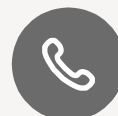
0032 498 27 56 37



quincy.oeyen@vlaanderen.be

Koen Dobbelaere

Digital archivist Digital Archives Flanders



0032 473 36 13 19



koen.dobbelaere@vlaanderen.be

The Agency for Facility Operations

Cluster Information Management

POLICY MAKER

Monitoring and creating legislation

PROCESS OPTIMISATION & DIGITISATION

Smart digitisation, towards paperless offices in the Flemish government

CENTRE OF EXPERTISE

From basic to advanced training, consultancy



DIGITAL ARCHIVES FLANDERS

Managing and preserving digital records

ANALOGUE ARCHIVE FOR THE FLEMISH GOVERNMENT

Managing and preserving the analogue archive of the Flemish government

FRAMEWORK CONTRACTS

Digital Archives Flanders

Shared service for managing and preserving digital records



Flemish government

- ✓ Agencies
- ✓ Departments
- ✓ ...



Local governments

- ✓ Provinces
- ✓ Municipalities
- ✓ ...



Bodies governed by public law

- ✓ Art. I. 3, 6° of the Administrative Decree

Series – Dossier – Items

An aggregation of files created and maintained by an owner that is in the same identifiable sequence, or result of the same accumulation or filing process and is of similar function, format or informational content

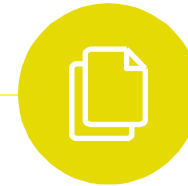


Series

Alle medical files in 1 hospital

Dossier

A medical file/folder about a specific case



A sequence of items which makes up an organizational/business activity.

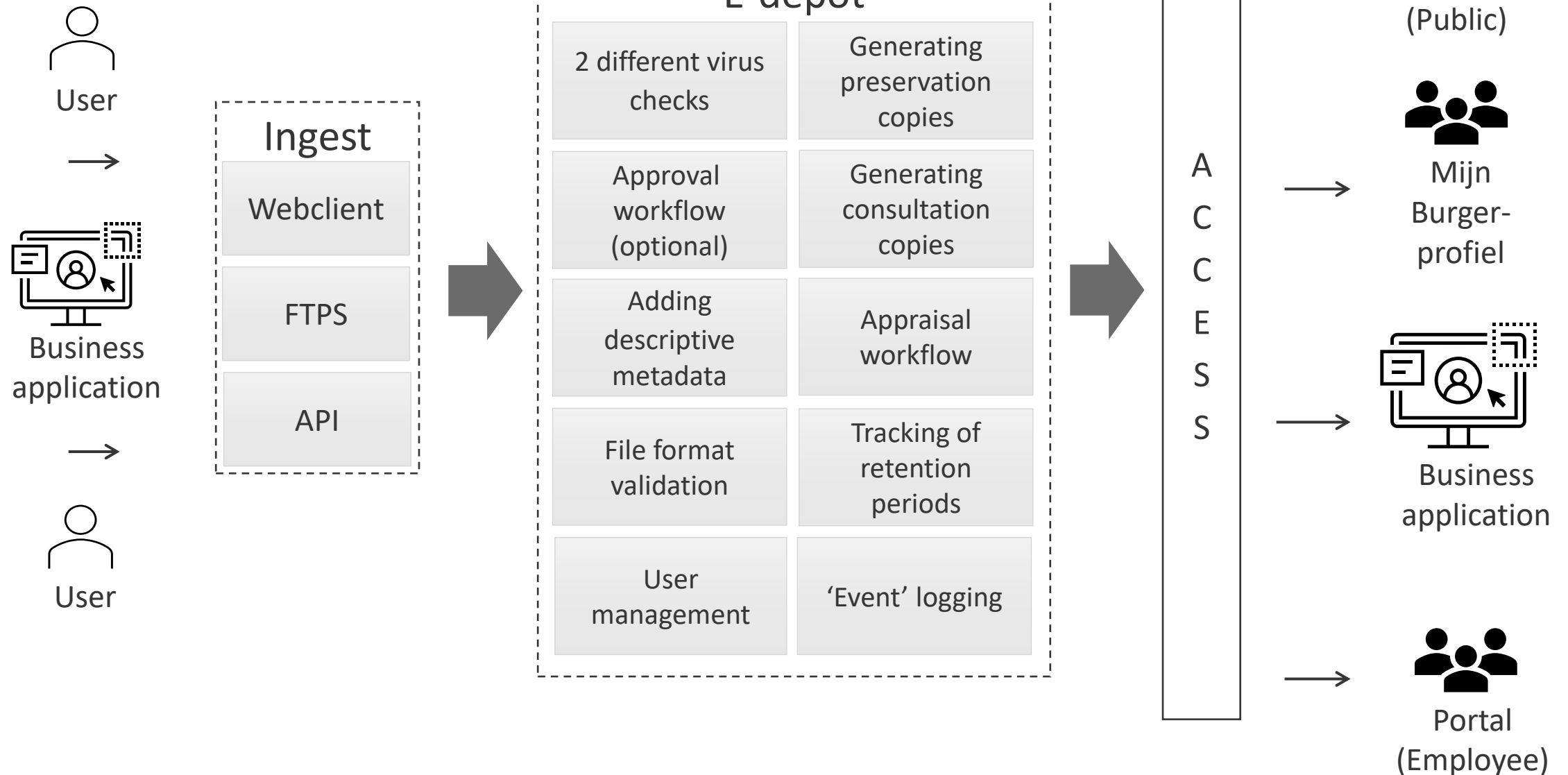
The smallest unit of records managed as an entity



Items

An e-mail containing a referral

High-level figure



Metadata

- Descriptive metadata
 - Purpose: retrieve information
- Technical metadata
 - Automatically extracted during ingest and processing
- Administrative metadata
 - Metadata inherited from the series
 - Exceptions are possible
- (meta) Data model is:
 - Basic but can be expanded or extended per serie
 - Based on standards (e.g., PREMIS, Dublin Core ...)

29_09_2016

Vergaderdossiers (Demo Series 2) Folder Published 18 Jun 2020 - 18 Jun 2020 - 18 Jun 2031

Contents **Details** Relationships Beveiligingsgroepen History

Identification		Series Identification	
Object Name	29_09_2016	Series Name	Vergaderdossiers (Demo Series 2)
Item description	-	Start	-
Openingsdatum	18 Jun 2020	Finish	-
Closing date	18 Jun 2020	Service	-
Unique ID	54318db89f3a44639aa78cc5414659fbf11269067bcf47018262cf3b07e28d90	Process	Meetings Meetings of selection committees Meetings of the Flemish Government Selection Committee
URI	https://digitaalarchief-ti.vlaanderen.be/id/dossier/54318db89f3a44639aa78cc5414659fbf11269067bcf47018262cf3b07e28d90	Bewaargever	Agency for Facility Services
Dossier code of the source application	-		

Details **Extensies** Relaties Versies Beveiligingsgroepen

Geef zoekterm in...

Extensie	Dossier	Stuk	Actief
AdresExtensie	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Dossier - Afzender	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Dossier - Behandelaar	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Dossier - Bestemming	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
VoorbeeldExtensie	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
VoorbeeldExtensieBoolean	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
VoorbeeldExtensieDate	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
VoorbeeldExtensieEnum	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
VoorbeeldExtensieLijst	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
VoorbeeldExtensieNummer	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>

VoorbeeldExtensie

Attribuut	Type	Verplicht
Tekst	Tekst	<input checked="" type="checkbox"/>
Nummer	Nummer	<input checked="" type="checkbox"/>
Boolean	Ja/Nee	<input checked="" type="checkbox"/>
Datum	Datum	<input type="checkbox"/>
Enumeration	Keuzelijst	<input type="checkbox"/>
Lijst	Lijst	<input checked="" type="checkbox"/>

SharePoint and E-depot: why?

- ▶ Everyone uses SharePoint (on prem, cloud)
- ▶ Our philosophy
 - As close to the administration as possible
 - Archive immediately when the dossier is closed
- ▶ We are not the 'typical' archive
 - We destroy records after the retention period
 - Some RM functionalities (i.e. legal logging)

Scenarios in SharePoint

- ▶ Library containing folders
→ 1 folder = 1 file
- ▶ Library containing document sets
→ 1 document set = 1 file
- ▶ Library containing documents structured by metadata, no folders
→ A file is defined by metadata

	2017 Herziening leidraad lokale besturen	Afgehandeld	1-1-2017	31-12-2017		15 april 2021
	2017 Herman Teirlinck WG informatiebeheer	Afgehandeld	1-1-2017	31-12-2017		15 april 2021
	2017 Ferraris	Afgehandeld	1-1-2017	31-12-2017		15 april 2021
	2017 Dienstencentra	Afgehandeld	1-1-2017	31-12-2017		15 april 2021
	2017 Code of conduct	Afgehandeld	1-1-2017	31-12-2017	Beleid	29 september 2021
	2017 Bestelbrief boekenluis	Afgehandeld	1-1-2017	31-12-2017		15 april 2021
	2017 Beslissingsmodel Hergebruik	Afgehandeld	1-1-2017	31-12-2017	Beleid	29 september 2021

	OVO000029 - Hoofdovereenkomst.pdf	Departement Financiën en	OVO000029		Hoofdovereenkomst		
	OVO000029 - Bestelformulier Documentenkuis.pdf	Departement Financiën en	OVO000029	Documentenkuis	Bestelformulier - Documentenkuis	3-5-2023	2023
	OVO000032 - Hoofdovereenkomst.pdf	Departement Kanselarij en	OVO000032		Hoofdovereenkomst	31-3-2022	2022
	OVO000032 - Bestelformulier - Serieregister.pdf	Departement Kanselarij en	OVO000032	Serieregister	Bestelformulier - Serieregister	31-3-2022	2022
	OVO000082 - Toetredings- en verwerkersovereenkomst DAV - Serieregister.pdf	Departement Landbouw e	OVO000082	Serieregister		23-6-2020	2020
	OVO000082 - Bestelformulier - Serieregister.pdf	Departement Landbouw e	OVO000082	Serieregister	Bestelformulier - Serieregister	14-2-2022	2020
	OVO000082 - DAV - Hoofdovereenkomst.pdf	Departement Landbouw e	OVO000082		Hoofdovereenkomst	14-2-2022	2020
	OVO000096 - Toetredings- en verwerkersovereenkomst DAV - E-depot.pdf	Departement Mobiliteit en	OVO000096	E-depot		28-10-2020	2020
	OVO000096 - Hoofdovereenkomst.pdf	Departement Mobiliteit en	OVO000096		Hoofdovereenkomst	7-2-2023	2020
	OVO000096 - Bestelformulier - Serieregister.pdf	Departement Mobiliteit en	OVO000096	Serieregister	Bestelformulier - Serieregister	7-2-2023	2020
	OVO000096 - Bestelformulier - E-depot.pdf	Departement Mobiliteit en	OVO000096	E-depot	Bestelformulier - E-depot	7-2-2023	2020
	OVO000049 - Toetredings- en verwerkersovereenkomst DAV - Serieregister.pdf	Departement Onderwijs er	OVO000049	Serieregister		18-9-2020	2020

What do we want? – Utopia

- ▶ 'Automated' integration of E-depot and SharePoint based on our API
 - Don't want to import a 'bulk' SIP
 - 'Create file' action in the archive where each dossier is created individually and the items are added
- ▶ 2 workflows:
 - Trigger-based (time or event-based) archiving
 - 'Click on button' archiving
- ▶ Anyone can decide to archive a file at any time

What do we have? - *Metadata Splitter & Collector*

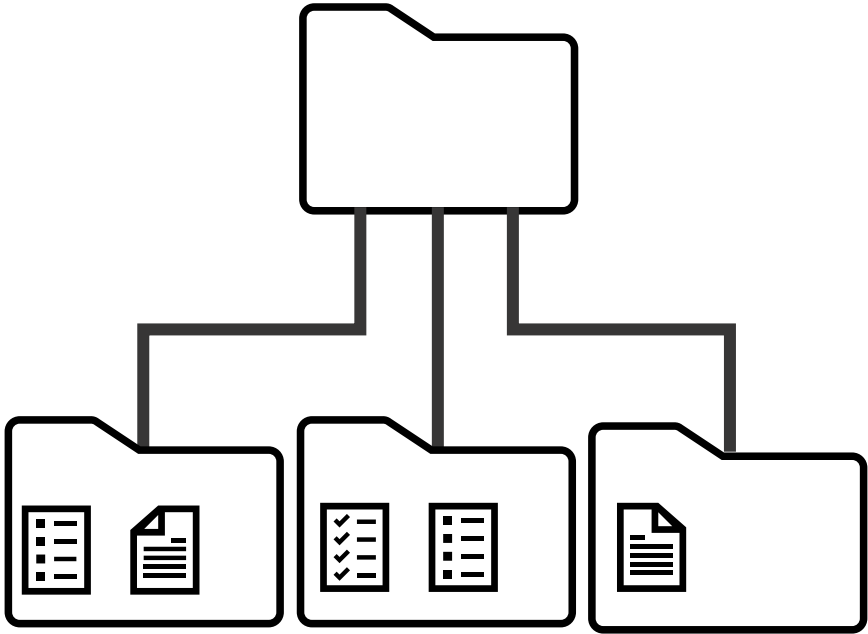
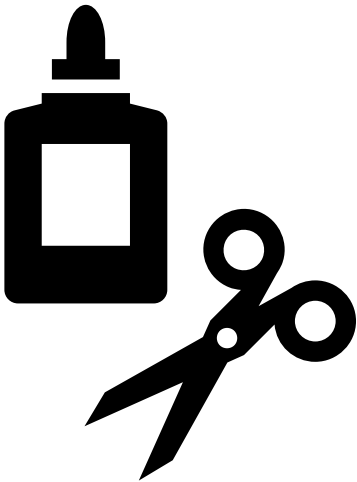
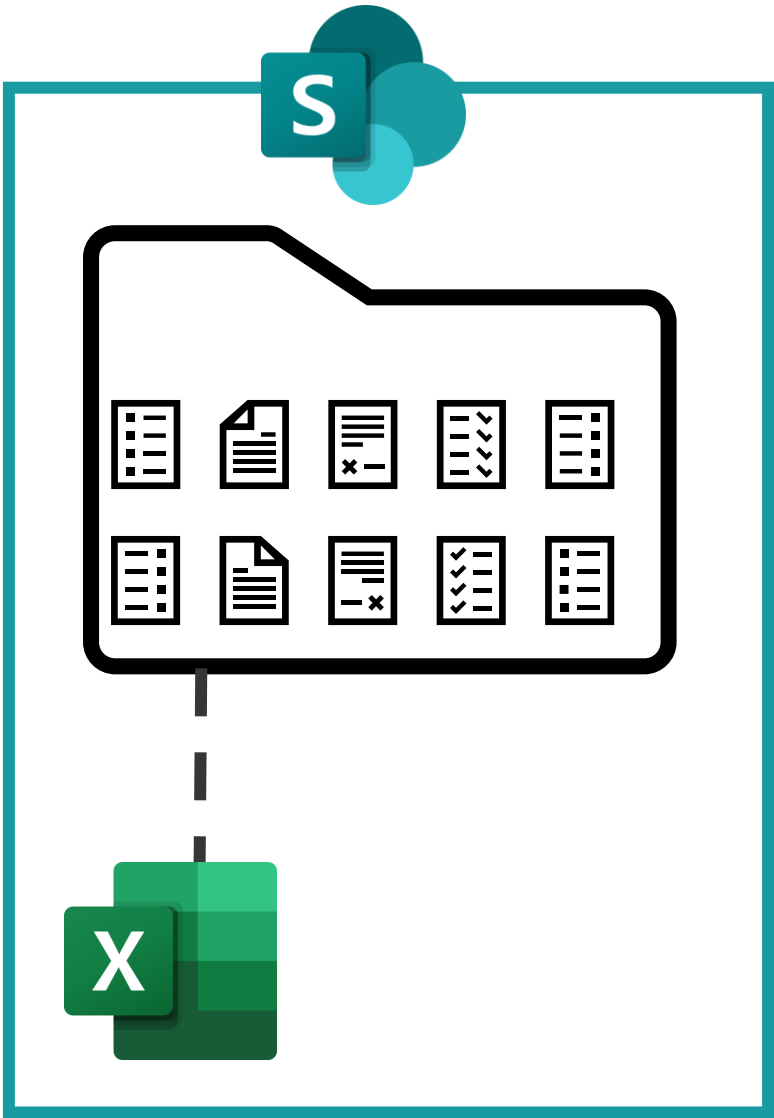
- ▶ A 'labour-intensive' work-around called "*Metadata Splitter & Collector*"
- ▶ Based on an export from SharePoint (data + metadata) creating a SIP that is E-depot compliant
 - We have built a (basic) tool that ensures we can
 - 1) SPLIT: restructure the exported data in a new folder structure based on the metadata.
 - 2) COLLECT: fill in a metadata Excel file based on technical metadata from the files from the selected folder and the folder structure itself
 - In both scenarios, the tool can map user created SharePoint metadata to our data model

[Download the Metadata Splitter & Collector](#)

Download

→ Split

→ New structure

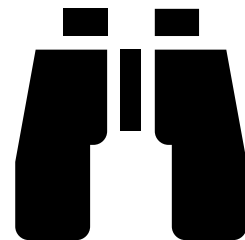
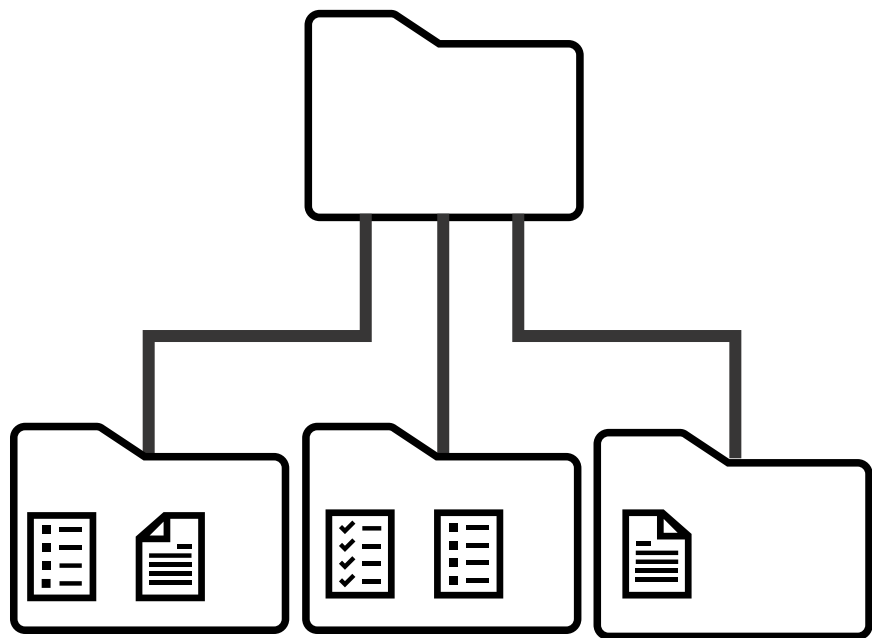


Folder structure

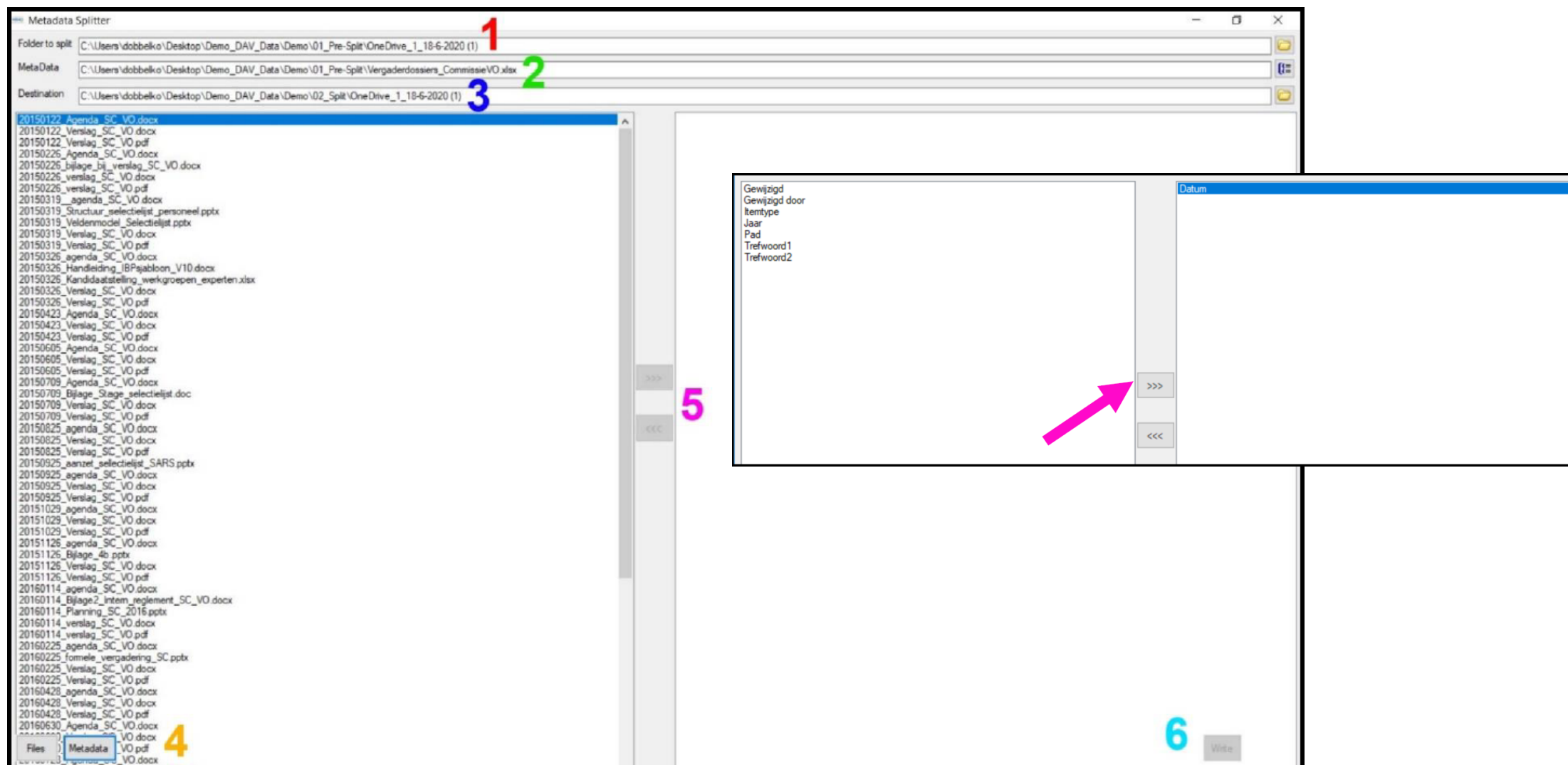


Collect

→ Metadata
file



Transform data that is structured based on metadata to our SIP requirements



Path in SIP	Type	DossierRef	Naam
Dossier_A	dossier	Dossier_A	Dossier_A
Dossier_A/1 Agenda/20150605_Agenda.docx	stuk	Dossier_A	20150605_Agenda.docx
Dossier_A/2 Verslag/20150605_Verslag.docx	stuk	Dossier_A	20150605_Verslag.docx
Dossier_A/3 Definitief verslag/20150605_Verslag.pdf	stuk	Dossier_A	20150605_Verslag.pdf
Dossier_B	dossier	Dossier_B	Dossier_B
Dossier_B/1 Agenda/20150709_Agenda.docx	stuk	Dossier_B	20150709_Agenda.docx
Dossier_B/2 Verslag/20150709_Verslag.docx	stuk	Dossier_B	20150709_Verslag.docx
Dossier_B/3 Definitief verslag/20150709_Verslag.pdf	stuk	Dossier_B	20150709_Verslag.pdf
Dossier_B/4 Bijlage/20150709_Bijlage_Stage.doc	stuk	Dossier_B	20150709_Bijlage_Stage.doc
Dossier_C	dossier	Dossier_C	Dossier_C
Dossier_C/1 Agenda/20160114_agenda.docx	stuk	Dossier_C	20160114_agenda.docx
Dossier_C/2 Verslag/20160114_verslag.docx	stuk	Dossier_C	20160114_verslag.docx
Dossier_C/3 Definitief verslag/20160114_verslag.pdf	stuk	Dossier_C	20160114_verslag.pdf
Dossier_C/4 Bijlage/20160114_Planning.pptx	stuk	Dossier_C	20160114_Planning.pptx
Dossier_D	dossier	Dossier_D	Dossier_D
Dossier_D/1 Agenda/20170915_Agenda.pptx	stuk	Dossier_D	20170915_Agenda.pptx

Road to Utopia: cobblestones aka challenges

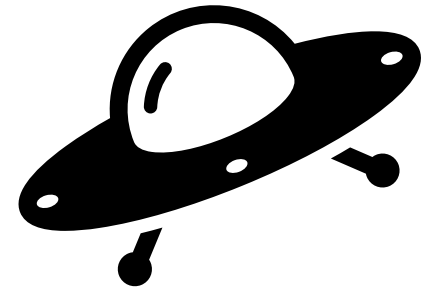
- ▶ Mapping metadata
 - How to level metadata fields (easily)?
 - × Dossier
 - × Document
 - A new mapping should be made for each library, as the metadata may differ
- ▶ A user-friendly transfer environment
 - User feedback
 - Error handling
- ▶ One-size-fits-all solution
 - Cover all user needs

Same problem, 2 solutions: pros and cons

Solution	Pros	Cons
Export and import (as is)	<ul style="list-style-type: none">• Easy to use• Low maintenance• Many improvements possible• More versatile	<ul style="list-style-type: none">• Labour intensive• Not 'sexy'• Error-sensitive (support)
Integration SharePoint and E-depot	<ul style="list-style-type: none">• Full automation• 'Sexy', looks good when working	<ul style="list-style-type: none">• High development and maintenance cost• Mapping is very labor intensive• Challenge to create a generic solution (multi tenant)• Stability• strict

Future plans (2024-...)

- ▶ Decide which solution
 - The 'labor-intensive' solution
 - The utopia solution
- ▶ Elements that will determine the decision
 - Cost (esp. the cost of upkeep)
 - UX and UI (usability)





INTERNATIONAL
CONFERENCE
ON DIGITAL
PRESERVATION
Ghent & Flanders, Belgium

[iPRES 2024 \(pubpub.org\)](https://pubpub.org)



INTERNATIONAL CONFERENCE
ON DIGITAL PRESERVATION
Ghent & Flanders, Belgium