

Flemish Heritage Database

Annex 3: data analysis

1.	Reading guide	1
2.	Data model objective.....	3
3.	Explanation of current solutions	5
3.1.	<i>Erfgoedplus.....</i>	5
3.1.1.	General.....	5
3.1.2.	Authorities.....	Fout! Bladwijzer niet gedefinieerd.
3.1.3.	Erfgoedregister registration module.....	8
3.1.4.	Digital Asset Management.....	9
3.1.5.	Fuseki triple store (harvesting)	9
3.1.6.	Public portal	9
3.2.	<i>Erfgoedinzicht.....</i>	9
3.2.1.	General.....	9
3.2.2.	Registration module.....	10
3.2.3.	Digital Asset Management.....	14
3.2.4.	Public portal	15
4.	Data model	15
4.1.	<i>General.....</i>	15
4.2.	<i>Set-up.....</i>	15
4.3.	<i>Expectations for future data model and management.....</i>	16
4.4.	<i>Explanation of the data analysis, including possible optimisations</i>	17
4.4.1.	Account and access profiles.....	17
4.4.2.	Provenance and audit trail	17
4.4.3.	Controlled value lists	17
4.4.4.	Collections	24
4.4.5.	Heritage objects	24
4.4.6.	Archives.....	25
4.4.7.	Publications	26
4.4.8.	DAM/meemoo archive.....	26
5.	Migration/start-up requirements	26

1. Reading guide

This annex is based on intensive data analysis of current data in the various environments of Erfgoedinzicht and Erfgoedregister. The results of this analysis can be found in this annex and in the file see Annex 6: data samples\mapping.xls.

Key data from the current solutions has been mapped to widely known standards in this file, with

- heritage objects and archaeological finds and their associated management processes mapped to [Spectrum 5.0](#), except for some information groups where this mapping needs additional explanation or differs from the models proposed in Spectrum 5.0;
- archive items have been mapped to a combination of [ISAD\(G\) 2nd ed.](#) and Spectrum 5, where [this documentation](#) can be used;
- publications have been mapped to [MARC21](#).

Column	Explanation
The columns under the entry book header refer to guidelines used by registrars when registering, for both Erfgoedplus and Erfgoedinzicht.	Not all fields are described in the current solutions. Deviations from the guidelines can often be observed in the data.
Spectrum procedure	The Spectrum procedure with which the fields from Erfgoedplus and Erfgoedinzicht can be mapped
Spectrum information group	The Spectrum information group with which the fields from Erfgoedplus and Erfgoedinzicht can be mapped
Spectrum information unit	<p>The Spectrum information unit with which the individual fields from Erfgoedplus and Erfgoedinzicht can be mapped</p> <p>When a field cannot be mapped to Spectrum, the column is read as follows:</p> <ul style="list-style-type: none"> • + [current tag: contractor may rename if desired]: to add to the information group • Legacy: to be included as a searchable record, which may be processed later, of this record (by including the necessary tags, IDs and content)
Spectrum additional information unit	Any additional Spectrum information units such as dates
Inherit	Data that must be 'inherited' by default when this Spectrum information group is displayed/linked in another intellectual entity (e.g. always show the title of a heritage object in input screens for loans)
Erfgoedinzicht - Tag	Current tag (including path) when requesting records via the Erfgoedinzicht API

Erfgoedinzicht - Multi	Multiple occurrences possible Y/N
Erfgoedinzicht - Type	Data type
Erfgoedregister - Tag export -	Current tag (including path) in the Erfgoedregister export
Erfgoedinzicht - Multi	Multiple occurrences possible Y/N
Erfgoedinzicht - Collection	Mapping on COMETA Group and Element
Erfgoedinzicht - Archives	Overview of the fields used in the archive module

Some complex exceptions are explained in more detail below.

For controlled value lists, such as agents, concepts, etc. this annex refers to known ontologies whenever possible.

Meemoo assumes that these descriptions enable the contractor to execute a realistic migration plan within the set deadlines.

2. Data model objective

As indicated, the tenderer may propose alternative standards, except for Spectrum. The tenderer must, however, provide sound justification for this.

In doing so, it is important that:

- the tenderer demonstrates how the following objectives can be achieved, whereby best practices for modelling data will be evaluated:
 - Accuracy and completeness: existing data can be migrated correctly and completely;
 - Usability: the data model must be (re)usable and easily accessible to those who need it. It also needs to be user-friendly and enable users to find, understand and use the data for its intended applications;
 - Consistency: the data model must be consistent and enable unambiguous meaning and interpretation of the data using standardised terminology, definitions and relationships between the data;
 - Flexibility: the data model must be flexible and able to adapt transparently to changing requirements and environments;
 - Traceability: the origin, evolution and changes of the data must be traceable;
- the tenderer must assume that, within the deadlines set, only minimal data cleansing by end users is possible. This means that:
 - because of the strict deadlines, the migration process will mainly focus on migrating a limited number of data models to the solution's final model, with limited qualitative evaluation of the data to be migrated;
 - a few variants of the migration procedures need to be implemented for the different existing solutions and tenants;
- the tenderer explains how meemoo can contribute to achieving the objectives described above, with a view to an efficient migration process in which the tenderer bears responsibility for:

- at least the technical implementation, going from the database engine, the data model implemented, naming of fields, etc. rests entirely with the contractor;
 - meemoo's analysts can, based on the data analysis and best-practices, (further) prepare scenarios in collaboration with the contractor;
 - this annex already provides a non-exhaustive set of observations and suggestions, with meemoo encouraging realistic and high-quality alternative solutions formulated by the contractor;
- meemoo ultimately wants to allow registrars to apply the [FAIR principles](#) where possible and has already included a limited set of requirements for this. Meemoo itself is not the owner of the data and content, but primarily a partner that supports organizations in implementing the FAIR principles. Meemoo wants to rely for further steps on a contractor that can offer expertise and support to apply these principles in a Linked Open Data strategy to be developed. A shared solid and transparent data model are a first necessary step for this.

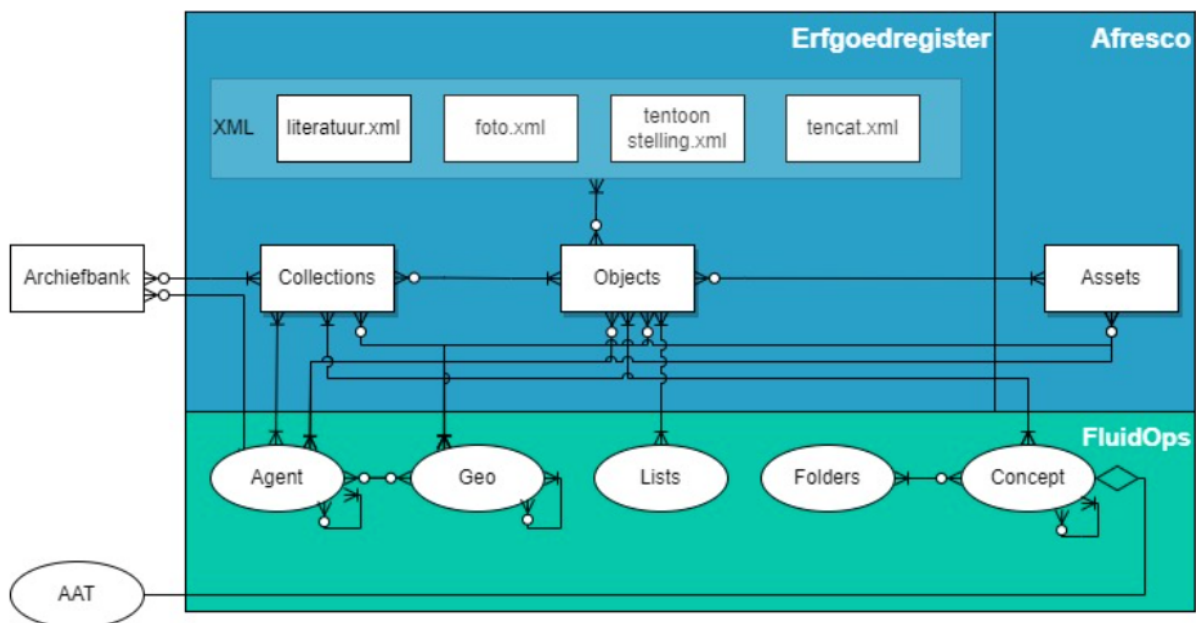
3. Explanation of current solutions

3.1. Erfgoedplus

3.1.1. General

Erfgoedplus collects information on heritage. All heritage that is registered the database is described in a similar, standardised way, and linked to each other. This makes Erfgoedplus a hub of heritage objects and their data.

Erfgoedplus is currently being used by 1,300 collection owners (mainly church administrations and local history societies) supported by service provision organisations with input from (voluntary) registrars and custom training, among other things.



Erfgoedplus contains a number of components located in different environments:

- authorities:
 - on the Achiefpunt environment, specifically Archivebench;
 - FluidOps workbench;
 - a number of additional XML files;
- a registration module, het Erfgoedregister, which end users interact with;
- Digital Asset Management in an Alfresco environment;
- an aggregation platform in a Fuseki triple store and publishing platform, which are not relevant in the context of migration.

We explain these in more detail in the following sections.

To help with the registration, authorities are used that are located both within the whole of the Erfgoedplus environment and on external platforms.

3.1.1.1. Archiefbank

3.1.1.1.1. General

Collections are created on Archiefbank that can currently be used as an external authority in Erfgoedplus. A number of agents are also registered that take up roles in managing the collections for this. This data has only been partially imported into Erfgoedplus.

The management of this data from Archiefpunt is being discontinued, however. The contractor will have to include this data in the data model in accordance with the guidelines, after which it will not be managed within the framework of the heritage database.

As indicated in Annex 1: Requirements are optionally requested to be able to register archives with Archiefpunt, but this is beyond the scope of this annex.

3.1.1.1.2. Collections

A separate AE_(Meemoo)_xxx.xml file is provided for each collection used in the Erfgoedregister, where the 'xxx' in the file name corresponds to the tag <ID>.

Tag	Comments
<ROWSET> <ROW num="1">	Each XML file contains only 1 ROW
<ROW num="1"> <ID>2875</ID>	ID = name of the XML file
<BWPL_ID>542</BWPL_ID>	Refers to the XML file BWPL_Meemoo.542.xml
<BWPL_XML_LINK> https://abv.icts.kuleuven.be/pls/abv/abv_erfgoedplus.get_xml_bwpl?p_id=542 </BWPL_XML_LINK>	The links in these files are often outdated and may be ignored.

There are no references to the IDs of collections in the Erfgoedregister. In Annex 6: data samples\mapping.xls we describe in detail which tags are relevant and how they map to the data from the existing environments described in this appendix.

3.1.1.1.3. Agents

For agents taking on a collection management role in Archiefbank, a separate BWPL_Meemoo.xxx.xml file is provided for each collection used in the Erfgoedregister, with the file name corresponding to the tag <ID>.

Tag	Comments
<BEWAARPLAATS>	Each XML file contains only 1 ROW
<ID>542</ID>	ID = name of the XML file
<IDENTIFICATIE_BLOCK_ROW num="1"/>	This element occurs only once.
<CONTACTPERSOON_BLOCK_ROW num="1"/>	This element can occur multiple times.

In the mapping table and below, we describe in detail which tags are relevant and how they map to the data from the existing environments described in this appendix.

3.1.1.2. FluidOps Information Workbench

The FluidOps Information Workbench is no longer supported, but contains a SPARQL endpoint that provides access to all data. This Workbench has an underlying RDF database [Sesame triple store](#). This triple store currently contains 22,952,507 triples. Sample data can be found in Annex 6: data samples\FluidOps.

In many cases, these data involve imports from other systems: for example, the class [Concept](#) is a (dated) near-copy of the [AAT](#).

The file (see Annex 6: data samples\Erfgoedplus\FluidOps\FluidOps schema.xlsx) lists all rdf types, their properties and the prefixes used.

Not all triples need to be migrated. We clarify later which triples do need to be migrated.

3.1.1.3. XML files

The Erfgoedregister also uses a number of XML files.

File	Content	Comments	#rec
Tentoonstelling.xml	Exhibitions	<Tentoonstelling xmlns:ext=" http://www.pcce.be/egb/spil-schema/extensions "> <object> <id> http://www.pcce.be/egb/event#KF.tentoonstelling.1 </id> <naam>Vlaamse Toeristenbond, Hasselt, september 1969</naam> </object>	110
tentcat.xml	Exhibition catalogues	<TentoonstellingsCataloog xmlns:ext=" http://www.pcce.be/egb/spil-schema/extensions "> <object> <id> http://www.pcce.be/egb/document#KF.tentcat.6 </id> <objectnr>6</objectnr> <naam>Leuven, 1959</naam> </object>	38
foto.xml	URLs to reference photos	<Foto xmlns:ext=" http://www.pcce.be/egb/spil-schema/extensions "> <object> <id> http://www.pcce.be/egb/photo#KF.reproductie.1 </id> <objectnr>A 61961</objectnr> <date>1944</date> <herkomst>KIK</herkomst> <uri> http://www.kikirpa.be/image/A/6/1/a61961_std.jpg </uri> </object>	9567
literatuur.xml	literature	<Literatuur xmlns:ext=" http://www.pcce.be/egb/spil-schema/extensions "> <object> <id> http://www.pcce.be/egb/document#KF.boeken.1 </id> <objectnr>1</objectnr> <naam>t Land van Ham, jg. 10, nr. 1</naam> </object>	456

The use of this is described below. Sample data can be found in Annex 6: data samples\Erfgoedregister\Sample records

3.1.2. Erfgoedregister registration module

The registration module for Erfgoedplus, the Erfgoedregister, is a custom-developed online module for registering heritage objects. The Erfgoedregister currently contains around 330,000 records.

The data is managed in an [Oracle Berkeley DBXML](#) environment. The data model has been aligned with [Spectrum 5.0](#) and [CIDOC-CRM](#).

This data is made available as an exported XML file, with the root <root> and an element <Artefact> for each record. All tags are always present, even if they do not contain a value.

Tag	Comments
<objectNumber collection="71022A00">21D</objectNumber>	Contains both identification of the object and the collection.
<access> <accessCategory>no access</accessCategory> </access>	Contains a number of values which have their prefLabel stored in Erfgoedplus, and not the identifier of the relevant concept in FluidOps. Details can be found in Annex 6: data samples\mapping.xls
<currentOwner actor=" http://www.erfgoedplus.be/collectiebeheerder/AB/2785 "> </currentOwner>	Contains a number of values with references to FluidOps. Details can be found in Annex 6: data samples\mapping.xls.
<document document="http://www.pcce.be/egb/document#KF.boeken.117"/>	Contains a number of values with references to XML files where http://www.pcce.be/egb/document#KF.boeken , refers to literatuur.xml http://www.pcce.be/egb/photo#KF.reproductie refers to foto.xml http://www.pcce.be/egb/event#KF.tentoonstelling refers to tentoonstelling.xml http://www.pcce.be/egb/document#KF.tentcat refers to tentcat.xml
<reproduction show="true" type="image" uri="http://www.pcce.be/egb/photo#R.1429013134283168"> </reproduction>	References to media files can be linked in various ways: Via Collection.UUID http://www.pcce.be/egb/photo#31005A08.9a9ace8a-0d61-09c7-e056-446783165e31 Via Objectnumber.UUID http://www.pcce.be/egb/photo#PA.008.K060.3587

	Other http://www.pcce.be/egb/photo#R.1310026277216807
<pre><production> (1..n) <maker> ... <maker actor="">Desaix </maker> </maker> </production></pre>	Agents, e.g. makers, are often registered as free text.

Sample data can be found in Annex 6: data samples\erfgoedregister\Sample records, the mapping of the fields in Annex6: data samples\mapping.xls

3.1.3. Digital Asset Management

An [Alfresco](#) environment is used for storing digital files.

Sample data can be found in Annex 6: data samples\Erfgoedplus\Alfresco\Sample data,

A description of the data in Annex 6: data samples\Erfgoedplus\Alfresco\Erfgoedplus - Alfresco Export (Documentation)(Specifications)

3.1.4. Fuseki triple store (harvesting)

The [Fuseki triple store](#) also harvests data from a number of external registration systems. This functionality and data is out of scope.

3.1.5. Public portal and Europeana

The portal <https://www.erfgoedplus.be/> uses the data in the Fuseki triple store. Records are published when it is explicitly stated that this may be done.

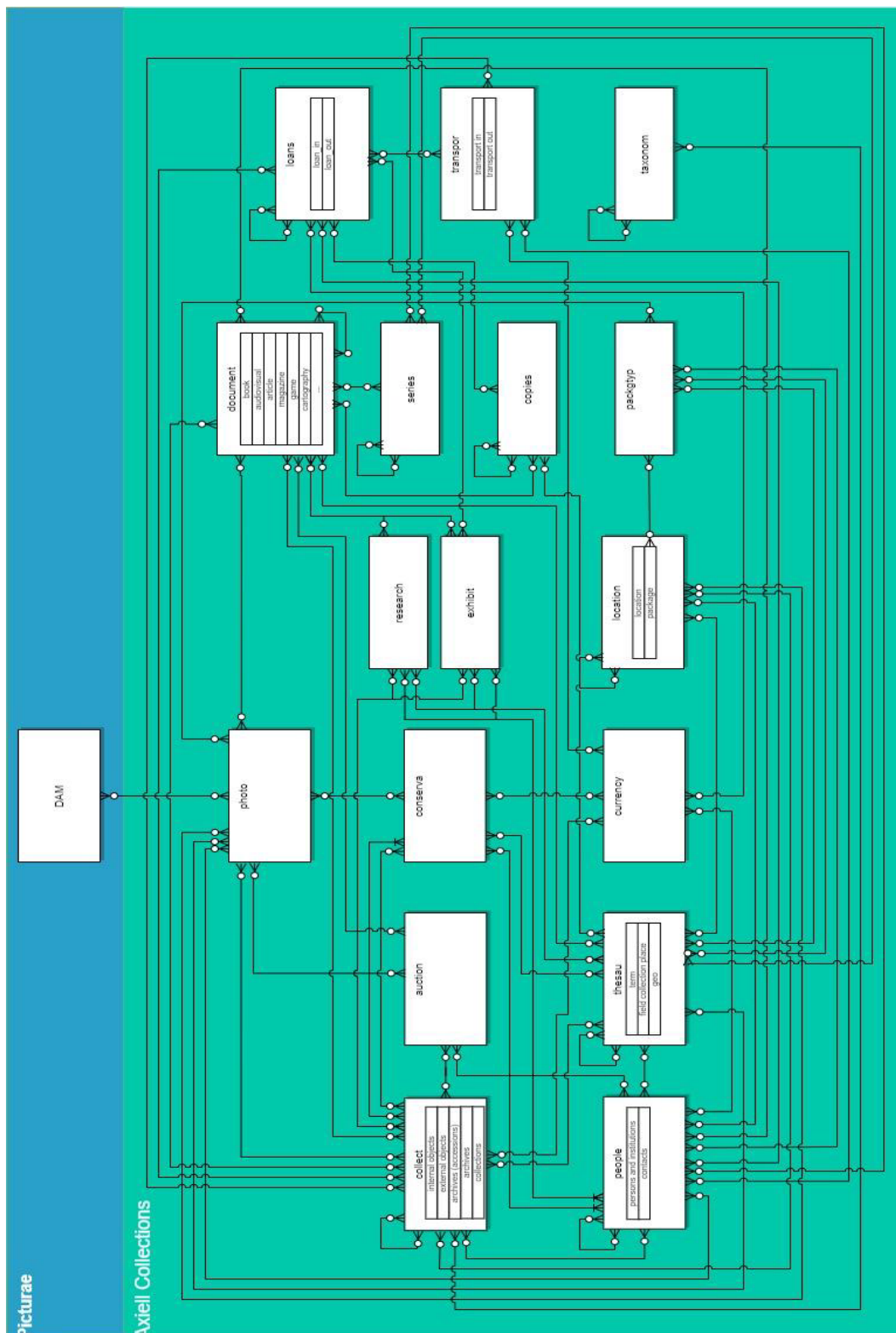
The same applies to the delivery of data on the Europeana platform.

3.2. Erfgoedinzicht

3.2.1. General

Organisations can use Erfgoedinzicht for digital registration, management and access to their tangible heritage collection.

Registration is done in Axiell Collections and Adlib for Windows, but use a single database. These applications mainly support the registration of Spectrum processes, archive management and library management. Data can be imported and exported via templates.



3.2.2. Registration module

The data is categorised into datasets (objects, publications, ...). Datasets can be sub-categorised again into information domains (e.g. objects under own management, objects

under third-party management, books, magazines, etc.). Each dataset contains the data for all Erfgoedinzicht tenants, but are strictly separated from each other with a few exceptions.

The most accurate way to distinguish data from Erfgoedinzicht tenants is by the identifier <preref>. A preref is only unique within a dataset. The same preref can (and will) appear in different datasets.

This preref has a fixed structure.

- The preref always begins with the id_number, a number in the range 1 - 999. No preceding zeroes are added to this, so the length of a <preref> can be variable, *toInt(substr(\$preref\$,0,length(\$preref\$)-7)* is therefore the part of the <preref> that indicates the id_number. <id_number> is also included with a separate tag in the datasets, but not always filled in. All selections at Erfgoedinzicht tenant level must therefore always be made on the <preref>.
- The digit after the <id_number> section in the <preref> indicates the information domain in the main 'collect' dataset, where:

Digit after id_number	Information domain	Explanation
0	Objects under own management/Internal objects	Heritage objects for which the registering organisation is responsible for the (physical) management.
1	Objects under third-party management/External objects	Heritage objects for which the registering organisation would benefit from the heritage object description, e.g. in the context of an incoming loan.
2	Collection records/Collections	Description of a logical set of heritage objects in accordance with Cometa . Not all Erfgoedinzicht tenants have this type of record.
5	Archive records/Archives	Description of archive components. Not all Erfgoedinzicht tenants have this type of record.

- The last part of the <preref> is a sequential number with preceding zeroes, consisting of 6 digits.

All records within the 'collect' dataset have the same basic structure. Depending on the information domain, other tags are filled in. The other datasets each have a unique structure.

The datasets/information domains are:

Information domain	Information domain	Dataset	Comments
Heritage objects under own management	Internal object catalogue	collect	e.g. 10000001-1000999
Heritage objects under third-party management	External object catalogue	collect	e.g. 101000001-101000999

Collection records	Collections	collect	e.g. 62000025 - 62000999
Archive records	Archives (catalogue)	collect	e.g. 205000006 - 205000999
Book	Books	document	e.g. 530000013
Audiovisual	Audiovisual materials	document	e.g. 531000013
Article	Articles	document	e.g. 532000013
[code 3 is not assigned to an information domain]	[code 3 is not assigned]	document	
Magazine	Serials	document	e.g. 534000013
Game	Games	document	e.g. 535000013
Loose-leaf	Loose-leaves	document	e.g. 536000013
Digital sources	Resources	document	e.g. 537000013
Desiderata	Desiderata	document	e.g. 538000013
Cartography	Cartography	document	e.g. 539000013
Thesaurus	Thesaurus	thesau	
Persons and institutions	Persons and institutions	people	
Locations/packages	Locations/packages	location	This dataset is managed by all tenants, provided they have the appropriate rights. The priref has an id_number according to the registrar.
Entry/Despatch	Entry/Despatch	transpor	
Incoming/outgoing loans	Incoming/outgoing loans	loans	
Assessments and treatments	Assessments and treatments	conserva	
Exhibitions	Exhibitions	exhibit	
Research/use	Research/use	research	
Auctions	Auctions	auction	

Copies	Copies	copies	
Taxonomy	Taxonomy	taxonom	
Package types	Package types	packgtyp	
Currencies	Currencies	currency	
Series	Serials	series	
Visual documentation	Visual documentation	photo	

The datasets will contain a number of fixed 'system fields' in addition to the <preref> and <id_number>:

Tag or attribute	Dataset	Explanation
<record preref="530011793" created="2014-06-23T13:35:37" modification="2023-01-13T14:05:09" selected="false" deleted="false">	all	<ul style="list-style-type: none"> • preref: see above • created: creation date of the record • modification: date of last change • selected: may be ignored • deleted: records are also included
<record_type @value=" ">	collect	Should indicate which type of record it is, but not always filled in.
<input_name>	all	Name of the user who was logged in and created the record, but this is not always filled in.
<input.time>	all	HH-MM-SS, time when the record was created.
<Edit>(0...n)	all	This XML element indicates when a record was modified. It consists of: <ul style="list-style-type: none"> • <Edit> • <edit.date>: date of modification • <edit.name>: user • <edit.source>dataset • <edit.time>: time • <edit.notes>: any notes • </Edit>
<diagnostic> ... </diagnostic>	all	Diagnostic details for the API call
<ACCESS_RIGHTS>(1..n)	collect	Access rights at record level
<record_access.owner>	collect	Identification of the current 'owner' of the data

References to other records within Erfgoedinzicht are structured as follows:

Dataset	Explanation
From the 'collect' dataset where this example refers to the 'conservation treatments' dataset	<code><Condition></code> <code><condition.lref>530005760</condition.lref></code> <code><condition.part>entire</condition.part></code> <code></Condition></code>
From other datasets	<code><object.object_number.lref>10003517</object.object_number.lref></code> <code><object.title>Table</object.title></code> <code></Object></code>
Exception dataset loans: <Relatedloan>	<code><record...</code> <code><loan_number>LIB_0002</loan_number></code> <code><Relatedloan></code> <code><related_loan>2013/14a</related_loan></code> <code><related_loan.relation_type>venue</related_loan.relation_type></code> <code></Relatedloan></code> Refers to <code><record...</code> <code><loan_number>2013/14a</loan_number></code>

Depending on the dataset that is referred to within a tag whose name ends in '.lref', one or more tags from the other dataset may be displayed within the .lref tag. The tags within the .lref tag are therefore not migrated for the dataset where the .lref appears.

Given the prirefs are not unique across the database, but only per dataset, a piece of information is missing here to create a good link. We clarify which dataset to address for each tag below.

The data from Erfgoedinzicht is made available via APIs. An account is made available to subscribers upon request. It is important to note here that the datasets each need to be queried separately to harvest all the data.

The base URI for an API call is <https://cjm-web.adlibhosting.com/cjmapi/wwwopac.ashx?database=collect&search=>. Explanation for the search can be found at <http://api.adlibsoft.com/api/functions/search>.

The API delivers XML by default. Adding '&output=JSON' will output JSON. You can find more info at <http://api.adlibsoft.com/documentation>.

Only tags containing data are displayed when using the API. More tags may therefore be present in some Erfgoedinzicht tenants, but each tag within each Erfgoedinzicht tenant has an identical name and meaning.

Exceptions and an explanation for how to handle missing data are provided below.

3.2.3. Digital Asset Management

Digital Asset Management from Picturae (part of the Memorix Maior solution) is used to manage (media) files. The metadata is made available via export. A sample of this can be found in Annex 6. The metadata is quite limited:

Tag	Explanation
-----	-------------

file_uuid	Unique UUID for the essence
file_name	Filename of the essence, possibly without extension
mimetype	Mimetype
path	Possible storage path from root within the DAM, with the folder UUID after the last '/'
Dam path	Possible subpath within path
damcode	Identification of the Erfgoedinzicht tenant, with its own code. We provide the contractor with a mapping file for this when migrating.
filesize	File size
modified_time	Date of last change

3.2.4. Public portal

The <https://erfgoedinzicht.be/> portal uses the data in the Adlib/Axiell Collections and Picturae solutions.

4. Data model

4.1. General

The data analysis shows that the current solutions have a number of modelling and data quality issues that conflict with the objectives for a sustainable logical data model. We provide a non-exhaustive overview of these issues in this section. This overview is accompanied by a number of suggestions and requirements, which are translated into the requirements in Annex 1: requirements.

4.2. Set-up

In order to enable the set-up as outlined in the specifications (III. Technical Requirements), meemoo and the current users express their desire to use standardisation to facilitate this in the solution.

The focus for the solution is on unambiguous and efficient registration and management of heritage data, and the ability to integrate with external solutions.

In practice, this means that:

- not every organisation needs the full capabilities of the solution. In particular, for registering heritage data, sometimes only parts of a logical data model are used. This means that organisations:
 - sometimes want to use (small) parts of the Spectrum information model while other organisations need comprehensive registration. This translates into the need for being able to provide different, standardised input screens to different organisations;
 - sometimes need automated location management, which translates into an optional take-up of automated location management for each participating organisation;
 - will sometimes have no need to manage publications or archives, which translates into an optional take-up of registration of publication and archive records for each participating organisation;

- sometimes already have their own DAM, and want to continue using this service, while other organisations want to see this in the solution, which translates into an optional take-up of a DAM for each participating organisation;
 - sometimes need to register contact details, such as addresses and phone numbers, which translates into an optional take-up of management for this data for each participating organisation, with attention to GDPR legislation. These are implicitly some of the Spectrum procedures that use detailed agent data. It is possible that some participating organisations will not use them.
- some tasks (both registration and management tasks) are placed as close as possible to the participating organisations' requirements, as indicated in the specifications (III. Technical Requirements);
 - there is little alignment, especially for Erfgoedinzicht, between the content in the different tenants' controlled value lists. Meemoo and the solution's users want to take steps forward here and integrate the controlled value lists with each other as much as possible;
 - the aim is to cause minimum interruption to service continuity. This translates into a (limited) number of migration scenarios that can be adapted or optimised in collaboration with the contractor (see Specification: A.3.4.2 Guidelines);
 - interpretation of the accuracy of registered intellectual entities is very challenging at this stage, given the large number of users. Accuracy in these specifications therefore translates into mere migrations from the current data model to the new model, unless specified or agreed otherwise between meemoo and the contractor.

4.3. Expectations for future data model and management

Given the above, as already stated, cleansing the existing data is only possible to a limited extent. Meemoo also sees opportunities to partially optimise the data during migration, however. Meemoo would like to engage in dialogue with the contractor about this so that further steps towards optimisation can be taken, without jeopardising deadlines. In particular, it is possible to take steps towards optimisation of controlled value lists.

Some of these expectations have been translated into requirements in Appendix 1: requirements, with modelling of controlled value lists based on applicable standards.

Meemoo refers here to an approach based on:

- (a subset of) [SKOS](#);
- registration of the authorities' provenance in accordance with (a subset of) the [PROV ontology](#).

The main features of this are:

- authorities are either available to all users or only within a cluster;
- authorities are multilingual;
- authorities may be relevant for different controlled value lists (which may require a poly-hierarchical structure);
- separate variants of shared authorities can be created within clusters (e.g. with their own skos:scopeNote or skos:broader). Variants must be able to refer to the shared authorities, for example via skos:exactMatch and skos:closeMatch;
- authorities can be reconciled with a reliable (linked data) source, e.g. via skos:exactMatch or skos:closeMatch;
- authorities can have at least a status of 'candidate', 'rejected' or 'approved';
- when creating or modifying authorities, the relevant provenance properties, such as 'creation date', 'modification date', and 'source references' are added;

Given the variable quality of these value lists, meemoo believe it is opportune to load some external resources as defaults for all users, including a URI pointing to that external source. This concerns sources such as (parts of) AAT, currencies, geographical data, dimension types, etc. for which meemoo has already drawn up a shortlist, with suggestions for Wikidata queries.

These 'default' authorities can then be supplemented by cluster-based authorities in accordance with the above principles. Clearly referenced authorities can be linked via skos:exactMatch.

With a view to using permanent relationships and data re-use, it is a requirement that key entities are identified by persistent identifiers (PID/pURLs). The data model supports the use of these PID/pURLs in accordance with a defined syntax, as well as the use of additional PID/URLs in accordance with other standards.

The use of these PID/pURLs at least applies to:

- all described (physical or conceptual) entities: objects, publications, exhibitions, concepts, persons and institutions...;
- the main procedures as described in the Spectrum procedures: loans, restorations, acquisitions, etc.

In the risk analysis, the tenderer pays sufficient attention to the aspects of data migration and data modelling, and the issues that (may) arise in the process.

4.4. Explanation of the data analysis, including possible optimisations

4.4.1. Account and access profiles

The existing user accounts and access rights do not necessarily need to be migrated. As indicated in Annex 1: requirements, the tenderer makes a proposal for this that facilitates the solution design.

4.4.2. Provenance and audit trail

The tenderer preferably uses the [PROV ontology](#) to identify the data origin.

The following restrictions on data migration in current solutions are permitted:

Erfgoedregister	Any restriction
The FluidOps audit trail, essentially the http://www.fluidops.com/ scheme.	This content may be ignored.
Erfgoedinzicht	Restriction
The audit trail tags for creation of a record, essentially the tags <input.xxx>	Are retained, but do not need to be matched with future user accounts.
The audit trail tags for modification of a record, essentially the tags <edit.xxx>	Are retained, but do not need to be matched with future user accounts.
The previous audit trail <edit.xxx> tags	May, if the contractor prefers, be made available to participating organisations in a non-integrated way, but are searchable for the registrar.

4.4.3. Controlled value lists

4.4.3.1. Erfgoedregister value lists('Lists' in FluidOps)

Erfgoedplus contains a number of value lists, mainly intended to manage data in a consistent way. These value lists have a limited set of properties, which are modelled in SKOS:

Properties
<u>preferred label</u>
<u>alternative label</u>
<u>definition</u>
<u>scope note</u>
<u>is in scheme</u>
<u>hidden label</u>
<u>rdf:type</u>
<u>skos:broaderMatch</u>
<u>has broader</u>
<u>has narrower match</u>

4.4.3.2. Erfgoedinzicht value lists

Some value lists in Erfgoedinzicht are not available in the export or registration modules. These are indicated as 'drop down' in Annex 6: data samples\mapping.xls.

4.4.3.3. Expectations for values lists

In consultation with the contractor, these value lists may be aligned with existing value lists in the solution, especially if they have an impact on the solution's functional behaviour. We are thinking here, for example, of the Spectrum information model for [Use Information](#), which defines the status of the registration, the possibilities for access, etc.

In any case, meemoo considers it a requirement that value lists can be managed by authorised users without any intervention from the contractor.

4.4.3.4. Erfgoedregister concepts ('folders')

The file Annex 6: data samples\Erfgoedplus\FluidOps\FluidOps schema.xlsx lists all the unique rdf types and their unique properties. In many cases, it uses external ontologies and schemas, so only a limited explanation is needed in this annex.

For assigning typologies of objects, materials used, techniques, etc. Erfgoedplus makes near-full use of the [Getty AAT](#), with the facets of materials, object names and styles, and periods loaded into FluidOps using an (obsolete) copy.

The triples are included in their own subcategory, the 'folders', where the folders themselves are modelled as concepts and included using 'has narrower match' AAT concepts. An example of this is:

Subject	Predicate	Object
<u>Materials</u>	<u>rdf:type</u>	<u>Concept Scheme</u>
<u>Materials</u>	<u>has top concept</u>	<u>coatings</u>
<u>Materials</u>	<u>has top concept</u>	<u>pottery</u>

The AAT concepts themselves remain unchanged and have the original 'broader match', which can therefore refer to another concept. The namespaces follow the [Getty Vocabulary Program ontology](#), with the exception of 'GVP', which is replaced by 'AAT'.

The contractor will preferably continue this approach, where these authorities are made available as commonly available concepts within the objectives for the data model. Given the outdated state of the current data, there needs to be a new import. The contractor may choose

- (in consultation with meemoo) to include only the relevant properties,
- or the entire model for AAT concepts.

This also applies for the languages specified as required in the requirements or for all languages. Meemoo is open to alternative approaches from tenderers as long as they meet the data model's requirements and objectives.

4.4.3.5. Erfgoedinzicht concepts ('terms')

This dataset is used to describe different types of intellectual entities (heritage objects, archive descriptions, publications) in a coherent way.

Based on an export (dated June 2022), we identify a total of 258,191 terms in the thesau dataset, of which 192,922 were unique terms.

It is notable here that, unlike SKOS which is concept-based, the thesau dataset is term-based (where a term can cover different content loads) and multiple terms can refer to the same concept.

The full structure is described in Annex 6: data samples\Erfgoedinzicht\Erfgoedinzicht Thesau en agent model.xls. We highlight some of the issues below. Given the potential impact of the optimisations for the solution and migration proposed below, meemoo wants to coordinate these aspects sufficiently together with the contractor in order to create a specific approach.

4.4.3.5.1. Status

All terms have a status (0..1) indicating the possibilities for their use. The status is indicated by a language-independent code (lang="neutral") and repeated per language. The languages are identified by a code (0=en, 1=fr, 2=de, 3=nl, 5=it).

Tags	Explanation
<code><term.status></code> <code><value lang="neutral">0</value></code> <code><value lang="0">undefined</value></code> <code><value lang="1">niet gedefinieerd</value></code> <code><value lang="2">non déterminé</value></code> <code><value lang="3">nicht definiert</value></code>	<p>This status is unique per term, but often unreliable. The following statuses were found for the language code 1 (nl):</p> <ul style="list-style-type: none"> • [niet ingevuld] • descriptor

<pre><value lang="5">indefinito</value> </term.status></pre>	<ul style="list-style-type: none"> • kandidaat • niet gedefinieerd • non-descriptor • verouderd • verworpen
--	--

Depending on the Erfgoedinzicht tenant's choice, these statuses are mapped to a similar status from the concept in the new solution.

4.4.3.5.2. Type

The tags <term.type> (0..n) indicate the domain (materials, techniques...) to which this term belongs.

Tags	Explanation
<pre><term.type> <value lang="neutral">COLLECTIONPLACE</value> <value lang="0">Collection place</value> <value lang="1">Vindplaats</value> <value lang="3">Fundort</value> </term.type></pre>	These may be multiple, and may indicate conflicting meanings in content, making de-referencing in the facts impossible

There are more than 100 types present for all Erfgoedinzicht tenants in total. It is notable here that:

- some types do not satisfy the definition of a concept:
 - descriptions of the sites where archaeological materials are found, for which address details, contacts and similar, are in the thesau dataset;
 - geo-data, see below in this annex;

The contractor must assign them a logical place, re-using existing models such as geographical data and agents.

- the thesau dataset's term-oriented approach makes it strongly preferable that terms are split, according to their type, into concepts by domain, based on each assigned type. This means that the data referring to these terms is impacted, and are required to follow this split.

4.4.3.5.3. Use/used_for

Preferred and non-preferred terms are indicated in the facts via the <use> and <used_for> tags. Each term has its own record and identifier (priref) here.

Tags	Explanation
<use>bike</use>	Use is a reference to the preferred term (skos:preflabel) used to identify a concept
<use.lref>12000001</use.lref>	Use.lref is a reference to the database record of the Use term
<used_for>bicycle</used_for>	Used for is a reference to the non-preferred term(s) (skos:altlabel) used to identify the concept
<used_for.lref>12000002</used_for.lref>	Use.lref is a reference to the database record of the Used for term

It is (strongly) preferred that the different terms are included together in one concept, creating skos:altLabels for non-preferred terms, for example. The other tags, except for <term.status>

(see above), may be ignored here.

(see also below for merging terms based on (indirectly) shared URI)

4.4.3.5.4. References to external sources

References to external authorities consist of a combination of

- a baseURL (which defines the external authority);
- an identifier (which defines the concept within the authority).

The combination of baseURL and identifier enables the creation of a unique (resolvable) URI (e.g. <http://vocab.getty.edu/aat/300249430>).

Tags	Explanation
<code><source>https://vocab.getty.edu/aat/</source></code>	Record source contains the base URL for the external vocabulary.
<code><term.number>300249430</term.number></code>	Source number contains the identifier of a concept in an external vocabulary.

Note: actual data may differ from this template, e.g.:

- the full URI is included in `<term.number>`;
- the full URI is included in `<source>`;
- the URI refers to an HTML representation (e.g. <https://vocab.getty.edu/page/aat/300249430>);
- the base URL is incorrect or incomplete (e.g. missing end slash, authority name instead of baseURL...)

The contractor preferably adapts some obvious deviations from the template during migration. Where it is not possible to make this adjustment, the combination of `<source>` and `<term.number>` are migrated to a field to be added to the model.

When these references produce valid URIs that match internal globally shared authorities (see above) within the same domain, they are expected to be linked during migration.

If possible, the contractor merges several terms within a cluster:

- when they have a shared URI referring to the same external resource
- when an equivalence relationship between different URIs is available in one of the external authorities (e.g. Identifiers in a Wikidata record).

This also takes into account the desire for organisations to be able to create their own variants of a concept, linked together by a shared URI and relating to the same concept (see above 4.3 Expectations for future data model and management, Annex 1: requirements and Annex 3: use cases (UC-13: Adding and using concepts)).

4.4.3.6. Erfgoedplus geo-data

The geodata in the FluidOps for Erfgoedplus always have an `rdf:type` that start at <http://data.erfgoedplus.be/def> and follow a clear structure.

- `egp:structure` is a distinct class of `structure` that is not logically part of geo-data, but is often used as a (storage) location in the Erfgoedregister;
- relationships between administrative locations are established via `egp:contains` and `egp:within`;
- there are two possibilities for relationships for historical buildings and streets:
 - you can either map something historical precisely onto something current: *situatedAt*;

- or situate it within something current: *situatedWithin*;
- streets can be 'within' a sub-district as well as 'within' a hamlet. a hamlet is 'within' a (non-administrative) municipality;
- Hamlets and (vernacular) municipalities are mapped to current administrative locations using the *situatedAt* and *situatedWithin* relationships;

4.4.3.7. Erfgoedinzicht geo-data

Geographic data is located in the *thesau* dataset and have a type *GEOKEYW*, *GEOKW*, *LOCAT*, *PLACE*, *geography* or *geographical keyword*

It is (strongly) preferred that the various terms are placed in a single model aligned with this data, together with the geographical data from Erfgoedplus. Meemoo is prepared to investigate together with the contractor how this data can be optimised.

4.4.3.8. Erfgoedplus agents

Agents can be found on three levels for the whole of Erfgoedplus:

Current solution	Current location in the solution	Comments
Archiefbank, AE_(Meemoo)_xxx.xml	<AE_REFERENTIE_ROW num="1"> ...	This tag refers to BWPL_Meemoo.xxx.xml, where the information is more complete
Archiefbank, BWPL_Meemoo.xxx.xml	[BWPL_Meemoo_xxx.XML]..<IDENTIFICATIE_BLOCK_ROW num="1"><OMSCHRIJVING><OMSCHRIJVING_ROW num="1">	See Annex 6: data samples\mapping.xls
FluidOps	http://xmlns.com/foaf/0.1/Person and http://www.w3.org/ns/org#Organization	Only used to describe people depicted and organisations depicted
Erfgoedregister	<production> <maker> <role>public institutions </role> <qualifier></qualifier> <maker actor="">PCCE (Provinciaal Centrum voor Cultureel Erfgoed)</maker> </maker>	Contains free text in many cases and, depending on <role>, may be the creator of the object or the creator of a representation of the object
Erfgoedregister	<conservationTreatment> <conservator actor=""> </conservator> </conservationTreatment>	Contains free text in many cases
Erfgoedregister	<insurance> <insurer actor=""> </insurer> </insurance>	Contains free text in many cases
Erfgoedregister	<loanOut> <borrower actor=""> </borrower> </loanOut>	Contains free text in many cases
Erfgoedregister	<aboutPerson> <role>glassworkers</role>	Informative, as these agents should not be created based on

	<code><person actor="http://data.erfgoedplus.be/person/140299#id"></person></code>	data in the Erfgoedregister: Reference to a person or organisation in FluidOps
--	--	---

Agents from the Archiefbank environment and FluidOps should be added to the solution's agents. The same applies for the distinct selection of free text variants for <maker actor>, <conservator actor>, <insurer actor> and <borrower actor>, with <maker><role> continuing to be assigned as a linked role (which is included in the concepts).

4.4.3.9. Erfgoedinzicht agents

The persons and institutions (people) dataset contains all agents used in the application to describe items and identify actors (e.g. maker, acquisition source, depicted person).

As with concepts, relationships can be established between agents in Erfgoedinzicht, agent records have a type and a status, and external resources can be referred to.

4.4.3.9.1. Status

All agent records have a status (0..1) indicating the possibilities for their use. The status is indicated by a language-independent code (lang="neutral") and repeated per language. The languages are identified by a code (0=en, 1=en, 2=fr, 3=de, 5=it).

Tags	Explanation
<pre> <name.status> <value lang="neutral">2</value> <value lang="0">approved non preferred term</value> <value lang="1">non-descriptor</value> <value lang="2">référence</value> <value lang="3">Verweisungsform</value> <value lang="5">termine non preferenziale approvato</value> </name.status> </pre>	<p>The following statuses were found. This may differ depending on the cluster:</p> <ul style="list-style-type: none"> 0 = undefined 1 = descriptor 2 = non-descriptor 3 = candidate 4 = obsolete 5 = rejected

Depending on the Erfgoedinzicht tenant's choice, these statuses are mapped to a similar status from the concept in the new solution.

4.4.3.9.2. Type

The tags <name.type> (0..n) indicate the domain (maker, acquisition source, depicted person, etc.) that this agent record belongs to.

Tags	Explanation
<pre> <name.type> <value lang="neutral">ACQUISITIONSOURCE</value> <value lang="0">acquisition source</value> <value lang="1">verwervingsbron</value> <value lang="2">source d'acquisition</value> <value lang="3">Erwerbungsquelle</value> <value lang="5">fonte di acquisizione</value> <value lang="6">πηγή απόκτησης</value> </name.type> </pre>	<p>The following statuses were found (not exhaustive).</p> <ul style="list-style-type: none"> ACQUISITIONSOURCE = acquisition source AUTHOR = author COLLECTOR = finder CORPORATE = corporate author CURATOR = curator DISTRIBUTOR = distributor FAMILY = family ILLUS = illustrator INST = institution MAKER = maker

	<ul style="list-style-type: none">• PERSON = person• PRINTER = printer• PUBL = publisher• SUPPLIER = supplier
--	--

These profiles are optionally linked to an external ontology as [GND](#) and roles from the FluidOps in Erfgoedplus.

4.4.3.9.3. Relationship preferred/non-preferred term, equivalent term

Preferred and non-preferred agent names are indicated via the <use> and <used_for> tags. Each term has its own record and identifier (priref) here.

Equivalent relationships between agent records are established via the <equivalent_name> tag. These equivalences may refer to alternative names for an agent (e.g. pseudonym) or names in another language.

These relationships are structured almost identically to the use/used_for relationships for concepts in Erfgoedinzicht (see above).

It is also strongly preferred here to include the variant names of one term together in one concept, adding the non-preferred terms as a skos:altLabel (see above).

4.4.3.9.4. References to external authorities

References to external authorities are highly analogous to the references to external authorities for concepts. It is highly desirable that these URIs are processed in the same way as those for concepts (see above).

Migration should take into account that agents may be part of a shared and/or local agent file. Agent records in the shared file must have a URI to an external authority (see above, 4.3 Expectations for future data model and management, Annex 1: requirements, and Annex 2: use case UC-12: Adding and using agents as authority).

4.4.3.9.5. Distinction between persons and organisations

As far as we can see, there is no guaranteed distinction between persons and organisations in the data; this is sometimes done by assigning a 'person' or 'organisation' <name.type>. Further interpretation of the data by meemoo or the contractor may provide more clarity here. This can be looked at in more detail in the migration phase. The tenderers indicate the extent to which this is problematic for migrating agent data.

4.4.4. Collections

As described above, collections in Erfgoedplus are created as external authorities at Archiefbank, and the Erfgoedregister only has limited descriptions.

In Erfgoedinzicht, collections can be described as a separate intellectual entity in accordance with the COMETA standard, see Annex 6: data samples\mapping.xls.

Not all Erfgoedinzicht tenants use this option. The 'thesau' dataset contains a 'collection' <term.type> that allows objects to be assigned to a collection, but without the ability to further describe or manage this collection.

All these collections should be modelled as separate intellectual entities in accordance with the COMETA standard, preserving their relationships with other intellectual entities and hierarchical relationships between them.

4.4.5. Heritage objects

These are mapped to Spectrum as indicated above in Annex 6: data samples\mapping.xls.

Meemoo requests that special attention is paid to:

- allocation to collections, where
 - records referring to a collection defined as separate intellectual entities use the form <Part_of>
<part_of_reference.lref>102000001</part_of_reference.lref></Part_of>;
 - records assigned to a collection using the 'thesau' dataset use the form
<collection.lref>340000004</collection.lref>
- the Spectrum information groups [Reproduction](#), [Reference](#), [Use object](#), [Use of collections](#), and the procedures for [Rights](#), [Rights in](#) and [Rights out](#).
Given the intertwined nature of these information groups and procedures, and their sometimes less than transparent relationships in Spectrum, meemoo expects tenderers to explain here how this can be configured.

Meemoo assumes that the registering of rights (both intellectual entities and representations and digital files) does not necessarily lead to automatic interpretation and implementation of (public) access rights, publication statuses, assumptions about the completeness of the registration as a function of making data accessible, etc. No complex business logic is therefore needed to automate interpretations.

Meemoo does expect, however, that registrars can (visualise and) manage all this in a simple way, and be able to use individual intellectual entities, representations and digital files on (a selection of) platforms, including for ingest into the meemoo archive

Meemoo also wants more flexibility to efficiently link media files, digital files and references to documentation, including descriptive metadata, etc. to IEs or Spectrum procedures;

- the Spectrum information groups [Place](#) and [Location](#), which are complex and sometimes not transparent in the current solution.
Meemoo wants to adopt a transparent model in the final data model, which must also allow automated location management;
- data related to exhibitions is crucial for various current users. Spectrum proposes including these in the [Use of collections](#) information model. The contractor may be able to use a different, but still Spectrum-based, model for this to help with registration efficiency.
- Erfgoedinzicht does not currently provide any clear place for registering PIDs/pURLs for intellectual entities. PIDs/pURLs have been registered in various places and forms in the current data model. Meemoo wants to be able to apply special exception rules for mapping this data, whereby meemoo can explain to the contractor which tags qualify for this for each Erfgoedinzicht tenant.

4.4.6. Archives

These are mapped as indicated above, see Annex 6: data samples\mapping.xls.

As indicated in the requirements, providers must make a substantiated proposal for a standard for registration and access.

Archive data appears exclusively (and to a very limited extent) in Erfgoedinzicht. Each tenant is free to use the archive modules, and therefore also to migrate this data.

As indicated above, these records are in the 'collect' dataset, and the 'accessioning' (acquisitions) and 'catalogue' (archive descriptions) types are possible. In practice, only a very limited number of records of the 'accessioning' type are available, with data analysis indicates that these are mainly test records. It may be agreed with current users that 'accessioning' type records will not be migrated.

Tags	Explanation
<code><Part_of></code> <code><part_of_reference.lref>206015608</part_of_reference.lref></code> ... <code></Part_of></code>	Referring to the archive level above, the referenced record always has a <code><Parts></code>
<code><Parts></code> <code><parts_reference.lref>206040113</parts_reference.lref></code> ... <code></Parts></code>	Referring to the archive level below, the referenced record always has a <code><Part_of></code>

In Erfgoedinzicht, parts of the [archive descriptions are registered as parts of Spectrum procedures](#). Meemoo wants to continue and optimise this strategy in collaboration with the contractor. This also means that the existing datasets from these Spectrum procedures contain references to archive records. If participating organisations decide to stop including their archive management in the solution, the references to these archive records in the relevant datasets must be captured with a minimal description of the original archive records.

4.4.7. Publications

These are mapped to MARC21 as indicated above in Annex 6: data samples\mapping.xls. As indicated in the requirements, providers must make a substantiated proposal for a standard for registration and access.

This type of data appears exclusively in Erfgoedinzicht. Each tenant is free to use the publication module, and therefore also to migrate the data. In some cases, the data will already be migrated to other solutions. If participating organisations decide to stop including their publication management in the solution, the references to these publications in the relevant datasets must be captured with a minimal description of the original publication.

4.4.8. DAM/meemoo archive

Asset management is done in different solutions as described above. The descriptive metadata always needs be located in the registration module. Meemoo is counting on technical metadata (mime type, file size, etc.) being created upon inclusion in the new solution.

The assets themselves are supplied in accordance with the contractor's requirements, but loaded into the solution by the contractor.

5. Migration/start-up requirements

Meemoo provides a number of validation steps at key milestones in the migration process (see Specifications B.2. PAYMENTS). For the migration phase, it is important that

- All components in Erfgoedplus are migrated as a whole;
- Erfgoedinzicht tenants are migrated in the order in which users sign up, starting with the pilot users.

Meemoo assumes here that the contractor:

- provides the facilities required for testing functionality and migration;

- the contractor may propose a phased approach for this, based on modules that the client can make available;
- sets up an iterative migration process that limits, initially, any interruption of the registration:
 - as already indicated, the contractor can count on additional support from pilot users and meemoo for this;
- takes the necessary steps, such as training, documentation, etc., to effectively allow organisations to get started in a timely manner;
- draws up a testing and acceptance plan in consultation with meemoo.

To facilitate testing and acceptance, the contractor provides logging for the migration that must enable meemoo and registrars to efficiently control manipulations (with high impact) that are required to include the data in the new solution. Meemoo is thinking here of records where fundamental changes are made compared to the current data, e.g. the authorities issue described above.

Meemoo will act as the Single Point of Contact for end-user acceptance, capturing and reporting issues, and be responsible for acceptance on behalf of the contractor.