



# **meemoo**

**VLAAMS INSTITUUT VOOR HET ARCHIEF**

## **Competitive procedure with negotiation**

for the purchase of

## **Software solution and services for Knowledge Graph based on metadata management**

Contract no. "meta-mnmt-2022"

## **TENDER SPECIFICATIONS**

### **Deadline date and time for the receipt of bids**

Deadline date: 30 September 2022

Deadline time: 10 am CEST

### **Disclaimer**

Transparency and knowledge sharing are core values for meemoo. That is why we share our main tender files in the domains of digitization and archiving. This way, everyone can see how we work or be inspired. Yet, we would like to include this warning and disclaimer.

This document has been drawn up in the context of a specific project, with a specific objective, timing and budget, which in turn are rooted in all kinds of circumstances, meemoo's vision on metadata management and so on. During or after the execution of the project, this vision, circumstances or other context elements may change or have changed. Furthermore, this document may also contain errors. The person who copies these documents in whole or in part is and remains fully responsible for assessing the consequences for his interests. It is best to obtain sound legal advice, especially - but not exclusively - for copying the legal provisions.

Meemoo can in no way be held liable for any disadvantage that any party may incur by making available or copying all or part of these documents.

# I. GENERAL PROVISIONS

## I.1. LIST OF DEVIATIONS FROM THE IMPLEMENTATION RD

These tender specifications do not deviate from the Implementation RD.

## I.2. CONTRACTING AUTHORITY

1. This contract is issued by meemoo, the Flemish Institute for Archives.

<p>meemoo vzw Ham 175 9000 Ghent VAT: BE 0644.450.380 Tel.: +32 9 298 05 01</p>
---

2. Contacts for this procedure

General follow-up:

- Miel Vander Sande
- [miel.vandersande@meemoo.be](mailto:miel.vandersande@meemoo.be)
- Tel.: + 32 492 832 129

Responsible for this contract:

- Matthias Priem
- [matthias.priem@meemoo.be](mailto:matthias.priem@meemoo.be)
- Tel.: + 32 486 035 418

All official correspondence regarding this contract must be sent by email.

***Questions about this contract are to be asked by email. Answers will be added as an additional document on the e-procurement platform to be available to all candidates.***

***We ask that candidates send questions before Octobre 7th, in order to formulate and publish an answer in a timely manner.***

***If necessary, a Q&A can be organized for this version,  
but for the time being we assume a written answer to any questions.***

3. Every bailiff's writ intended for the contracting authority must be served on meemoo vzw, Ham 175, 9000 Ghent. It is irrelevant whether it concerns the service of a summons, a court order or other writ.

### I.3. OBJECT AND CLASSIFICATION OF THE CONTRACT

Meemoo manages a large quantity of mainly audio-visual material from more than 160 partners in cultural heritage and media. More than 6 million objects are currently stored, ranging from digitised newspapers, photos, video, and audio. In addition, we offer a number of access platforms that make the digitised content available to specific target groups, including teachers, students, professional re-users, or the general public.

Metadata is a key element in all our processes. An important part of our activities is to collect, integrate, manage, and search a large variety of heterogeneous metadata across the archived content. The scale of this has increased enormously, making our current approach to volume, flexibility, and accessibility inadequate.

One of the specific challenges here is the data created in the ERDF-GiVE metadata project, in which data from machine learning algorithms (speech recognition, face detection and entity recognition) will have to be modelled and stored for reuse.

Therefore, meemoo wants to focus on a modern Knowledge Graph-based infrastructure that offers application-independent, uniform access to (meta)data that is spread across various systems and formats.

Specifically, meemoo wants to purchase a software solution that can store, manage, and query a large volume of metadata in a flexible graph model and integrate it with our existing systems and processes.

The scope of this assignment consists of two parts:

- the supply of a software solution, consisting of:
  - a basic solution;
  - 10 possible options;
  - integration into the meemoo access infrastructure.
- the provision of consultancy services after delivery.

For a further explanation of the contract, see part IV. Technical instructions.

This contract is a contract for services within the meaning of art 2, 21° of the Public Procurement Law of 17 June 2016.

Applicable CPV code: 72322000-8; see e-procurement platform.

## I.4. AWARD METHOD

This contract will be awarded by competitive procedure with negotiation, on the basis of Article 38, § 1, in particular the following exceptional cases of the Public Procurement Law of 17 June 2016:

- The services include design or innovative solutions
- The technical specifications must be accurate

The most economically advantageous tender is determined on the basis of the best price-quality ratio, which is determined on the basis of the award criteria as indicated in A.5.

## I.5. APPLICABLE LEGAL PROVISIONS

### a) Government procurement regulations

- Public Procurement Law of 17 June 2016 (hereinafter: Public Procurement Law);
- Royal Decree of 18 April 2017 awarding public contracts in classical sectors (hereinafter: Awarding RD);
- Royal Decree of 14 January 2013 determining the general implementing rules for public contracts (hereinafter: Implementation RD);
- Law of 17 June 2013 on the grounds, information and remedies regarding public contracts, certain contracts for work, supplies and services and concessions.

These regulations can be found at:

<http://overheid.vlaanderen.be/regelgeving-overheidsopdrachten>

### b) Environmental, social, and labour law

Social and labour law as referred to in Article 7 of the public procurement law is understood to mean, among other things:

- the decree of 10 July 2008 containing a framework for the Flemish equal opportunities and equal treatment policy; the law of 10 May 2007 to combat certain forms of discrimination, the law of 10 May 2007 amending the law of 30 July 1981 to punish certain acts motivated by racism or xenophobia and the law of 10 May 2007 to combat discrimination between women and men;
- the law of 4 August 1996 on the well-being of employees in the performance of their jobs, more specifically chapter Vbis. Special provisions regarding violence, harassment, and sexual harassment at work.

## I.6. OTHER GENERAL APPLICABLE PROVISIONS

1. This contract has a duration of 3 years, commencing from the date stated at the time of closing.
2. In addition to the above-mentioned term, the contract may be extended twice for a period of 1 year for reasons of continuity of service, on the basis of Article 57, § 2, of the Public Procurement Law. Meemoo explicitly chooses a maximum duration of 5 years (instead of 4) because the solution will be an important part of the meemoo metadata infrastructure, where stability and continuity are highly desirable. The extra year also offers us the opportunity to make full use of the solution for 4 years and then have one (1) year to inquire with the market again, without jeopardising continuity.

The extension means that the contractual conditions remain unchanged.

This extension will be done tacitly unless a registered letter to the contrary is sent by meemoo at least 1 month before the expiry date of the contract.

3. This contract is a framework agreement as referred to in art. 2, 35° of the Public Procurement Law of 17 June 2016, whereby contracts will be awarded as and when required. The framework agreement will be concluded with one participant.

### Maximum value of the contract

*The maximum value of the contract (over the entire duration) is set at €750,000 excluding VAT. The maximum value of the framework agreement is determined on the basis of the maximum quantities to be ordered across all orders that can be placed within the framework agreement.*

4. The candidate may use both Dutch and English in their oral and written relationship with meemoo. The government may require a translation, possibly a sworn translation, of documents that are only available in another language.
5. The use of electronic means for the exchange of written documents is mandatory, both in the context of the awarding and the implementation of the contract. However, registered mail does not have to be electronic. Tenders must be submitted in accordance with the requirements set out in A.3.2. Bidders must provide one or more e-mail addresses on the tender form for electronic communication.
6. The contracting authority ensures that this contract is carried out in accordance with anti-discrimination legislation. See the provisions on non-discrimination under B.7.1.
7. All personal data included in the tender / requests to participate in response to the requirements of the specifications will be processed by meemoo in accordance with the provisions of the General Data Protection Regulation. Personal data will be processed solely for the purpose of awarding and executing the contract. Access to and inspection of the documents shall be limited to meemoo staff and the authorities responsible for administrative and budgetary control, for whom such access and inspection is necessary in connection with the award and implementation of the

contract. Personal data will be processed (stored, etc.) on the IT systems of meemoo. The bidder has the consent of the person concerned to add this personal data to the tender. In accordance with Article 164 §4 of the Public Procurement Law, and like all elements of an award file, personal data is kept for 10 years.

## II. ADMINISTRATIVE PROVISIONS

### A. AWARDING OF THE CONTRACT

#### A.1. SELECTION

No grounds for exclusion may apply to the bidder (Articles 67 to 69 of the Public Procurement Law). This includes the mandatory grounds for exclusion, the grounds for exclusion related to fiscal and social debts, and the optional grounds for exclusion.  
The bidder can prove corrective measures.

The above provision applies individually to the participants who submit a tender together as a consortium.

If the tenderer has relied on the capacities of subcontractors or other entities in the context of the qualitative selection (when submitting the request to participate), the indications regarding the part of the contract to be performed by these subcontractors must be consistent with the information in the request to participate.

These subcontractors or other entities whose capacities the tenderer has relied upon, may not be excluded either.

At this stage of the procurement procedure, the tenderer must continue to meet the qualitative selection criteria as specified in the selection guidelines.

If the tenderer has not yet electronically signed the European Single Procurement Document (ESPD) in e-Tendering when submitting the request to participate, the tenderer must add the ESPD to the tender again.

This also applies to the ESPD for each participant in a consortium or any entity whose capacities the tenderer has relied upon.



## A.2. MODALITIES

### A.2.1. LOTS (ART. 58 LAW, ART. 49-50 AWARDDING RD)

This contract will not be divided into lots, given that the object of the contract is one specific service, the components of which cannot be offered separately.

### A.2.2. VARIANTS ((ART. 56 LAW)

There are no required or allowed variants.

### A.2.3. OPTIONS (ART. 56 LAW, ART. 48 AWARDDING RD)

For a detailed description of the options, please refer to the requirements (III.6).

#### **Required options**

The tenderer is obliged to submit a bid for each required option, otherwise there will be a penalty of substantial irregularity of their bid.

Failure to comply with the minimum requirements of an option also entails the substantial irregularity of both the basic bid and the option.

- **Option A:** GraphQL access
- **Option B:** Link with search index

#### **Allowed options**

The submission of permitted options is optional. An authorised option may be accompanied by an additional price or any other consideration. Offering permitted options improves the quality of the bid and is, therefore, included in the award criteria (see A.5.1.).

- **Option C:** ETL framework
- **Option D:** Graphic administration interface
- **Option E:** Automatic annotation of Dutch text
- **Option F:** IIIF 3.0 Presentation API
- **Option G:** graph data analysis
- **Option H:** Integration with the SAML Identity Provider
- **Option I:** Reconciliation API
- **Option J:** OAI-PMH API

### Free options:

The submission of free options is prohibited.

It is not permitted to submit a bid for an option without submitting a basic tender or, where appropriate, a variant (see A.2.2.).

The offer for options shall be stated in a separate section of the tender. The tenderer shall always indicate clearly that these are options.

## A.3. TENDER – OPENING, SUBMISSION, FORM AND CONTENT

### A.3.1. DEADLINE DATE AND TIME FOR THE RECEIPT OF TENDERS AND OPENING (ART. 92 AWARDDING RD)

Deadline date and time: see cover page.

This deadline date and time determine the timely submission by the tenderers. Any tender received on or after this time shall be considered late. Late tenders will not be accepted.

### A.3.2. SUBMISSION OF TENDERS (ART. 14 LAW)

Bids must be submitted electronically via the e-Tendering website <https://eten.publicprocurement.be/>, an electronic platform in accordance with Art. 14, § 7 Public Procurement Law.

More information on the use of e-Tendering can be found on <http://www.publicprocurement.be> or via the e-procurement helpdesk at +32 (0)2 740 80 00, or [e.proc@publicprocurement.be](mailto:e.proc@publicprocurement.be).

### A.3.3. SIGNING OF TENDERS (ART. 42-44 AWARDDING RD)

The tenderer is advised that their tender, submitted via e-Tendering, must be signed electronically with a valid **qualified electronic signature**.  
A scanned signature is insufficient!

The electronic signature must be placed **on the submission report** in e-Tendering.

This electronic signature must originate from the **person(s) authorised or entitled to bind the tenderer**. The tenderer also attaches the necessary documents proving the authority to bind the company (extracts from the articles of association, power of attorney, etc.).

In case a tender is submitted by a **consortium** of companies, an electronic signature must be supplied **for each participant** in the consortium **by the person(s) authorised or entitled to bind the participant**.

A qualified electronic signature can be given by means of a Belgian e-ID, or a qualified certificate that can be purchased from private parties.

For more information on purchasing a qualified certificate, see:

<http://overheid.vlaanderen.be/gekwalificeerde-certificaten>

Note: for foreign companies, the certificate cannot be in the name of the legal entity (electronic seal). In accordance with EU Regulation 910/2014 (eIDAS Regulation), this cannot result in a binding electronic signature of the tender.

For legal entities established in Belgium, however, signing by means of an electronic seal is possible, in view of art. XII.25. §3 of the Belgian economic law.

By tendering for this contract, the tenderer undertakes to use their movable and immovable property to perform the contract described in those specifications, in accordance with its terms and conditions. If a tender is submitted by a consortium without legal personality, each participant in the group is jointly and severally bound.

#### A.3.4. FORM AND CONTENT OF THE TENDER

The tenderer's attention is drawn to the fact that he must complete their tender and inventory on the form attached to these specifications.

The following is a non-exhaustive list of all the documents that must accompany the tender, in addition to the tender form:

- If not yet signed electronically when the request for participation is submitted:
  - o completed European Single Procurement Document for the tenderer;
  - o completed European Single Procurement Document for the subcontractors or other entities whose capacity the tenderer relies upon and for all participants in the consortium without legal personality;
- the necessary documents proving the authority of the persons placing an electronic signature to bind the company (A.3.3.);
- The documents relating to the assessment based on the award criteria (A.5) and Technical instructions, in particular:

- o A detailed price list for all elements included in A.5.1. A detailed price list for the required options as well as the allowed options, where applicable. Price lists must be supplied by means of a completed price matrix according to the template in **EN-appendix1-pricematrix.xlsx**. The template may not be modified;
- o The project plan for the delivery of the software, consisting of at least:
  - a proposal for an RACI matrix that defines the responsibilities between meemoo and the supplier;
  - a detailed project plan with timings and tasks for the delivery of the solution.
- o Answers to the business cases in III.4 and the requirements in III.7, consisting of:
  - a completed matrix with an answer to each mandatory requirement of the basic solution and options offered;
  - a short description of a possible solution per business case.
- o Description of the software development consisting of:
  - an architecture diagram and description of the components;
  - the overall roadmap with planned features, improvements and milestones;
  - the strategy or method of communication with the customer, software development, the release of product versions and guidance on updates;
  - documentation for users and developers, or a reference to it.
- o A contract proposal for support and maintenance (see requirements in III.8)
- o A Service Level Agreement (SLA) proposal (see requirements in III.9)

### A.3.5. COMMITMENT PERIOD (ART. 58 AWARDING RD)

Tenderers remain bound by their tenders for a period of ninety calendar days, commencing on the day after the deadline for receipt of the tenders.

The submission of revised tenders during the negotiations shall, in each case, lead to a renewed commitment period.

## A.4. PRICE

### A.4.1. PRICE DETERMINATION (ART. 26 AWARDING RD)

This contract is a mixed price contract:

- the basic solution (without options), the support contract and the integration and configuration services of the basic solution must be offered at a global price;
- options should be offered against a price list

#### A.4.2. QUOTATION (ART. 29 AWARDING RD)

If required for the accuracy of the unit prices, the tenderer may specify them to four decimal places.

The tenderer shall indicate the value added tax (VAT) as a separate item and include it in the price of the tender.

#### A.4.3. INCLUDED PRICE ELEMENTS (ART. 32, § 3 AWARDING RD)

The following costs, charges, services, etc. are included in the unit prices and global prices:

- travel, transport and insurance;
- the documentation related to the services;
- the training necessary for use.

#### A.4.4. PRICE OR COST INVESTIGATION (ART. 35 AND 37 AWARDING RD)

At the request of the contracting authority, the tenderer shall provide all the information necessary for the price investigation of their tender.

The contracting authority may either proceed itself or designate a person to carry out all verifications of the accounting documents and all on-the-spot inspections, in order to verify the accuracy of the information provided by the tenderer in the context of the price investigation.

### A.5. AWARD CRITERIA

#### A.5.1. NEGOTIATION PROCEDURE – AWARD CRITERIA (ART. 81 LAW)

For the award of the contract, the quality of the regular offers will be assessed on the basis of the award criteria below and in accordance with the assigned weighting percentages. The framework contract will be awarded to the most economically advantageous tender, from the point of view of the contracting authority, taking into account the award criteria.

When assessing the tenders, the contracting authority may ask the tenderers for clarification, if necessary, without this resulting in changes to the tender.

In order to clarify the criteria, the elements that will be taken into consideration during the evaluation are explained, without necessarily being exhaustive. The criteria are:

### **The price of the solution. (weighting: 60%)**

The total price at which the solution will be assessed is the sum of

- (70 pts) the **'total cost of ownership' of the software solution** over a period of **3 years**, consisting of:
  - the total price (incl. license) for an SaaS solution for the requested data volume and number of users (for requirements, see the growth path in III.6 and the functionalities in III.7) ;
  - the total hosting cost (insofar as not included in the SaaS price);
  - the total cost for installation, start-up, configuration of the solution on INT, QAS and PRD environments (for information about environments see III.2);
  - the total cost of a basic support contract (see requirements in III.7);
  - the total price for basic training of developers and administrators.
- (30 pts) a **budget estimate for services** based on an example scenario (see below), of which meemoo commits to a **minimum of 50 days** over the duration of the assignment.

This criterion is tested against the 'total cost of ownership' over a three-year period for the solution and services offered. The first year **starts on January 1st, 2023**. Any additional costs such as project management, administrative costs or others should be included in the price. Prices must be offered via a completed price matrix in the appendix (see A.3.4).

The cost-effectiveness of the solution will be assessed for each of the elements above using the following formula:  $P_t = P_{t,max} \times (P_{r,min} / P_{r,tender})$ , whereby:

- $P_t$  = points awarded to the element
- $P_{t,max}$  = maximum weighting of this element
- $P_{r,min}$  = lowest price of the quote found to be regular
- $P_{r,tender}$  = price of the tender

Then, meemoo will add up the two scores and reduce them to the weight of this criterion (see weighting).

If there is a suspicion of speculative pricing or abnormal pricing, meemoo reserves the right to investigate these prices and, if necessary, to reject the tender as irregular. See the rules on the regularity investigation and abnormal prices in the RD and section A.4.4.

*Consultancy services example scenario*

Since no specific assignments can be described at this stage, a total price for services cannot be requested. In order to make a price comparison possible, the budget will be based on the following hypothetical scenario:

A project consisting of 300 man-days, divided according to the following profiles and tasks.

- Senior Metadata Modeller: 10%
- Senior Analyst: 10%
- Junior Developer: 50%
- Senior Developer: 20%
- Project Manager: 10%

In your bid, include the price per profile (price per 8-hour day, including travel expenses) as well as the total price for this example scenario.

### **The quality of the solution (weighting: 40%)**

The following assessment elements are taken into account, without being exhaustive:

- completeness
  - the extent to which the basic solution and the required options meet the requirements set out in section III.7 by comparing the completed matrix with the specified minimum requirements;
  - the number of individual options allowed and the extent to which they meet the requirements of section III.7 by comparing the completed matrix with the specified minimum requirements;
  - the coverage of the provided solutions on the business cases from section III.4;
- transparency and interoperability
  - the open nature of the solution, APIs, data formats and query language, i.e. the specifications are openly licensed and/or the source code is open-source.
  - the extent to which the solution, APIs, data formats and query language are based on internationally applicable standards;
  - the existence of an exit path, in particular the extent to which the data in the system can be exported to another system;
  - the quality of the proposed project approach and software roadmap;
- capability
  - the performance and scalability of the solution under the circumstances described in III.4 and III.7, based on, among other things, the maximum data volume at which it can perform queries, example cases of similar production environments and an understanding of how the entire solution can be scaled;
  - the technical quality of the solution offered, in particular the quality of the APIs as evidenced by the documentation, the general ease of use, the existence of libraries, etc;
  - the ease of use of the solution, i.e. deployment, documentation, testing, APIs, graphical interfaces;
  - other additional features that can be related to the management, enrichment, import or export of the (meta)data processed and stored in the solution.

## A.5.2. NEGOTIATIONS

The contracting authority negotiates with one or more tenderers. The contracting authority has the option of conducting the negotiations in successive phases, thus limiting the number of tenderers with whom the contracting authority negotiates by applying the award criteria.

If a tender contains a substantial irregularity, the contracting authority will have this substantial irregularity regularised before the negotiations are started.

A tender that is submitted late or a tender that does not allow an initial substantive assessment cannot be regularised.

When the contracting authority wishes to conclude the negotiations, it shall notify the remaining tenderers accordingly. Remaining tenderer refers to the tenderers with whom the contracting authority is still negotiating at that moment. These remaining tenderers are invited to submit a final tender.

During the course of the negotiations, the contracting authority can indicate how the adapted tenders should be submitted. The submission of the final tenders will always take place via e-Tendering (see A.3.2. for more information).

The submission of intermediate tenders will take place by email.

The final tender is non-negotiable. If a final tender is substantially inconsistent, it cannot be regularised.

This version of the tender document concludes negotiations and is an invitation to submit final quotes.



## B.IMPLEMENTATION OF THE CONTRACT

### B.1. GENERAL IMPLEMENTATION CONDITIONS

#### B.1.1. SERVICE LOCATION (ART. 149 IMPLEMENTATION RD)

The services must be provided from the meemoo office in Ghent. In consultation, this service can also be partially provided remotely.

#### B.1.2. GUARANTEE (ART. 25 UPTO AND INCLUDING 33 IMPLEMENTATION RD)

For this contract, a guarantee is not required.

#### B.1.3. FIXED OR MINIMUM QUANTITIES (ARTS. 148 AND 151, § 5 IMPLEMENTATION RD)

For this contract, there are no fixed or minimum quantities.

### B.2. PAYMENTS

#### B.2.1. PAYMENT METHOD FOR THE PRICE (ART. 66 IMPLEMENTATION RD)

The software solution license and maintenance services are paid for with deducted annual payments. Other costs related to the software solution will be paid in percentage instalments according to the provisional and final delivery as described in B.6.1. This distribution is further determined in consultation with the candidate supplier, whereby at least 15% is paid at the start. For the time being, the following is assumed

- 30% at the start;
- 30% upon first provisional acceptance;
- 20% upon second provisional acceptance;
- 20% upon final delivery.

The services related to building integrations (i.e. effectively linking the software solution with meemoo applications or platforms that are in production) are paid for by deducted monthly payments.

### B.2.2. PROCEDURE (ARTS. 150, 156 AND 160 IMPLEMENTATION RD)

A verification period is not foreseen. The invoice shall serve as a claim.

Payment will be made within a period of 30 days from the date of receipt of the invoice, or 30 days after the termination of the services if the invoice is received before termination or the date of receipt of the invoice is uncertain.

### B.2.3. METHOD OF INVOICING

Invoices are sent via email to [boekhouding@meemoo.be](mailto:boekhouding@meemoo.be)

## B.3. CHANGES DURING IMPLEMENTATION

### B.3.1. PRICE REVISION (ART. 38/7 IMPLEMENTATION RD)

A revision of the prices for license fees and support contracts is not allowed.

### B.3.2. UNFORESEEN CIRCUMSTANCES ON THE PART OF THE SERVICE PROVIDER (ARTS. 38/9 AND 38/10 IMPLEMENTATION RD)

- a) Where the service provider can demonstrate that the contractual balance of the contract is disrupted to their detriment due to circumstances that are external to the contracting authority and that could not reasonably have been foreseen when the tender was submitted, that could not be circumvented and the consequences of which could not be remedied despite the service provider having taken all the necessary steps to that end, the service provider can claim the following revision:
- term extension, with the exception of a term extension for deliverables that can be linked to the import of the ERDF-GiVE metadata that has a December 2023 deadline. The precise deadlines for implementation are part of the negotiation.

- in the event of a very significant disadvantage, another form of revision (damages, for example), or termination of the contract.
- b) If the contractual balance is disrupted in favour of the service provider for any reason beyond the control of the contracting authority, the contract may be revised:
- either by shortening the implementation terms on the part of the service provider;
  - or when there is a very important advantage for the service provider, through some other form of revision or termination of the contract for the benefit of the contracting authority.
- c) The disadvantage or advantage incurred by the service provider is deemed to reach the threshold of the very significant disadvantage/advantage if the disadvantage or advantage amounts to at least 15% of the initial order amount.

#### **B.3.4. REPLACEMENT OF THE SERVICE PROVIDER IN THE EVENT OF BANKRUPTCY (ART. 38/3 IMPLEMENTATION RD)**

In the event of bankruptcy of the service provider, the contract can be transferred to a company proposed by the bankruptcy trustee, for example, to the subcontractors. However, the contracting authority has the right to terminate the contract and start a new contract award procedure.

### **B.4. INTELLECTUAL RIGHTS AND CONFIDENTIALITY**

#### **B.4.1. INTELLECTUAL RIGHTS AND KNOW-HOW (ART. 19 AND 10 IMPLEMENTATION RD)**

*The contractor shall transfer to the contracting authority all their property rights in the computer program developed in the performance of their contract (hereinafter “the developed software”); the transfer will also include the property rights in the material accompanying and preparing the developed software, such as documentation, design, analysis, and any other literary work and/or document stored in a durable manner or in machine language.*

*The transfer of all property rights applies to the contractor and to all persons on whom the contractor relies, such as their staff or a subcontractor, or will rely on for the implementation of the contract. The compensation for this transfer of rights is included in the total amount of the offer.*

*The contractor grants the contracting authority permission to disclose the computer program developed for the implementation of this contract to the public under the name of the contracting authority and to exploit it under that name.*

*The contractor guarantees to have all the rights and the necessary authorisations to transfer the aforementioned property rights to the contracting authority and that this transfer will not infringe the intellectual or any other rights of third parties.*

*Without additional costs, the contractor undertakes to make the source code of the developed software available to the contracting authority in the form of a usable development and production environment, and to keep it permanently up-to-date, as well as a copy of the preparatory and accompanying material (including all the technical specifications).*

*If the use of the developed software also requires the use of third-party software, the contractor will ensure that the necessary user rights for the third-party software are also transferred with the developed software transfer.*

*The contracting authority will grant the contractor a free, non-exclusive, and non-transferable right of use for the developed software, to the extent and for the duration necessary for the performance of the contract. Subject to the prior consent of the contracting authority and under the same conditions as those that apply to the contractor, the contractor may sub-license the use of the software referred to in the first paragraph to their subcontractors to the extent and for the duration necessary for the performance of the contract.*

#### **B.4.2. EXISTING INTELLECTUAL PROPERTY RIGHTS (ART. 30 AWARDING RD)**

The tenderer is required to indicate in their tender the intellectual property rights of which they are the holder or for which they must obtain a licence from a third party for all or part of the services to be provided.

The purchase price and the fees payable for the licenses to use these intellectual property rights must be included in the prices offered.

#### **B.4.3. CONFIDENTIALITY (ART. 18 IMPLEMENTATION RD)**

The information provided by the contracting authority in the context of this contract may not be used for other purposes or communicated to third parties.

The service provider shall take all necessary measures to ensure that the confidentiality of the information, data and research results provided, is stored securely by both themselves and by anyone who has access to them.

The service provider shall also include these confidentiality obligations in their contracts with the subcontractors.

#### **B.4.4. PROCESSING OF PERSONAL DATA**

The implementation of this contract involves the processing of personal data. From 25 May 2018, the requirements of the General Data Protection Regulation (hereinafter: GDPR) must be followed.

See the processing agreement in Annex 1 (in Dutch; translation can be provided upon request after awarding).

### **B.5. PENALTIES AND LEGAL ACTIONS**

#### **B.5.1. LEGAL ACTIONS (ART. 73, § 2 IMPLEMENTATION RD)**

Any legal action by the service provider is brought before a Dutch-speaking Belgian court, except in the case of an action for intervention in an existing lawsuit.

### **B.6. INSPECTIONS AND DELIVERY**

#### **B.6.1. DELIVERY (ART. 64 AND 156 IMPLEMENTATION RD)**

The delivery of this contract consists of 3 milestones:

Milestone	Date	Criteria
<b>first provisional delivery</b>	01 February 2023	<ul style="list-style-type: none"><li>the software solution is available on INT, QAS and PRD environments (see III.2);</li><li>there is a production-ready GraphQL implementation for the link between the platforms and the Knowledge Graph (either by providing Option A or by supporting an implementation by meemoo);</li><li>there is a production-ready ETL system to load the data needed for the platforms into the Knowledge Graph (either by providing Option B or by supporting an implementation by meemoo);</li><li>there is a production-ready GraphQL API for metadata about organisations;</li></ul>

		<ul style="list-style-type: none"> <li>the load test*, as agreed with the service provider for this phase, has been successfully performed.</li> </ul>
<b>second provisional delivery</b>	01 April 2023	<ul style="list-style-type: none"> <li>there is a production-ready (GraphQL) API for the Archives 2.0 version 3;</li> <li>the load test*, as agreed with the service provider for this phase, has been successfully performed.</li> </ul>
<b>final delivery</b>	01 June 2023	<ul style="list-style-type: none"> <li>there is a production-ready IIIF 3.0 API for the dissemination of high-resolution ERDF-GiVE images;</li> <li>the metadata generated by the ERDF-GiVE projects (GiVE-1452) on facial recognition, speech-to-text and named-entity recognition are included in the Knowledge Graph;</li> <li>the load test*, as agreed with the service provider for this phase, has been successfully performed.</li> </ul>

\* A load test is performed on the PRD environment and covers the path from a GraphQL query to the graph data storage. Meemoo will prepare the set of queries (the 'load') together with the candidate supplier and base it on a real scenario of one of the meemoo access platforms.

The requirements for the provisional and final delivery can still be refined in the negotiations.

The dates are indicative and will be agreed by mutual consent.

From the date of full delivery, determined in accordance with the rules of the contract documents, the contracting authority has 30 days to complete the delivery formalities and to notify the service provider of the result. This period will start insofar as the contracting authority is simultaneously in possession of the list of services performed.

If the services are finished before or after this date, the service provider will notify the leading official by registered letter and ask them to proceed with the delivery. In such case, the thirty-day period starts from the date of receipt of the service provider's request.

## B.7. IMPLEMENTATION CONDITIONS

### B.7.1. NON-DISCRIMINATION

In implementing this contract, the service provider undertakes to not discriminate on the basis of gender, age, sexual orientation, marital status, birth, wealth, religion or philosophy, political opinion, language, health condition, disability, physical or genetic characteristics, social position, nationality, alleged race, colour, descent, national or ethnic origin or trade union affiliation. They shall ensure this

both among their staff and with regard to third parties, such as participants, visitors, external collaborators, etc.

As far as is reasonable, the service provider undertakes to make adjustments at the request of persons with disabilities, that offset the limiting influence of an unsuitable environment on the participation of a person with a disability (see article 19 of the Decree of 10 July 2008 on a framework for the Flemish policy on equal opportunities and equal treatment).

The service provider undertakes to inform employees and third parties, including participants, visitors, external collaborators, etc. that they will not take into account questions or requirements of a discriminatory nature.

If a staff member of the service provider is guilty of discrimination, bullying, violence, or sexual harassment during the execution of the contract, the service provider will take the necessary measures to put an end to this behaviour and, where necessary, restore the victim's honour. The employees with hierarchical responsibilities shall ensure compliance with this commitment.

In the event of any complaint against the service provider in this respect, the latter will fully cooperate with any investigation that is conducted by a discrimination contact point, or any other organisation appointed by the Flemish government.

The service provider also asks all its staff members to be alert to discrimination, bullying, violence, or sexual harassment, in the sense that they should immediately report the cases they witness to an employee with hierarchical responsibility.

The service provider undertakes not to put pressure on their own staff, who are victims of discrimination, bullying, violence, or sexual harassment by a customer or a third party, to refrain from submitting a complaint or initiating a legal action before the court in this regard.

The service provider ensures that the subcontractors that they may engage for the contract, also comply with these implementation conditions.

# III. TECHNICAL INSTRUCTIONS

## 1. CONTEXT

### 1.1. ABOUT MEEMOO, FLEMISH INSTITUTE FOR ARCHIVES

[Meemoo](#) is a non-profit organisation that, with the help from the Flemish Government, is committed to the digital archiving of organisations in culture, media, and government. Together with our partners, we bring the past back to life and prepare it for the future. We safeguard their archive content digitally and make it accessible and useable. Meemoo's core tasks are:

- Digitisation, digital preservation, and management
- Making content accessible;
- Gathering and sharing knowledge;
- Advice about digital cultural heritage processes;
- Setting up collaborative projects with (content) partners and other stakeholders.

Meemoo is a service provider in the field of digitisation, archiving and accessibility, with a focus on service and customer orientation. As a result, meemoo has years of experience in setting up large-scale digitisation projects in which efficiency, economies of scale, and quality are paramount. Until now, the focus has mainly been on audio-visual material (moving images and sound).

At present, meemoo's main clients (also known as *content partners*) are:

- The Flemish commercial, public and regional broadcasters;
- Flemish cultural heritage institutions: the institutions recognised and subsidised by the Cultural Heritage Decree:
  - the institutions recognised by the cultural heritage decree;
  - city archives of the main cities.
- Heritage units;
- Flemish public administrations;
- Arts organisations from the Flemish performing arts sector.

This list may be extended or modified at any time.<sup>1</sup>

---

<sup>1</sup> An up-to-date overview of meemoo's content partners can be found on <https://meemoo.be/nl/contentpartners>.



## 1.2. METADATA MANAGEMENT AND INTEGRATION

In addition to its role as an archive, meemoo also has the role of being a data aggregator: not only does it currently collect audio, video, and images, but it also collects the metadata of more than 170 content partners. Thanks to the presence of this metadata, the archive material is searchable across all collections of the various content partners. This is the driving force behind various forms of access and the possibility for content partners to curate their own archive material. But also for meemoo, metadata is crucial for the creation and implementation of efficient internal processes and services.

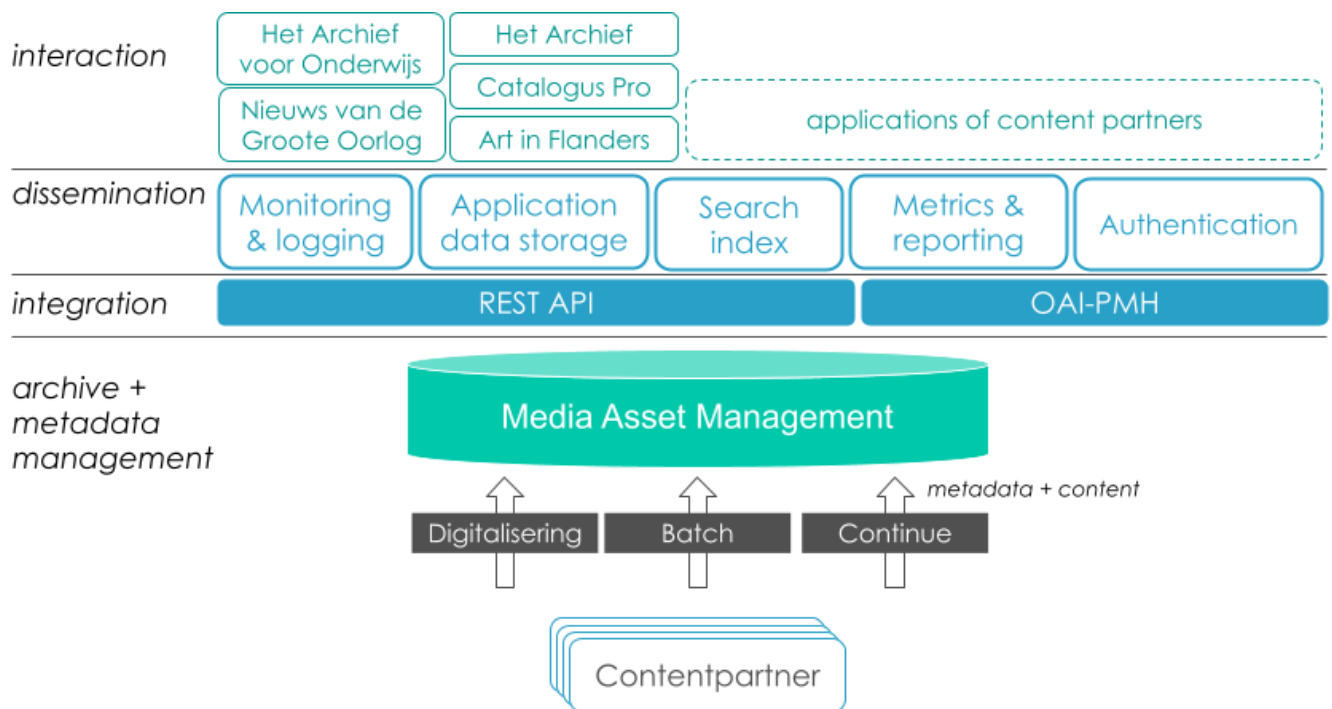
However, this metadata is not imported in a uniform way. Most content partners have their own collection management systems and their own (meta)data structure in which they organise information within their collections. The result is a variety of different types of information, data models and file and metadata formats. Therefore, meemoo also has the less visible and rather underexposed role of data integrator: it standardises the various metadata so that they can be accessed in an unambiguous way.

Due to the general increase in the amount of data, the complex task of data integration has become one of the biggest challenges in various domains in recent years. Also at meemoo, metadata is growing and changing continuously due to new (types of) collections, additions, new insights, and corrections by content partners. In the future, efforts will also be made to automatically catch up with a large backlog of unannotated material through AI applications. To do this, meemoo must be able to manage and process this amount and variety of metadata correctly.

Specifically, meemoo has recently started a machine learning project under the impetus of ERDF resources, which will work on a large scale with speech recognition, face detection, facial recognition, and entity recognition on text (ref. GiVE-1452). In particular, meemoo wants to

- perform speech recognition on 160,000 hours of audio
- perform entity recognition on the texts from speech recognition
- perform facial recognition on a select set of people in 88,000 videos.

Importing this data is one of the most important challenges and milestones within these tender specifications. The data is currently being created and needs to be stored and managed in the knowledge graph for searchability and accessibility.



**Figure 1: Current infrastructure for metadata management with a central Media Asset Management system**

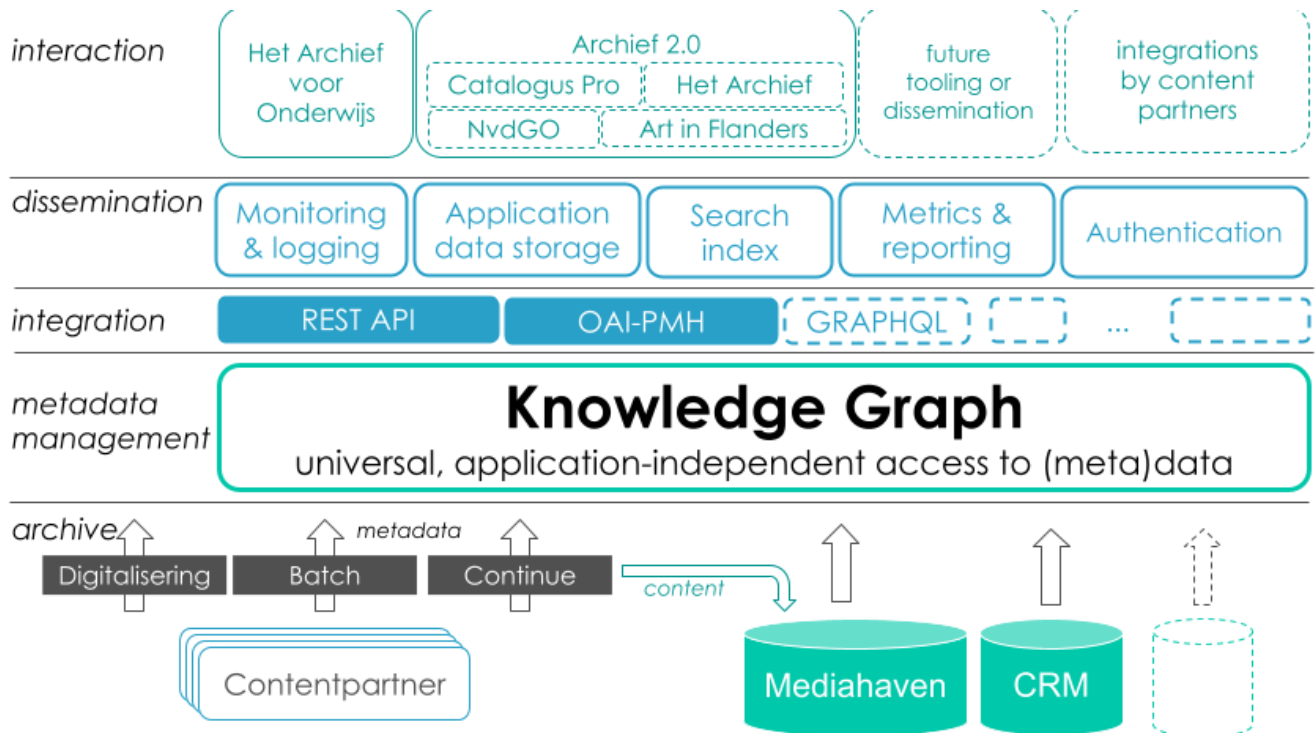
As illustrated in Figure 1, our current infrastructure is highly centralised around the Media Asset Management system (MAM) with a common access infrastructure above it, consisting of a series of reusable components. Several applications either let metadata flow in or take metadata out:

- the platforms for internal use, including the AMS registration system or the digital intake tool;
- the access platforms for external use, including The Archive for Education and The Archives;
- or
- external systems of content partners, for example, a self-organised solution with spreadsheets or a collection management system such as Adlib.

These metadata flows are organised as a customised translation between the MAM and the connecting platform. So, our systems are linked with point-to-point integration and these are difficult to scale. Not only does each new system create a separate, maintainable integration, but they are also very fragile: any change on either side means a software development, configuration, and testing effort.

Different platforms and applications organise and give meaning to (meta)data based on the functionality they offer the user. They are often not equipped to exchange their data on a large scale and, therefore, form so-called data silos. This inevitably leads to a difficult migration of data between two different platforms with complex and error-prone data transformations. This also causes confusion among the users of these platforms, both within meemoo and with external partners. Information that does not fall within the logic of the application has difficulty finding a place in the metadata model of that application. As a result, a lot of knowledge within the organisation often cannot be converted into

interpretable metadata. Therefore, a lot of knowledge remains implicit, and this creates dependencies between people, teams, or organisations.



**Figure 2: Intended infrastructure for metadata management with a Knowledge Graph as the universal, application-independent view of metadata.**

A Knowledge Graph solution should facilitate sustainable integration of the various systems by creating an unambiguous representation and interpretation of metadata, as illustrated in Figure 2. These tender specifications are part of making the following sustainable, robust, and uniform:

1. **metadata storage and organisation:** choosing database technology and framework;
2. **the machine-readable representation of metadata:** the design and use of data models and formats;
3. **the integration of metadata with user applications:** the way in which (internal) tools or access platforms can search, add, modify, and retrieve metadata.

### 1.3. ACCESS PLATFORMS

As can be seen in the architecture sketch in section 1.1, the metadata supports multiple access platforms. These platforms typically open up a different sub-collection from the underlying archive system to a specific target group, such as education, research or the wider public, and are developed

by a third party on behalf of meemoo. The table below lists all the platforms that are currently operational or under development:

Platform	Description	# records	Launch date	Target group
Archive for Education	Access to audio-visual material for use in the classroom. Visitors can create limited content here, such as teaching assignments and collections.	25K	January 2020	Teachers or people working in the educational sector.
Hetarchief.be	Access of metadata and/or audio-visual material to the general public.	1.7M	2016	People with an interest in heritage, amateur documentary makers, scientists, non-experts, casual visitors.
News of the Great War	Access to digitised newspapers with metadata from the First World War	52K	2018	Historians, or people with a specific interest in WWI.
Catalogus Pro	Professional reuse of metadata and audio-visual material between a select group of content partners.	1.7M	2018	Collection managers at the content partners.
Art In Flanders	Image library with high-resolution images of art and heritage for commercial and non-commercial use.	22K	2018	Professional re-users of works of art, for example, for printing.
Visitor tool	Access to the 'personal' collection via an application that can only be used in the reading room of the content partner.	2M	June 2022	The visitors of the individual content partner.
Archive 2.0	New platform for making metadata and/or audio-visual material available to the general public, professional reuse and in the reading room of the content partner. Integrates and eventually replaces Visitor Tool, hetarchief.be, Catalogus Pro and Art In Flanders.	2-2.5M	April 2023	All of the above with the exception of teachers or people working in the educational sector.

The platforms operate on a shared infrastructure consisting of the MAM, an Elasticsearch search index, and an SAML authentication server, complemented by an application database in PostgreSQL and ETL and synchronisation services. Currently, the relevant subset of metadata from the MAM is first synchronised to the application database and presented via a Hasura GraphQL Engine. From the application database, the metadata is indexed in Elasticsearch.

## 1.4. METADATA MODEL

The metadata of content partners are classified in the current meemoo metadata model upon inflow. This metadata model has the structure of a flat list with approximately 178 possible fields and is used in XML. Some fields are keys nested under the same field or contain values that come from a controlled list. All information about this model can be found at <https://developer.meemoo.be/docs/metadata/> (only in Dutch).

A number of domain models are currently under development to replace the current metadata model. Each domain model is conceptually elaborated and then formalised into:

1. a light ontology expressed in [RDFS](#) with all classes and properties;
2. an application profile expressed in [SHACL](#), which lays down the rules and limitations of use;
3. (If applicable) a taxonomy expressed in [SKOS](#) for controlled lists of terms.

The domain models refer to each other and reuse existing schemas, ontologies, or standards as much as possible. The table below provides an overview:

Domain	Information about	Base
<a href="#">Organisations</a> (in development)	Organisations, content partners, contracts, contacts, addresses and account managers	<a href="#">W3C Organization Ontology</a>
<a href="#">Objects</a> (in development)	Structure of archival objects, digital reproductions, media copies (archive master, mezzanine, low resolution), file integrity, identification, and the relationship with the original/content	<a href="#">PREMIS OWL 3</a>

Right of use (planned)	Licences and permissions of meemoo and users on the material/metadata.	<a href="#">ORDL 2.2</a>
<a href="#">Searchability</a> (in development)	Search terms and content description (Title, Keywords, Copyright).	<a href="#">Schema.org</a>
<a href="#">Source</a> (planned)	Entry process, the digitisation chain, the source of the metadata and when the metadata was modified and by whom.	<a href="#">PROV-O</a> <a href="#">PREMIS OWL 3</a>
Education (in development)	Learning Object Model, educational structure, key competencies, building blocks and subjects.	<a href="#">SKOS</a> <a href="#">Vlaamse onderwijs</a>
Registration (to be planned)	Analogue carriers, logistics and quality control.	<a href="#">EBUCore</a> <a href="#">EBU:StorageMediaTypeCodeCS</a> <a href="#">PBCore</a>
Digital media (to be planned)	Media encoding (codec) and format (media type, aspectRatio, imageSize)	<a href="#">EXIF</a> <a href="#">Ontology for Media Resources</a> MPEG
Projects (to be planned)	Collaborations and project structure	<a href="#">DINGO</a> <a href="#">FOAF</a>

## 1.5. INFLOW

Material and metadata can flow in on a project basis via the following processes:

- **digitisation of inflow:** content partners first register a series of analogue carriers in meemoo's registration tool (AMS) and deliver this material to meemoo. Then, meemoo has

this material digitised by an external supplier and imported into the meemoo archive system. The metadata is based on the registration and the digitisation process;

- **digital batch inflow:** the content partner delivers digitally born material with metadata as a defined set to meemoo, who then imports it into the meemoo archive system;
- **digital continuous inflow:** the content partner does a regular upload of digitally born material with metadata to the meemoo archive system.

The way in which material and metadata should be offered to the meemoo archive system is determined by the SIP definition that is specific to the project or media type:

- In essence, for each object, there is always a media file (essence) accompanied by an XML file (sidecar) that contains all metadata.
- A complex object involves multiple media files that are accompanied by a [METS-XML](#) that contains the file structure, the descriptive metadata and [PREMIS](#) information about the source, digitisation, or inflow process.

A uniform SIP specification is currently being worked on based on [E-ARK standard](#); see [https://developer.meemoo.be/docs/diginstroom/sip\\_specification/](https://developer.meemoo.be/docs/diginstroom/sip_specification/). Meemoo is also considering an alternative based on a [JSON\(-LD\)](#) format.

## 1.6. DATA WAREHOUSE

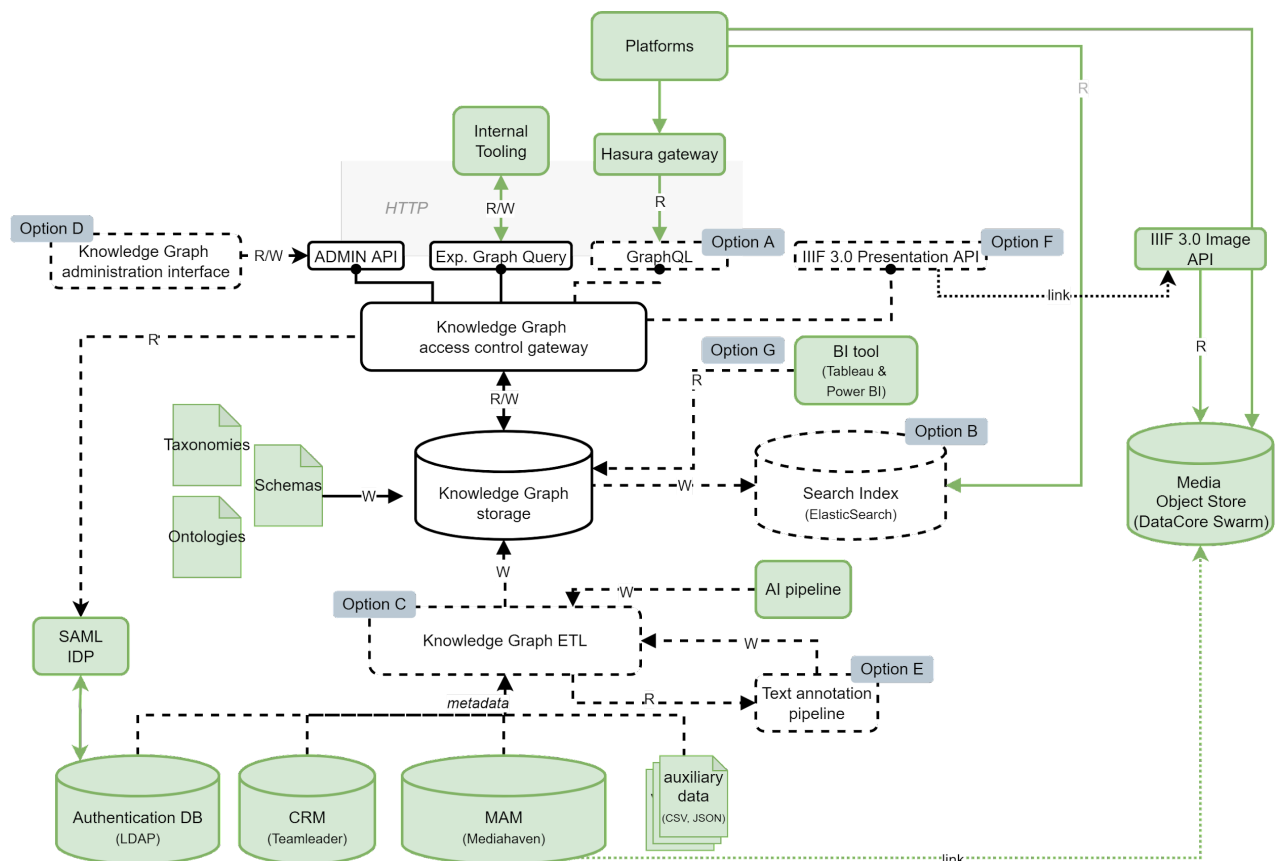
As part of the KPIs and the collaboration with the content partners, meemoo regularly publishes reports with figures (for example, on the state of the archive and use of the platforms). The reports are prepared in the BI applications Tableau and Microsoft PowerBI, which retrieve the data from the meemoo data warehouse.

The data warehouse integrates the main and relevant data sources with an ETL process (Python/SQL) in a PostgreSQL database divided into various facts and dimensions. The following data, among others, is integrated:

- metadata about analogue carriers from the registration tool (AMS);
- descriptive and technical metadata about archival objects from the MAM system;
- data on organisations and people from the CRM system (Teamleader) and LDAP;
- user actions from the application databases of the access platforms (PostgreSQL);
- content created by visitors to The Archive for Education (AvO);
- metadata from various processes such as video transcoding, digitisation...

## 2. ARCHITECTURE

The following figure outlines a possible architecture that positions the software solution within the existing software infrastructure of meemoo. Although they are drawn as separate components, they can also be offered as one or more integrated wholes.



**Figure 3: Example of a possible architecture and context. The green components and arrows are offered by meemoo. The white components and solid arrows are part of the requested basic solution. Components and arrows marked with a dotted line are part of the options.**

Components or data sources offered by meemoo:

Internal Tooling	Applications to support meemoo's internal processes, including extracting and importing material, logging and diagnostics, validation, etc.
Platforms	Web platforms for accessing metadata and archive material (see III.1.2)



Taxonomies, ontologies, and schemas	Machine-readable definitions, multilingual labels and descriptions of classes, relationships, attributes, controlled lists, and restrictions used to model the metadata.
Search Index	An Elasticsearch instance that indexes a subset of the metadata and supports the search functions in the platforms.
SAML IdP	The SSO SAML2.0 Identity Provider that authenticates users to applications. Meemoo uses SimpleSAMLphp ( <a href="https://simplesamlphp.org/">https://simplesamlphp.org/</a> ).
Authentication DB	An LDAP containing user data for authentication.
BI tool	The applications Kibana, Tableau and PowerBI for creating reports and dashboards.
CRM	The customer relationship management system Teamleader for managing information about staff, content partners, suppliers, etc.
Auxiliary data	Structured information that is not managed in an application, but in separate files such as spreadsheets. Currently there are no actual files in scope here, but this may change later.
AI pipeline	A process for automatic annotation of archive material consisting of facial recognition, speech recognition... generating metadata to be integrated.
Media Object Store	On-premise <a href="#">DataCore Swarm</a> instance for delivering media archive objects over the S3 protocol.
IIIF 3.0 Image API	Standardised API according to the <a href="#">Image API 3.0</a> specification that provides access to high-resolution images.

Components that are part of the requested solution, whether as an option or not (see III.7 for the specific requirements):

Knowledge Graph Storage	Component where the metadata, thesaurus, ontologies, and schemas are available as a knowledge graph. Preferably, this is one or more databases where metadata is indexed for performance reasons, but this is also partially filled in with a virtualisation layer if the situation so requires.
Admin API	An (HTTP) API for administrative functions such as creating, modifying, emptying, or deleting databases and/or graphs; user management and access control; and statistics on status, performance, and CPU/memory usage.
Knowledge Graph administration interface	A graphical application that facilitates interaction with the Admin API in a user-friendly and non-technical way.
Expressive Graph	An HTTP API with which the underlying graph data can be queried with structured

Query	queries in SPARQL, OpenCypher, GQL or another expressive query language.
GraphQL	A GraphQL API with which the underlying graph data can be queried.
Knowledge Graph ETL	An ETL framework or application that can convert large volumes of JSON, XML and RDB data into graph data (based on the taxonomies, ontologies & diagrams) and load them into the Knowledge Graph.
Text annotation pipeline	A component or process that recognises entities in Dutch text, links them to the entities present and thus enriches the existing Knowledge Graph.
IIIF 3.0 Presentation API	Standardised API according to the <a href="#">Presentation API 3.0</a> specification that provides access to information about digital representations of (heritage) objects.

Systems within meemoo are always offered in three environments:

- **Integration (INT):** the environment for testing during software development
- **Quality assurance (QAS):** the staging environment in which the latest changes are tested before they go into production
- **Production (PRD):** the environment in which the system runs in production.

### 3. SCOPE

The scope of this assignment consists of two parts:

1. the supply of a **software solution**, consisting of:
  - a. a basic solution;
  - b. 10 possible options;
  - c. integration into the meemoo access infrastructure.
2. the **provision of consultancy services** after delivery.

#### SCOPE OF THE SOFTWARE SOLUTION

The scope of the basic solution and the options consists of offering a high-performance solution for the application-independent storage of metadata, adapted to the high volume of metadata, the multitude of users and the necessary flexibility in querying and modelling. In specific terms, this is limited to the requirements in III.7.

The scope of the integration includes making the software solution production-ready for use in the access platforms and is described in III.10.

## SCOPE OF THE CONSULTANCY SERVICES

The specific scope of the consultancy services will be determined in consultation with the service provider and may include, for example, the following activities:

- integrating and/or developing AI pipelines for the automatic annotation of archive material;
- integrating the Knowledge Graph into the backend of our web platforms or internal applications;
- redesigning our architecture, including determining the relationship and integration between the MAM, the Knowledge Graph, and the object store (media storage location).

## NOT IN SCOPE

- development of data models, thesaurus, ontologies and diagrams;
- integration with the inflow process or input formats such as the SIP definitions (excluding making the ETL solution available, if offered);
- development of or integration with the web platforms for accessibility to end-users;
- offering or developing a Knowledge Graph administrative interface where, for example, content partners or other external users can consult and edit metadata;
- offering and managing an identity provider;
- complete replacement of the data warehouse;
- processing the archive material itself.

For the above, meemoo is always looking for support and, therefore, an appropriate framework in the form of software libraries, documentation, (graphic) tools and technical expertise is part of the scope.

## 4. BUSINESS CASES

### 4.1. MAKING METADATA ACCESSIBLE TO USERS

The metadata of the archive material is made accessible via a number of web platforms. Visitors to these platforms search for archive material using a number of 'traditional' search functions and navigation, such as:

- full-text search, mainly keyword searches, boolean operators...;
- filters and faceted search to refine search results;

- sorting search results by date, relevance, or alphabet;
- browsing within a category (for example, audio, video, photo);
- choosing from controlled lists (for example, different providers).

Therefore, it is crucial to be able to offer such basic functionalities in a high-performance manner. As each platform focuses on a specific target group (see III.1.3), this functionality can differ per platform or increase in importance.

In addition to traditional search, meemoo also looks at explorative search: the user does not (yet) have a specific goal and is looking to make discoveries. Based on meaningful relationships in the metadata, users navigate between archive objects and other entities by following links or searching paths.

Opening a search result usually yields a detail page, containing (substantive) information about the archive object (for example, title, description), the reproductions in the archive (for example, high-resolution and/or low-resolution copy), the context (for example, actors, producers, provider, rights) and underlying or related archive objects. The design of such a page depends on the media type. For example, a television programme, a film, and a newspaper are ideally presented in a different and customised way, with attention to user-friendliness. In addition, metadata and field labels can also be multilingual. A platform must be able to deal with this dynamically.

Finally, the metadata is also to be accessed via an API to enable the building of integrations. Depending on the nature of the integration, the metadata is offered with a different protocol (OAI-PMH, REST, GraphQL) or in a different metadata model (for example, [Europeana Data Model](#) for synchronisation with Europeana; [CIDOC-CRM](#) for scientific access; [Schema.org](#) for SEO). In addition, it should also be possible to easily create a relevant cross-section of the metadata, tailored to the integration (for example, only photo material of paintings). Here too, it is essential to be able to take into account the rights that apply to the material or the metadata (see III.4.3).

## 4.2. FLEXIBLE INFLOW OF METADATA

Meemoo aggregates metadata from 170+ partners from different sectors that differ in:

- the nature of their activities: broadcasters, archives, heritage libraries, government institutions, heritage cells, performing arts organisations, museums;
- the meaning and name of certain metadata fields;
- the information that is or is not relevant and, therefore, is or is not kept.

In order to avoid errors or loss of information during inflow, the metadata model of the archive system must be able to deal with this diversity. As much information as possible must get a place. In addition, different data models must also be able to co-exist. In this way, fields that do not immediately have a place can still be saved and used later.

Validation is also essential here. In order to monitor the integrity of the archive, metadata that does not meet expectations must be detected automatically upon inflow.

The inflow of metadata also needs to be sufficiently flexible to scale to future use cases. Meemoo plans a large-scale addition and enrichment of metadata with AI techniques:

- automatic annotation of text, images and audio-visual material;
- automatic linking and associating of entities in the metadata;
- Automatic deriving and completing of metadata based on rules.

### 4.3. FINE-GRAINED RIGHTS AND USER RESTRICTION

The material and metadata are subject to a complex rights structure, which is why by default, they are not made available to the public or to other content partners. Therefore, the public [hetarchief.be](https://hetarchief.be) only contains a limited set of metadata and no playable media.

*An archive object on [the Art In Flanders platform](#) is a digital reproduction of a painting. Therefore, there is copyright on the painting and the photographic registration with two different rights holders. In addition, the object is also transferred to meemoo with a licence that indicates on which platform (i.e. for which target group) the object and/or the metadata may be made accessible. This also determines the user rights of visitors to this platform, i.e. what they can do with it (depending on the platform): view, play, download in low resolution, download in high resolution, etc. So, the painting can be in the public domain, while the photo is still copyrighted. Despite the public domain status, the content partner can also choose not to disclose the image. Currently, this information is very sparse or unstructured in the metadata, which makes it unusable in queries and the logic is contained in the code of platforms.*

**Figure 4: Example of a problem related to rights**

However, meemoo's task is to bring as much material as possible into the public domain and thus to make public what can be made available. Some of the material can also be made available to the education sector on the basis of copyright exceptions. In the private sphere, we still have professional reuse, where a limited group of content partners are given access to each other's extensive set and the archive material itself.

Therefore, an access platform must be able to easily access cross-sections based on:

- one, several or all responsible content partners;
- the copyright on the material and/or individual metadata;
- the explicit right of use that meemoo receives from the content partner.

Since the metadata can also be made accessible via API, these cross-sections must also be made accessible in this way.

In time, granular permissions for the users of the platforms will be added, i.e. what the user can do with the material and under which licence (for example, view for home use, but no downloading).

## 4.4. INDICATION OF DATA ORIGIN

Metadata originates from the content partner, but can be enriched by meemoo during and after archiving:

1. During the inflow process, metadata relating to the registration of the carrier, the digitisation process and the preservation of the material is added.
2. To increase retrievability, domain-specific metadata (for example, events and artists in the performing arts) from external databases is integrated.
3. Content partners make adjustments and additions once the material is in the archive.
4. To catch up, a large volume of material is automatically annotated by AI techniques.

*Example 1: For archive material from the performing arts or archive material from digitised film, meemoo relies on external specialised organisations for metadata. Afterwards, it is often not clear whether a title, description or keyword was entered by the content partner, meemoo, or the external organisation. This sometimes leads to conflicts about certain categorisations of material (for example, dance or cabaret) or a lack of source reference. This information is not provided for in the data model and cannot be properly maintained. Therefore, different perspectives cannot coexist, which is sometimes desirable.*

*Example 2: Meemoo performs facial recognition on photo material from different content partners and thus generates an inaccurate list of people per photo. Some content partners want this metadata to be submitted to them first for validation. Others want this metadata to be part of the search functionality but marked as inaccurate.*

Figure 5: Examples of problems related to data origin

For reasons of liability and source attribution, it is essential to be able to distinguish which metadata was created by whom and when. This also enables a transparent administrator application. It must also be possible to indicate the reliability, in the case of automatically generated metadata, for example, so that it can serve as a search criterion or as an indicator for a manual validation of the metadata.

## 4.5. SHARED GLOSSARIES AND VOCABULARIES

Several applications at meemoo rely on shared controlled lists of terms. They are either managed separately from an application (for example, in a spreadsheet or documentation page), or managed from the authoritative source system (for example, the analogue material registration tool AMS, which contains a list of analogue media).

These lists are mainly used:

- to clarify what the possible values are for a particular field;
- to ensure that the same value is used everywhere;
- to ensure that the same label is always shown.

Some applications also make use of the (hierarchical) relationship between the terms, mainly when classifying and suggesting archive materials for education.

It is important that these lists can be reused efficiently. The meaning, labels and interrelationships of the terms must be clear and unambiguous in all places. They must also be easy for any system to read and use. For the time being, we are trying to use [VocBench](#) for this.

## 5. PLANNING

The planning of this contract is similar to the renewal of the [hetarchief](#) web platform. This new version includes the current platform, Catalogus Pro, the visitor tool, and access via IIIF (see III.1.3). The development will take place in 3 phases, with the delivery of phase 1 (visitor tool) scheduled for April 2022. The retrieval of metadata from a Knowledge Graph is provided in phase 3: access via IIIF (June 2023) and the integration of the Catalogus Pro platform (April 2024). Therefore, these last two milestones determine the delivery date of the solution.

The following planning is used for this dossier:

Step	Activity	Deadline
1	Publication of the tender specifications	13 May 2022
2	Q&A session <ul style="list-style-type: none"> <li>• <b>deadline questions by email</b></li> <li>• live session by video call</li> <li>• publication of answers on TED</li> </ul>	<b>25 May 2022</b> 01 June 2022 02 June 2022
3	<b>Submission of offers</b>	<b>21 June 2022 10:00 CET</b>
4	First negotiation phase and offer selection <ul style="list-style-type: none"> <li>• Publication of additional specifications</li> </ul>	21 June up to 12 September 2022 12 September 2022
5	<b>Submission of Best and Final Offer (BAFO)</b>	<b>30 September 2022 10:00 CET</b>
6	Planned award	<b>October 28, 2022, at the latest</b>
7	First provisional delivery (see B6.2)	1 February 2023 (subject to change)
8	Second provisional delivery (see B6.2)	1 April 2023 (subject to change)
9	Final delivery (see B6.2)	1 June 2023 (subject to change)

Further clarification of the procedure:

- Step 4: meemoo will select a maximum number of candidate suppliers from the offers received in step 1 (see A.5). The selection of these offers will be made according to the criteria that you can find further on in this document.
- Step 5: If necessary, a second selection round is organised with the selected candidate suppliers. At the least, this will include a presentation of the offer and the solution. Based on discussions with this supplier, a written addendum to the original quotation will be distributed. Based on the original documents and this addendum, the candidate supplier must submit a best and final offer (BAFO).
- Step 7: meemoo will evaluate the amended offers based on the criteria stated in this document to award the contract to the candidate supplier with the highest score.

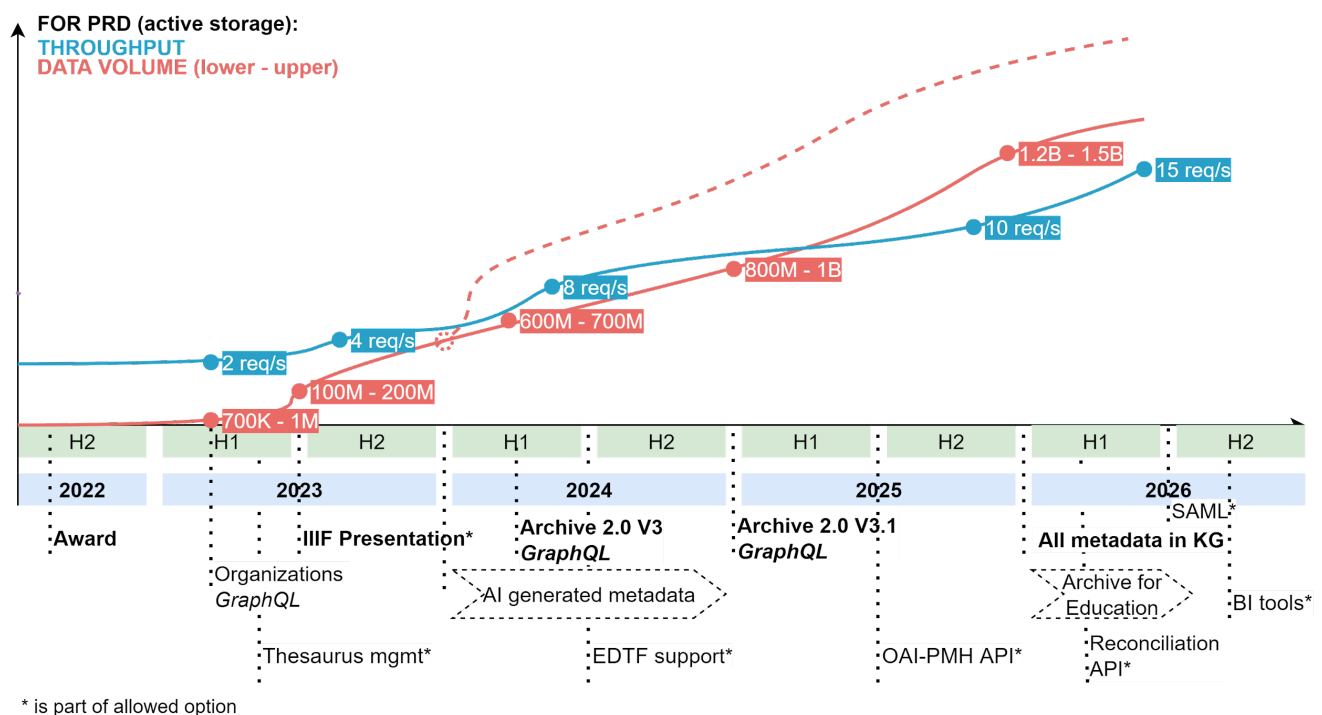
Meemoo reserves the right to make changes to the above procedure as long as they do not harm competition as determined by law and are communicated transparently to the (selected) candidate suppliers.



## 6. GROWTH PATH

The software solution will be gradually integrated into the software infrastructure of meemoo. The size and use of the Knowledge Graph starts with the minimum viable and grows steadily throughout the duration of the assignment. In addition, sudden increases due to unforeseen opportunities, of which the magnitude and quality of the results are still unknown (e.g. the ERDF-Give AI annotation projects) are possible.

Figure 4 illustrates the projected growth path based on the key milestones for meemoo, including those that are part of the deliverables in B.6.1. In addition, the time axis contains a number of particular functionalities and options from III.7; these are purely indicative. The markers on the time axis indicate when certain functionalities will be available at the latest; these are independent of who ultimately offers them (the tenderer or meemoo).



**Figure 4: Growth path for the Knowledge Graph based on significant milestones (marked in bold) and options. The dotted line indicates a potential deviation from the growth path due to a sudden rise.**

The table below summarizes the growth path in estimated numbers on active data volume, throughput and the amount of active users. Tenderers should base their proposal with regard to the architecture, scaling and hosting of their solution on this, albeit limited to the years 2023 - 2025 as described in A.5.1. Information about the corresponding Service Level is described in III.9. For the sudden increase scenario, a description of the strategy and options for upscaling the solution is sufficient.

		April 2023	Juli 2023	April 2024	Januari 2025	Januari 2026
		Organizations GraphQL goes in production	IIIF 3.0 Presentation API goes in production	Archive 2.0 V3 retrieves public metadata from the Knowledge Graph using GraphQL	Inclusion of the metadata from Art in Flanders and Catalogus Pro in Archive 2.0 V3.1	All archive metadata is managed in the Knowledge Graph
PRD	Data Volume (nodes/edges or triples)	700.000 - 1 million	100 - 200 million	600 - 700 million	800 million - 1 billion	1.2 billion - 1.5 billion
	Throughput excl. network latency (requests/s)	2	4	8	8	10
	Active users	10	10	20	40	40
QAS	Data Volume (nodes/edges or triples)	105.000 - 150.000	15 - 30 million	90 - 105 million	120 - 150 million	180 - 225 million
	Throughput excl. network latency (requests/s)	2	2	2	2	2
	Active users	5	5	10	10	10
INT	Data Volume (nodes/edges or triples)	35.000 - 50.000	5 - 10 million	30 - 35 million	40 - 50 million	60 - 75 million
	Throughput excl. network latency (requests/s)	1	1	1	1	1
	Active users	2	2	2	2	2

## 7. FUNCTIONALITIES

The following matrices describe the desired functionalities of the software solution and the minimum requirements, divided over the basic solution and the different options. Each functionality can have the following status:

REQUIRED	The solution or option must contain this functionality and it must meet the minimum requirements described.
----------	---

OPTIONAL	The solution or option is not required to include this functionality. If the functionality is offered, it must meet the described minimum requirements.
----------	---

## BASIC SOLUTION (S): GRAPH-BASED METADATA MANAGEMENT SYSTEM

S1	Persistence of graph data	REQUIRED	In 2023, meemoo wants to start storing the metadata as a knowledge graph in accordance with the growth path from III.6. The solution must be able to store and index these volumes. It must also be possible to add data in large batches (~1 million nodes/edges or RDF triples) and in more granular quantities (~10-100 nodes/edges or RDF triples).
S2	Expressive, open, and widely supported query language	REQUIRED	<p>The graph must at least be queried in SPARQL 1.1, openCypher (V9), or another expressive query language of which the open character and broad adoption can be substantiated. The query language supports at least:</p> <ul style="list-style-type: none"> <li>- reading, adding, modifying, and deleting data in a graph;</li> <li>- deduplicating, sorting, paginating, and limiting of results;</li> <li>- filtering and aggregating graph data;</li> <li>- dealing with missing data and NULL values.</li> </ul> <p>It is also desirable that the query language can be extended with its own functions (whether or not implemented by the supplier).</p>

S3	Acceptable query response times	REQUIRED	<p>Meemoo is currently building the Archive 2.0 user platform that will consolidate the functionality and data of a number of other platforms. By the end of 2023, this platform should be fed by the metadata from our Knowledge Graph. In addition, a number of other existing or new applications will make use of the Knowledge Graph, such as the Archive for Education and/or internal tooling.</p> <p>Although these applications do not have to query the Knowledge Graph directly - they can use an application database or cache, if necessary, queries must be answered at a sufficiently high frequency (for example, queries without complex functions such as pathfinding under 100ms). Usually, meemoo works with 1 application to query the underlying database. Users do not query the solution directly, but the performance of the intermediate application must be guaranteed. Our most used platform currently handles 4000 visitors per day with peaks of up to 800 visitors per hour during working hours.</p> <p>This performance must also remain constant with the estimated annual growth in datavolume from III.6 and 5% in platform visitors. The solution must also offer the ability to quickly identify inefficiencies, for example by being able to request query plans. Currently it is about 80% simple queries (a context of 3 hops) and 20% more analytical queries. Queries with optional/NULL data are common. Meemoo doesn't expect this to change in the first three years. Concrete examples of queries can be found in appendix-2.</p>
----	---------------------------------	----------	---

S4	Data access for multiple users	REQUIRED	<p>By default, a content partner's metadata is not accessible to anyone outside of meemoo. That is why parts of the metadata must be protected if, for example, they are queried or changed via API. Thus, a permission management and access control mechanism such as DAC, MAC, RBAC, ABAC or PBAC must be present. There must be access for multiple users.</p> <p>The level of granularity depends on how the metadata is subdivided in the solution, but this is likely to be at least at the level of a graph or a particular set of nodes/edges or RDF triples. Ideally, this can be done down to the level of a number of nodes/resources. Different parts of the graph can be protected from certain groups or users so that they cannot see or query each other's metadata. A user who has access to several graphs can also query them together. Also, on different subsets, certain users can be restricted to only having read or write permissions.</p> <p>The solution must be protected against data corruption, for example, by a concurrency control system for manipulation and reading of data by different connections or users.</p>
S5	System administration API	REQUIRED	<p>For the day-to-day management of the infrastructure or when problems arise, meemoo must be able to gain insight into the performance of the system and do basic administration.</p> <p>There should be an API that enables at least the following administrative functions:</p> <ul style="list-style-type: none"> <li>- starting, stopping, or restarting the solution;</li> <li>- change the configuration or settings;</li> <li>- creating, modifying, emptying, or deleting databases and/or graphs;</li> <li>- managing users/groups/roles and access control;</li> </ul> <p>Other desired features:</p> <ul style="list-style-type: none"> <li>- management and a status overview of the cluster and its nodes;</li> <li>- retrieving and managing running queries;</li> <li>- gain insight into ongoing tasks and locks;</li> <li>- indexing functionality, for example, full or partial reindex.</li> <li>- management of computer resources such as thread pool and memory usage.</li> </ul> <p>The API is at least available via CLI or HTTP.</p>

S6	Queries about several graphs	REQUIRED	<p>Depending on how the system can and will divide data, the metadata of different content partners is divided into individual graphs, databases, or repositories. In addition, there are a number of graphs, databases, or repositories with general metadata, for example, information about organisations, schemas, and taxonomies.</p> <p>As a minimum, it should be possible to query these different metadata individually or together. It should also be possible to query all metadata together. For the time being, it is not a requirement to also be able to query external databases, for example, Knowledge Graphs or APIs that are publicly accessible on the web.</p>
S7	Annotating metadata	REQUIRED	<p>In addition to the metadata itself, it is important to maintain fine-grained information about:</p> <ul style="list-style-type: none"> <li>- who or what (for example, AI) added or changed certain metadata;</li> <li>- when certain metadata was added or last modified;</li> <li>- what the reliability score of this data is.</li> </ul> <p>It should also be possible to select or filter metadata based on the above information. Therefore, the internal graph model should allow to annotate facts in the metadata with its own metadata, such as by supporting RDFstar/SPARQLstar, edge properties, hyperedges...</p>
S8	Importing and exporting graph data	REQUIRED	<p>Meemoo must be able to connect the Knowledge Graph to the inflow processes or integrate new data sources via ETL. Sometimes static data must also be added to a file (for example, a taxonomy). Also, content partners must always be able to export their metadata as a data dump.</p> <p>In addition, the data as a whole must be transferable to other systems as easily as possible without loss of information.</p> <p>Therefore, datasets, graphs and the results of queries must be able to be serialised to a standardised data format for graphs (for example, Turtle, JSON-LD or GraphSON). It should also be possible to import these formats.</p>

S9	Support for lists and sets	REQUIRED	<p>Some (internal) applications rely on the order of terms in controlled lists (such as the education taxonomies) and have high performance requirements. Therefore, operations on lists in the query language or APIs (such as retrieving in sequence, searching, or sorting) must be performed as efficiently as possible.</p> <p>Both the query language and the data model should ideally provide direct support for sorted data structures. If the query language does not support this by default (SPARQL, for example), then proprietary extensions are a possible solution.</p> <p>The primary use case is dealing with ordered collections in SKOS. The necessary functionalities are therefore limited to:</p> <ul style="list-style-type: none"> <li>- storing nodes, resources or values as sequences and open ended lists</li> <li>- quickly retrieve all nodes or resources in order and use them in a query</li> </ul>
S10	HTTP API for queries	REQUIRED	<p>In order to integrate the (web) applications with the Knowledge Graph, it must be possible to perform queries over HTTP. It should also be possible to offer an API per graph or database.</p>
S11	Support for multilingual metadata	REQUIRED	<p>The data model must allow for different values in different languages for the same metadata field. It should be possible to select the metadata in a certain language.</p>
S12	Good availability and redundancy	REQUIRED	<p>In order to guarantee a good service, meemoo must keep the downtime of platforms or processes as low as possible.</p> <p>So, the solution must at least:</p> <ul style="list-style-type: none"> <li>- be able to perform an automatic fail-over;</li> <li>- offer automated continuous backups with point-in-time recovery</li> <li>- allow a degree of clustering or redundancy.</li> </ul> <p>Are also desirable:</p> <ul style="list-style-type: none"> <li>- equipped with round robin or other load balancing.</li> <li>- the possibility of rolling deployments and upgrades.</li> </ul>

S13	Good support for developers and DevOps	REQUIRED	<p>The software solution receives regular updates and has a transparent release cycle. The supplier provides excellent communication and guidance with every transition to a new version. Bugs and critical issues such as security risks are resolved promptly.</p> <p>There is also very comprehensive documentation for developers, so that they can build integrations smoothly. The service provider also has a structured approach for asking questions or reporting/handling issues (for example, support tickets, available by phone, etc.).</p>
S14	Available in different environments	REQUIRED	<p>As a minimum, the software solution is available in the following environments:</p> <ul style="list-style-type: none"> <li>- PRD: for applications in production;</li> <li>- QAS: for testing integrations and new functionalities;</li> <li>- INT: instance with predictable behavior for automatic testing</li> </ul> <p>See the growth path in III.6 for requirements related to data volume, throughput and number of users.</p>
S15	Monitoring and logging	REQUIRED	<p>The solution can forward operational and debug information to the ELK stack of meemoo. An integration with logstash is preferred.</p>
S17	Python and/or JavaScript software libraries	OPTIONAL	<p>There are mature software libraries that abstract the administrative and query APIs. They can</p> <ul style="list-style-type: none"> <li>- programmatically connect with the software solution;</li> <li>- perform administrative functions (as described in requirement S5) or queries.</li> </ul> <p>The libraries are available at least in Python 3 or Javascript/Typescript. Compatibility with other well-known software libraries (for example, RDFLib, RDF.js, Tinkerpop, Pandas, Apache Spark) is desirable.</p>



S18	Transaction history tracking	OPTIONAL	<p>The solution keeps a transaction log with all changes that happen on the data. This information can be used when querying the graph. The main use case is the fine-grained tracking and retrieval of the metadata origin.</p> <p>A possible future use case is to query, provide, or restore a previous version of a graph.</p>
S19	Reasoning about data	OPTIONAL	<p>There is support for data enrichment by automatically deriving metadata based on the existing metadata. This can be done by the presence of a (rule) reasoning engine (for example, Datalog, N3, RDFS, OWL, SHACL AF) or by sufficient expressiveness in the query language (for example, SPARQL CONSTRUCT).</p> <p>This mainly is about:</p> <ul style="list-style-type: none"> <li>- deriving subclasses;</li> <li>- implementing simple and independent rules with a view of small internally known patterns or corrections;</li> <li>- merging equivalent nodes;</li> <li>- applying extensive or complex ontologies is not applicable;</li> <li>- we start under the closed-world assumption; the open-world assumption is less important for the time being.</li> </ul>
S20	Validation of data integrity	OPTIONAL	<p>Data integrity can be checked by validating the data against a schema definition, such as in SHACL or SHEX.</p> <p>This can be done on a large scale, either after the import or during the import.</p> <p>It is also possible to reject data that does not validate.</p>
<b>OPTION A: GRAPHQL ACCESS</b>			
A1	Configuration of GraphQL APIs	REQUIRED	<p>The access platforms are built by external suppliers and use GraphQL to integrate with the meemoo infrastructure. Therefore, the Knowledge Graph must also be accessible via GraphQL. The software solution allows, for example, to configure one or more GraphQL APIs that give access to specific subsets of the Knowledge Graph (for example, metadata from a certain domain) from a certain user or with a certain licence.</p>

A2	Configuration of GraphQL schemas	REQUIRED	<p>It is essential that the GraphQL schema, and thus the API, can be defined to suit the consuming application and, therefore, independently of the graph model in the Knowledge Graph.</p> <p>This means that:</p> <ul style="list-style-type: none"> <li>- it should be possible to draw up a GraphQL schema manually with complete freedom in types, properties and their naming;</li> <li>- the schema should not be dependent or limited by the Knowledge Graph data model (for example, mandatory alignment with the existing class structure or properties);</li> <li>- the automatic generation of a GraphQL schema is, therefore, of less importance (although it is useful in scenarios where the automatic schema is sufficient).</li> </ul>
A3	Compatible with the GraphQL ecosystem	REQUIRED	<p>The API shall conform to the GraphQL October 2021 specification (<a href="http://spec.graphql.org/October2021/">http://spec.graphql.org/October2021/</a>), including schema introspection, and, therefore, be compatible with tools from the GraphQL ecosystem such as Hasura and Apollo. The semantics of the GraphQL queries should not depend on the semantics of the graph query language if this would lead to incompatible behaviour, for example, a difference in the handling of missing and NULL values compared to other GraphQL APIs.</p>
A4	Integration with access management Knowledge Graph	OPTIONAL	<p>The permissions or restrictions that apply to the Knowledge Graph can also be applied to the GraphQL endpoint, for example, user A has permission B and, therefore, cannot request/see field C in response.</p>
A5	Interactive test environment	OPTIONAL	<p>The API also provides an interactive environment in which GraphQL queries can be tested against the API, for example, GraphiQL (<a href="https://graphql-dotnet.github.io/docs/getting-started/graphiql/">https://graphql-dotnet.github.io/docs/getting-started/graphiql/</a>).</p>
A6	Support for transforms and custom resolvers	OPTIONAL	<p>The GraphQL implementation offers transforms and custom resolvers, for example, for filtering and visualising the <a href="#">EDTE</a> date notation.</p>
<h2>OPTION B: INTEGRATION WITH A SEARCH INDEX</h2>			

B1	Providing a search index	REQUIRED	<p>The access platforms currently use an Elasticsearch search index to search and filter metadata. The same functionality should be present when managing the metadata in the Knowledge Graph.</p> <p>The solution can offer this in two ways:</p> <p>1) as an integrated part of the solution: there is a search and filter API over the Knowledge Graph with similar functions and performance to the current Elasticsearch setup;</p> <p>2) as a link/integration with the existing Elasticsearch.</p> <p>In both cases, the (re)indexing of graph data to the index must be automatic, for example, a change (CRUD) in the Knowledge Graph must flow to (the document in) the index in real time. If the meemoo Elasticsearch solution is used, the provided solution must be able to provide metadata in Elasticsearch-readable JSON documents and support at least Elasticsearch version 7 or higher.</p>
B2	Schema and lexicalisation	OPTIONAL	<p>Meemoo mainly has metadata in Dutch, offers the access platforms in Dutch (in the future also English/French) and, therefore, mainly has Dutch-speaking users. The Elasticsearch, which supports the search functions in the platforms, is configured specifically for this situation. Offering a schema and mapping of the metadata in the Elasticsearch is allowed, but not required.</p>
B3	Support for EDTF	OPTIONAL	<p>There is a function for interpreting and indexing dates in the <a href="#">EDTF</a> notation.</p>
<b><u>OPTION C: ETL FRAMEWORK</u></b>			

C1	ETL process to the Knowledge Graph	REQUIRED	<p>An ETL (or ELT) should facilitate the frequent transfer of data from different sources to the graph.</p> <p>It must be able to handle at least a large volume of data records:</p> <ul style="list-style-type: none"> <li>- read JSON, XML and tables stored in files or applications;</li> <li>- perform basic transformations on values such as dividing, restructuring or correcting;</li> <li>- map to a graph model based on mapping definitions (applying schemas or ontologies);</li> <li>- upload in the Knowledge Graph.</li> </ul> <p>The ETL can be offered as:</p> <ul style="list-style-type: none"> <li>- an integrated application;</li> <li>- a workflow of actions or tasks (for example, as components of an existing ETL framework);</li> <li>- framework with which an ETL can be built.</li> </ul>
C2	Scalable and easy management	REQUIRED	<p>The largest volume of data to be processed consists of approximately 7 million JSON records and is growing every day. The ETL will usually have to process this incrementally, but sometimes a full synchronisation is done.</p> <p>Therefore, executing the ETL should be both scalable (i.e. limited memory usage) and user-friendly. For example, pipelines can be developed and executed iteratively; pipelines are (partly) reusable; you can test locally and it's easy to debug. A graphical interface to analyse and manage the ETL is desirable.</p> <p>Meemoo is currently investigating a possible adoption of the Prefect 2 (<a href="https://www.prefect.io/orion/v2/">https://www.prefect.io/orion/v2/</a>) framework. The current ETL processes are custom Python and Ruby implementations and will be migrated in due course. Recurring processes are scheduled via crontabs in the OpenShift container platform.</p>
C3	Intuitive reusable mapping definitions	REQUIRED	<p>Defining transformations or mappings is easy and fast, for example, because of an intuitive and readable syntax and/or because there is a helpful interface. The mappings will mainly be prepared by data engineers and/or developers.</p>

C4	Validation of output	REQUIRED	Before loading, the ETL result can be validated against a schema definition (for example, in SHACL or SHEMA) or with a number of tests that check the data quality.
----	----------------------	----------	---

## OPTION D: GRAPHIC ADMINISTRATION INTERFACE

D1	Graphical management application	REQUIRED	There is a graphical application that allows an administrator within meemoo to perform all the functions of the system administration API in a user-friendly manner.
D2	Query management	REQUIRED	The interface makes it easy to create queries (for example, auto-complete, syntax highlighting...), and test them. Queries can also be saved.
D3	Management of ontologies, vocabularies, taxonomies, and data catalogues.	OPTIONAL	The application offers the possibility to manage (SKOS) taxonomies that are used in the Knowledge Graph. In addition, the management of data models in the form of (RDFS) controlled vocabularies and (SHACL) restrictions is also desirable. Meemoo currently does this in VocBench ( <a href="http://vocbench.uniroma2.it">http://vocbench.uniroma2.it</a> ) and GitHub.
D4	Visual data exploration and navigation	OPTIONAL	The interface allows the user to visually explore the data in the Knowledge Graph, for example, with an interactive graph visualisation or a number of connected web pages.
D5	Support for media & heritage data	OPTIONAL	The interface provides additional functionality for modelling and describing data from media, archives, or other cultural heritage organisations.
D6	Support for audio-visual and photographic material	OPTIONAL	The interface can handle and visualise audio-visual and photographic material

## OPTION E: AUTOMATIC ANNOTATION OF DUTCH TEXT

E1	Automatic tagging of Dutch text	REQUIRED	This component can automatically recognise entities in Dutch text and assign them a type (for example, Person, Location, Artwork...).
----	---------------------------------	----------	---

E2	Linking entities	REQUIRED	Identified entities are linked to the content of the Knowledge Graph and/or to external sources such as Wikipedia, Wikidata or taxonomies such as AAT ( <a href="https://www.getty.edu/research/tools/vocabularies/aat/">https://www.getty.edu/research/tools/vocabularies/aat/</a> ).
E3	Automatic retrieval of textual descriptions	OPTIONAL	The Knowledge Graph will contain many textual descriptions such as title, description, OCR, or transcripts. This component can automatically retrieve them for analysis, for example, based on a query.
E4	Adaptability of the system	OPTIONAL	The internal models can be adapted or (partly) retrained based on the meemoo data/context. It is also possible to expand or correct the list of possible entities.

## OPTION F: IIIF 3.0 PRESENTATION API

F1	IIIF 3.0 presentation API	REQUIRED	IIIF 3.0 is an important standard for access in digital heritage and essential for the operation of meemoo (for example, artinlanders) and the dissemination of ERDF-GIVE metadata. The software solution makes it possible to provide an IIIF 3.0 (Presentation) API ( <a href="https://iiif.io/api/presentation/3.0/">https://iiif.io/api/presentation/3.0/</a> ) on the metadata in the graph.
----	---------------------------	----------	--

## OPTION G: ANALYSIS OF GRAPH DATA

G1	Integration with BI software	REQUIRED	A connection can be made from Tableau and/or Microsoft PowerBI with the Knowledge Graph for generating reports.
G2	Efficient counting and aggregation of graph data	REQUIRED	<p>The reports based on the Knowledge Graph are mainly relating to gaining insights about and based on the metadata, such as:</p> <ul style="list-style-type: none"> <li>- How much metadata is present or missing?</li> <li>- Which metadata fields are popular?</li> <li>- How many different media types are there?</li> <li>- Which authors often appear in the archive?</li> <li>- How much material is in the public domain?</li> </ul> <p>The software solution must be able to answer these queries, which often involve aggregating and counting. The time this takes is less important as long as it remains workable (for example, hours, but not days) and does not affect the execution of other queries.</p>

G3	Execution of graph algorithms	OPTIONAL	<p>The software solution supports a number of graph algorithms:</p> <ul style="list-style-type: none"> <li>- finding possible paths between two nodes, for example, with A*;</li> <li>- clustering of nodes according to certain properties;</li> <li>- ranking of nodes, for example, according to popularity (in &amp; out degree);</li> <li>- calculating statistics about links such as pagerank.</li> </ul>
<b><u>OPTION H: INTEGRATION WITH SAML IDENTITY PROVIDER</u></b>			
H1	Integration with SAML Identity Provider	REQUIRED	<p>The authentication of users in the solution can be done with the meemoo LDAP and/or SAML 2.0 IdP. Authentication via bearer token (jwt) is a plus.</p>
<b><u>OPTION I: RECONCILIATION API</u></b>			
I1	Implementation of the reconciliation API	REQUIRED	<p>Meemoo can offer content partners the option to link their data to the meemoo Knowledge Graph, for example, via OpenRefine. Therefore, the software solution can offer a Reconciliation Service API v0.1 (<a href="https://reconciliation-api.github.io/specs/0.1/">https://reconciliation-api.github.io/specs/0.1/</a>) that can link text to entities from the Knowledge Graph.</p>
<b><u>OPTION J: OAI-PMH API</u></b>			
J1	Implementation of an OAI-PMH API	REQUIRED	<p>Content partners can retrieve their own metadata from the Knowledge Graph via the OAI-PMH protocol</p>

## 8. MAINTENANCE

The software solution must be accompanied by an annual renewable maintenance contract to be entered into with effect from the day after the first provisional delivery (see B.6.2). If necessary, this maintenance contract is carried out by the service provider and includes corrective, adaptive and preventive maintenance with regard to the applications.

The tenderer will be the single point of contact for the entire delivered solution, including hosting. Support contracts for the hosting are set up by the candidate supplier itself in function of meemoo's requirements regarding maintenance from this section and SLA (see III.9).

Candidate suppliers are requested to provide an action plan for three levels of urgency:

- High priority issues: A response time of 4 hours within office hours (Mon-Fri, 9am - 5pm CE(S)T) is required. Work on the issue will begin within 4 hours of the issue being registered.
- Medium priority issues: Response time within the next business day.
- Low priority issues: Response time within 5 business days.

Candidate suppliers may propose changes to these requirements, for example, to align with certain internal processes.

Candidate suppliers should also provide prices for work outside office hours and weekend work.

## 9. SERVICE LEVEL

The solution should only be highly available for read operations in the PRD environment. For write operations, the PRD environment does not have to be highly available. There are no high availability requirements for the QAS and INT environment. In concrete terms, this translates into the following 'uptime':

	Read operations	Write operations
<b>PRD</b>	≥ 99.99%	≥ 98%
<b>QAS</b>	≥ 95%	≥ 95%
<b>INT</b>	≥ 95%	≥ 95%



The above uptimes are excluded from foreseen unavailability. The candidate provides sufficient measures to guarantee uptime, such as internal monitoring and alerting systems. Meemoo has the right to measure availability with an internal monitoring system.

In consultation with meemoo, times for maintenance can be provided. During this maintenance, the application may be temporarily unavailable. To this, the following preconditions apply:

- The maintenance must be scheduled well in advance (at least 1 month) in consultation with meemoo.
- Adaptations are performed sequentially on the different environments, with meemoo having sufficient time per environment to test the adaptation.
- Unavailability of the production environment is kept to a minimum.
- Maintenance of the production environments will always take place outside office hours (after 6 pm and preferably on a weekend).
- For urgent updates such as hot fixes or things that, from a certain point of view, need to be deployed quickly (e.g. fixes for issues with a very high impact for end users, issues that relate to the security of the solution, ...), an accelerated procedure can be used. Its use is done in consultation with meemoo and is only possible with the explicit consent of meemoo.

## 10. INTEGRATION OF THE SOFTWARE SOLUTION

This contract also includes the provision of services for integrating the software solution into meemoo's access infrastructure. Specifically, this involves the following tasks:

1. Positioning the solution in meemoo's access infrastructure, i.e. a concrete interpretation of the architectural picture in III.2. For example, tasks of the different components, necessary APIs, mapping of data flows, etc.
2. Making the solution ready for metadata import. In principle, the data is delivered in RDF by meemoo, but this can still change based on the options offered (for example, ETL), the nature of the solution (for example, property graph), or the approach suggested by the service provider.
3. Connecting the solution with the necessary meemoo components:
  - a. make the content of the Knowledge Graph queryable via GraphQL and integrate it into the Hasura setup of meemoo. If the GraphQL implementation is provided by meemoo, the supplier supports its connection to the solution.
  - b. provide a synchronisation between the taxonomies of meemoo and the Knowledge Graph. These are provisionally managed in a VocBench instance (<http://vocbench.uniroma2.it/>). This synchronization happens according to fixed release cycles, so not continuously.
4. Perform an end-to-end load test as agreed with meemoo, including configuring and optimising the software solution for this test.

Some platforms provide specific admin or CMS functionality to users, which is not considered metadata; that data remains stored in the application's PostgreSQL database and remains accessible via Hasura. Hence, there are two possible integration scenarios: a) the Hasura serves as a proxy for the KG GraphQL via remote schema stitching or b) both co-exist and are directly accessible. Note that this does not affect the requirements: the service must act as a high-performance, fully compatible GraphQL endpoint.

The design of the price matrix should be used as-is and may not be altered.  
An option's required functionality is indivisible and is therefore priced as one unit.

The design of the price matrix should be used as-is and may not be altered. An option's required functionality is indivisible and is therefore priced as one unit.				Year 1 (jan - dec 2023)	Year 2 (jan - dec 2024)	Year 3 (jan - dec 2025)
Included in the price comparison						
Basic Solution (S)	Required functionalities	PRD	License	€0,00	€0,00	€0,00
			Hosting	€0,00	€0,00	€0,00
			Support	€0,00	€0,00	€0,00
		QAS	License	€0,00	€0,00	€0,00
			Hosting	€0,00	€0,00	€0,00
			Support	€0,00	€0,00	€0,00
		INT	License	€0,00	€0,00	€0,00
			Hosting	€0,00	€0,00	€0,00
			Support	€0,00	€0,00	€0,00
		Installation, start-up, configuration.			€0,00	
Training			€0,00			
Services	day price	Senior Metadata Modeler			€0,00	
		Senior Analyst			€0,00	
		Junior Developer			€0,00	
		Senior Developer			€0,00	
		Project Manager			€0,00	
NOT included in the price comparison						
Required options: the tenderer is obliged to submit a bid for each required option. '0' = included in the basic solution						
Option A	Required functionalities	A1 (Configuration of GraphQL APIs), A2 (Configuration of GraphQL schemas), A3 (Compatible with the GraphQL ecosystem)		€0,00	€0,00	€0,00
	Optional functionalities	A4 (Integration with access management Knowledge Graph)		€0,00	€0,00	€0,00
		A5 (Interactive test environment )		€0,00	€0,00	€0,00
		A6 (Support for transforms and custom resolvers)		€0,00	€0,00	€0,00
Option B	Required functionalities	B1 (Providing a search index)		€0,00	€0,00	€0,00
	Optional functionalities	B2 (Schema and lexicalisation)		€0,00	€0,00	€0,00
		B3 (Support for EDTF)		€0,00	€0,00	€0,00
Allowed options: the submission of permitted options is optional. 'N/A' = not submitted; '0' = included in the basic solution						
Basic solution (S)	Optional functionalities	S17 (Python and/or JavaScript software libraries)		€0,00	€0,00	€0,00
		S18 (Transaction history tracking)		€0,00	€0,00	€0,00
		S19 (Reasoning about data)		€0,00	€0,00	€0,00
				€0,00	€0,00	€0,00
		S20 (Validation of data integrity)				
Option C	Required functionalities	C1 (ETL process to the Knowledge Graph), C2 (Scalable and easy management), C3 (Intuitive reusable mapping definitions), C4 (Validation of output)		€0,00	€0,00	€0,00
Option D	Required functionalities	D1 (Graphical management application), D2 (Query management)		€0,00	€0,00	€0,00
	Optional functionalities	D3 (Management of ontologies, vocabularies, taxonomies, and c		€0,00	€0,00	€0,00
		D4 (Visual data exploration and navigation)		€0,00	€0,00	€0,00
		D5 (Support for media & heritage data)		€0,00	€0,00	€0,00
		D6 (Support for audio-visual and photographic material)		€0,00	€0,00	€0,00
Option E	Required functionalities	E1 (Automatic tagging of Dutch text), E2 (Linking entities)		€0,00	€0,00	€0,00
	Optional functionalities	E3 (Automatic retrieval of textual descriptions)		€0,00	€0,00	€0,00
		E4 (Adaptability of the system)		€0,00	€0,00	€0,00
Option F	Required functionalities	F1 (IIIF 3.0 presentation API)		€0,00	€0,00	€0,00
Option G	Required functionalities	G1 (Integration with BI software) G2 (Efficient counting and aggregation of graph data)		€0,00	€0,00	€0,00
	Optional functionalities	G3 (Execution of graph algorithms)		€0,00	€0,00	€0,00
Option H	Required functionalities	H1 (Integration with SAML Identity Provider)		€0,00	€0,00	€0,00
Option I	Required functionalities	I1 (Implementation of the reconciliation API)		€0,00	€0,00	€0,00
Option J	Required functionalities	J1 (Implementation of an OAI-PMH API)		€0,00	€0,00	€0,00