

Output 5.1.1
Data Management Plan
version 2.0
30 june 2021

Project Name: CoGhent; Collections of Ghent
Funding programme: Urban Innovative Actions Call 2019
Contract Number: UIA05-249
Work Package: 5

Contributors: Olivier Van D'huynslager, Sofie Teugels

Acronyms	5
Executive Summary	6
Concepts, tools and standards	7
1. Data description	9
1.1 Data types and formats	9
1.1.1 Cultural Heritage Data	9
1.1.2 Narrative frames	9
1.2 Data sources (provenance)	10
1.2.1 Descriptive metadata from CMS	10
1.2.2 Data from file servers cultural heritage institutions	10
1.2.3 Data input by end users	10
1.2.4 LDES Client	10
1.3 Data streams (and transformations)	11
1.3.1 From sources to DAMS	11
1.3.1.1 From CMS to DAMS	11
1.3.1.2 Data from file servers cultural heritage institutions	12
1.3.1.3 Data input by end users	12
1.3.2 From DAMS to end point	12
1.3.3 From DAMS to CMS	12
1.4 Size of the data (storage and back-up)	13
1.4.1 Size of cultural heritage data	13
1.4.2 Size of the narrative frames	13
1.5 Data preservation	13
2. Promote reuse of cultural heritage data	15
2.1 PSI open data and Open Access Policy	15
2.2 FAIR	15
2.2.1 Making data findable, including provisions for metadata	15
2.2.2 Making data openly accessible	19
2.2.3 Making data interoperable	20
2.2.3.1 Descriptive metadata from CMS	20
2.2.3.2 Data from file servers cultural heritage institutions	21
2.2.3.3 Data input by end users	21
2.2.4 Making data reusable	21
2.2.4.1 Descriptive metadata from CMS	21
2.2.4.2 Data from file servers cultural heritage institutions	21
2.2.4.2.1 Workflow rights clearance	22
2.2.4.3 Data input by end users	24
3. Privacy, ethics, trust and security	25

3.1 Privacy	25
3.1.1 First hand personal data	26
3.1.2 Second hand personal data	26
3.2 Ethics	26
3.3 Trust	26
3.4 Security	26
4. Roles and responsibilities	27
4.1 Legal	27
4.1.1 Descriptive metadata from CMS	27
4.1.2 Data from file servers cultural heritage institutions	28
4.1.3 Data input by end users	28
4.1.4 LDES client	28
4.1.5 Personal data	28
4.2 Operational	28
4.2.1 Descriptive metadata from CMS	28
4.2.2 Data from file servers cultural heritage institutions	29
4.2.3 Data input by end users	29
4.2.4 LDES client	29
4.2.5 Personal data	29

Acronyms

AF: Application Form
API: Application Programming Interface
AV: Audiovisual
CDL: Cultural Data Lab
CH: Cultural Heritage
CHI: Cultural Heritage Institution
CHO: Cultural Heritage Object
CMS: Collection Management System
CPT: Cultural Participation Toolkit
DAMS: Digital Asset Management System
DMP: Data Management Plan
DGP: Data Governance Plan
DPA: Data Processing Agreement
ERC: European Research Council
FAIR: Findable, Accesibile, Interoperable, Readable
IIIF: International Image Interoperability Framework
LDES: Linked Data Event Stream
LDF: Linked Data Fragments
MUA: Main Urban Authority
OAP: Open Access Policy
UIA: Urban Innovation Actions
WP: Work Package
GLAM: Galleries, Libraries, Archives and Museums
RiC: Records in Contexts

Executive Summary

The Executive Summary will be written in the third and final version of the Data Management Plan.

The Data Management Plan describes the lifecycle of the data used and generated within the CoGhent project. It sets the framework for the management of the data, from acquisition to curation to dissemination to preservation, post-project data management included.

The Data Management Plan is prepared as part of 'Work Package 5: Data', and is labelled as 'Output 5.1.1'. It is considered a living document as the Data Management Plan will be regularly updated based on new insights. Updated versions will be submitted to the business steering committee for approval.

CoGhent is funded by Urban Innovative Actions (UiA), an initiative of the European Union that promotes pilot projects for sustainable urban development.

Concepts, tools and standards

The Data Management Plan mentions several concepts, tools and standards that are used and applied throughout the CoGhent project. To ensure a clear and uniform understanding, the DMP starts with an explanation of these concepts, tools and standards.

The CoGhent project is an urban innovation project that aims to investigate how cultural heritage data can be used and reused in a civic context to promote social inclusion and social cohesion. The main objective of CoGhent is to enhance cultural participatory practice.

The foundation of the CoGhent project is the data of the cultural heritage collections of the five participating cultural heritage institutions: STAM, Design Museum Gent, Industriemuseum, Huis van Alijn and Archief Gent.

CoGhent, when using the term **cultural heritage**, refers to both tangible and intangible expressions that have shared meanings and values within communities and within a current frame of reference, and that are therefore passed onto future generations.

CoGhent defines **cultural heritage data** as all data relating to the identification and interpretation of tangible and intangible cultural heritage objects as well as their digital reproductions. Data consists of two elements: descriptive metadata and digital reproductions (images, audiovisual fragments). Cultural heritage data are the assets for the narrative frames CoGhent wants to collect and present.

For the duration of the CoGhent project, the cultural heritage data of the collections will be optimized and opened up, based on this Data Management Plan and existing (inter)national standards for collection and information management (**SPECTRUM**, **CIDOC-CRM**, **OSLO**), using the **FAIR principles**¹. The cultural heritage institutions are supported in this by the **Data Intervention Team** (5 full time project employees). The preparations and workflow of the Data Intervention Team are documented in a [CoGhent manual](#) (via Gitbook), available for the participating cultural heritage institutions and interested third parties, e.g. cultural heritage institutions that want to join in the future.

The optimized data is collected via a **Linked Data Event Stream (LDES)** and published as **Linked Open Usable Data (LOUD)** in a newly designed **Digital Asset Management System (DAMS)**. Not only do the participating cultural heritage institutions have access to DAMS, the public as well can add and adapt data via a **web platform**, connected to DAMS. CoGhent actively reaches out to the public to harvest unknown footage, personal narratives, annotations and hidden histories.

By publishing cultural heritage data as linked open usable data the project wants to open up data for use in participatory processes such as co-creation and co-curation activities which the project incentivizes and facilitates. CoGhent wants to facilitate the creation of new and engaging digital applications for a broad audience; applications that help enhance cultural participatory practice and increase social cohesion, the project's main objective.

¹ <https://www.go-fair.org/fair-principles/>

The cultural heritage data (**assets**) are enriched with narratives (**narrative frames**) that give context. The narratives are harvested by presenting the assets to a broad public via the CoGhent web platform and the CoGhent box (in situ), and by crowdsourcing initiatives. The narratives are captured via the web platform and are created by the public, by the end users.

CoGhent aims to increase social cohesion by facilitating online and offline conversations about cultural heritage. It is our goal that these conversations will subsequently result in collecting even more cultural heritage (data) and narratives. By 2023, **CoGhent** wants to digitally present and publish the data of **100.000 cultural heritage objects**, including a minimum of **300 narrative frames**.

CoGhent is working closely with meemoo to develop a **PID strategy** with attention to the integration of the Flemish URI standard. This will eventually be included as a use case in the best practice guide on the use of persistent URIs in the cultural heritage sector.

1. Data description

1.1 Data types and formats

CoGhent actively collects, processes, reuses and generates different types of data: **cultural heritage data (descriptive metadata and digital reproductions) and narrative frames (created by end users)**. The cultural heritage data is **derived or compiled data** from different sources. The narrative frames are **experimental data**, collected under controlled conditions, in situ or via the web platform of CoGhent.

Both data types are linked with **reference data** (external authorities; published and/or curated peer-reviewed data sets), as the FAIR principles are implemented.

For the format of both data types open and documented standards are used, e.g. **OSLO cultuur object** and **OSLO cultuur event** (standards developed by the Flemish government) and **IIIF**. The initial format (input) of the data before the transformation to an open and documented standard (output) is explained below (*1.1.1 Cultural heritage data and 1.1.2 Narrative frames*).

In the process, **personal data** as well will be collected. Personal data and privacy policy are discussed in chapter 3 'Privacy and ethical aspects'.

CoGhent wants to maximise the possibilities of reusing the data. To facilitate the intake, it does not impose any restrictions on file formats but does provide the necessary strategy for converting to more current formats if there is a high risk of a format becoming obsolete in the near future.

1.1.1 Cultural Heritage Data

Cultural heritage data is collected via the participating cultural heritage institutions and the end users. The data consists of both **descriptive metadata** and **digital reproductions**. A digital reproduction provided with descriptive metadata is considered an **asset**.

The descriptive metadata is preserved in a collection management system (software and online). An export of the metadata results in a csv-format.

The digital reproductions are JPEG (.jpeg, .jpg), TIFF (.tiff, .tif), Adobe Portable Document Format (.pdf), *aanvullen met audio en video (file-based or directory-based data format)*.

1.1.2 Narrative frames

Apart from descriptive metadata and digital reproductions, the end user can also create **narrative frames** via the CoGhent web platform. Narrative frames link assets by adding relevant context. For CoGhent, the narrative frames are an important instrument to reach the project's objective of increasing social cohesion. The narrative frames (or stories) will expose unknown connections between assets and end users.

The narrative frames are created via the web platform in the format of an open and documented standard.

1.2 Data sources (provenance)

CoGhent installs multiple data streams, collecting and presenting data from various sources, going from collection management systems (descriptive metadata) and file servers (digital reproductions) from the participating cultural heritage institutions, to direct input in DAMS via the web platform by the end users (descriptive metadata, digital reproductions and narrative frames). The descriptive metadata from CMS and the digital reproductions from the file servers belong together and form a whole.

Data enters the project either in a structured way (when derived from a dedicated CMS) or not (input end users). Both data entries need a different approach (see further).

1.2.1 Descriptive metadata from CMS

All participating cultural heritage institutions use Adlib (software based) or Axiell Collections (online based), a full-fledged collection management system, for the management of the **descriptive metadata**.

The descriptive metadata compiled from a **CMS** is already optimized by the data intervention team, meaning that the FAIR principles and CIDOC-CRM standard are implemented, before it is mapped according to open OSLO standards.

Depending on the outcome of the crowdsourcing activities, the use of other collection management systems by end users can surface.

1.2.2 Data from file servers cultural heritage institutions

The digital reproductions of the collections (incl. audiovisual content) are stored on the **file servers** of the participating cultural heritage institutions. For the import of the reproductions in DAMS, a migration is set up by Inuits. The intention is to migrate all reproductions to DAMS, where it will be managed in the future.

1.2.3 Data input by end users

The data input by **end users** includes both descriptive metadata, digital reproductions and narrative frames. The data from end users is collected via the web platform and in situ, e.g. via crowdsourcing initiatives. Input via the web platform is possible when the end users create a profile and give informed consent (see chapter 3 'Privacy, ethics, trust and security'). The data entry is directly and manually in **DAMS** and lands in a validation layer. Before the data is made public, the Data Intervention Team implements an optimization of the data, based on the Data Management Plan.

1.2.4 LDES Client

Ultimately, CoGhent will publish data as linked open data via **Linked Data Event Streams** (LDES), creating a fourth source. End users can harvest data and create their own indexes on top of it.

Link IIF images and metadata coming from LDES.

1.3 Data streams (and transformations)

The two main data streams are 1) the migration and input of data from various sources into the Digital Asset Management System (DAMS) and 2) the Linked Data Event Stream departing from DAMS to be published on a public endpoint. An additional data stream from DAMS back to CMS, upon request of the participating cultural heritage institutions, is taken under consideration.

1.3.1 From sources to DAMS

1.3.1.1 From CMS to DAMS

Via the node service eventstream API the data is collected from Adlib/Axiell Collections and imported in DAMS. There are five eventstreams, one for every participating cultural heritage institution.

Data received from the participating cultural heritage institutions² is data already used in professional collection management processes, where authenticity and integrity of the data is important. The integrity of the data should be monitored and ensured at all times. That's why the content transformation and optimization takes place as close as possible to the source.

Transformations at source level:

- All data is standardized uniformly with the standards adopted by the project in correspondence with the partnering institutions (see chapter 2)
- Intake of this metadata into the infrastructure of the project is only allowed when it meets a certain set of expectations, in adopting standards and linked open data principles. These rules are the result of a collaborative process with the partnering institutions to ensure continuity and mitigation of major changes in current registration processes.

Transformation to Linked Open Data format (adlib-backend)

- Data is recurrently fetched via a sync job every 24hours at approximately 1AM.
- After a first pull of the data, only data that is altered at the source (CMS) will be fetched.
- The adlib2evenstream service transforms and structures data into linked data.

See 2.2 for further documentation on the Linked Data Event Stream specs.

² Within the duration of the project, we mean the five cultural heritage institutions registered as content partners: *STAM*, *Design Museum Gent*, *Industrie Museum*, *Huis van Alijn* and *Archief Gent*.

1.3.1.2 Data from file servers cultural heritage institutions

The digital reproductions (both images and other audiovisual material) are currently stored on the different file servers of the participating cultural heritage institutions. The digital reproductions are managed differently by each institution: the composition of file names, the available information in EXIF, the used licences,... First of all, it's important that the file name holds the CMS record number in order to match the digital reproduction with the correct and corresponding descriptive metadata. Second, if EXIF information needs to be extracted, the information needs to be correct. And thirdly, the digital reproductions must be cleared from copyright.

CoGhent wants to ensure that the import of this content into DAMS happens in a standardized manner. To facilitate this, the necessary standardized import mechanisms (both for bulk and individual intake) are foreseen. Bulk import is done via a csv file, prepared by the Data Intervention Team. In this process, the size of the digital reproductions is also checked. The csv file is stored on the file servers of the institutions where Inuits will track the listed digital reproductions and import them in DAMS. There is a general approach for all the participating cultural heritage institutions, meaning that every csv file holds the same columns, requiring the same information. There will be a recurrent check for newly added data by Inuits but the cultural heritage institutions remain in control of what is fetched and when.

Based on their file name, the reproductions are linked with the corresponding descriptive metadata in DAMS.

1.3.1.3 Data input by end users

The end users are stimulated to enrich the available cultural heritage data in DAMS with metadata, digital reproductions (incl. audiovisual), annotations and narrative frames. The end users put the data directly in DAMS, where the data is collected in a validation layer. CoGhent chose only to filter or reject data based on ethical, political, and privacy sensitivities, to create space for relevance within a broader framework, relevance that's not detectable short-term. The screening of the input is done by the Data Intervention Team.

1.3.2 From DAMS to end point

rol Inuits, District09

Via a Linked Data Event Stream, the data is available as open data in the District 09 triple store, Virtuoso.

1.3.3 From DAMS to CMS

This particular data stream is still in the research phase and will be highly dependent on the result of the first months of data input by the end users. The data stream from DAMS to the CMS refers to adapted and added metadata by the end users in DAMS that is approved by

the Data Intervention Team and that the cultural heritage institutions want to register in their CMS.

1.4 Size of the data (storage and back-up)

1.4.1 Size of cultural heritage data

To accommodate the incoming data streams, starting in the second quarter of 2021, the project will provide an initial disk of storage capacity. As such, 10TB is provided for data from the participating cultural heritage institutions. For data from the end users, 2TB is provided. Storage-side and server-side will be hosted and maintained by District09. As the project moves forward, the amount of data will increase. Towards 2023, a minimum of 100.000 images of cultural heritage objects will need storage, taking approximately 25TB.

The storage capacity needed to accommodate the intake of cultural heritage data, originating from professional providers, is estimable as the size of collections and their respective digital reproductions are known. The storage capacity needed for data from the end users is unknown as it is dependent on various external factors, such as the results of participatory trajectories initiated by partners. CoGhent will monitor the size of the data input of end users and adjust the available storage as such. Once the web platform is open for input by end users, an efficient monitoring can be set up, based on (to be determined) measurement indicators. This ensures a realistic growth forecast and a sustainable storage policy.

CoGhent wants to eliminate technical limits and rather focus on scalable solutions to ensure all input is stored properly. The 25TB storage is assured until the end of the project in 2023.

As to storage after 2023, CoGhent wants to establish an agreement with District09/professional providers/City of Ghent/... (depends on value contribution) to secure a sustainable solution.

1.4.2 Size of the narrative frames

In research phase.

1.5 Data preservation

CoGhent wants to preserve the data that is optimized and transformed into linked open data, in other words, the Linked Data Event Stream. For the duration of the project all versions of an object in the Linked Data Event Stream will be stored by District09.

2. Promote reuse of cultural heritage data

This section explains how the FAIR principles, making data findable, accessible, interoperable and reusable, will be enforced during and implemented after the project.

2.1 PSI open data and Open Access Policy

Cultural heritage data will be transformed and optimized by the participating cultural heritage institutions, supported in this by the Data Intervention Team. It will do so in such a manner, to meet the needs and expectations of all activities and best serve all stakeholders within the project. The data will be exposed via a Hypermedia Web API to facilitate interactions between computer programs. For this purpose, the project recognizes the importance of **Public Sector Information Open Data**³. As a result of these optimizations, data published via the Hypermedia Web API (endpoint) will be enriched, linked and open.

“Citizens should be empowered to make better decisions based on insights gleaned from non personal data. And that data should be available to all - whether public or private, big or small, start-up or giant.”⁴

CoGhent acknowledges the notion that merely opening up data is not enough. In order to truly accomplish Linked Open Useable Data (LOUD) it has to financially incite, inspire and stimulate cross-domain discovery and reuse.

By adopting an Open Access Policy (OAP) and opening up information, the project wants to optimise the impact of publicly-funded research and innovation during and after the project.

Expected impacts of opening up scientific information:

- better research and cultural heritage practice by building on previous results
- more efficient research by avoiding duplication and promoting reuse
- economic growth (accelerated and open innovation)
- improved transparency (involving citizens & society)

2.2 FAIR

CoGhent adopts and implements the FAIR principles during the full lifecycle of the data. From first creation to publication as LOD.

2.2.1 Making data findable, including provisions for metadata

Findable - data should be identified using globally unique, persistent and resolvable identifiers, and should include machine-actionable contextual information that can be indexed to support human and machine discovery of the given data.

³ <https://overheid.vlaanderen.be/informatie-vlaanderen/producten-diensten/open-data>

⁴ Communication from the commission to the European Parliament, the council, the European economic and social committee and the committee of the regions, *A European Strategy for data*, 1 (https://ec.europa.eu/info/sites/info/files/communication-european-strategy-data-19feb2020_en.pdf)

CoGhent provides a DCAT catalogue for all the collections/eventstreams and uses the OSLO DCAT-AP standard⁵ which is aligned with the DCAT Application profiles for data portals in Flanders and Europe (DCAT-AP).⁶ This catalogue hints where third-parties can retrieve the latest versions of objects in a dataset. Consequently, third-parties are able to replicate the objects on their system and expose more advanced indexes or even query interfaces. This way, making data findable becomes a collaborative responsibility.

F1: CoGhent assigns the cultural heritage data, both provided and collected, a globally unique and persistent identifier.

In Coghent, Linked Open Data (LOD) is a first-class citizen with the use of the Linked Data Event Streams (LDES). Following the second LOD principle, HTTP URIs must be used so others can look up a certain resource. Also, a persistent identifier is necessary to reconcile versioned objects of the LDES to the same object (materialize an object with a certain version, retrieve an object's history).

Generation and storage of persistent URI's

It will be possible to generate and use persistent URI's for the following entities: DCAT catalogue, LDES datasets, records of objects (artefacts), agents, concepts and images.

For the LDES process (adlib-backend) it is important that persistent URI's are a part of the metadata. This process needs to fetch an existing persistent URI from the CMS (Adlib API) if the cultural heritage institutions already generate persistent URI's via their own solution or - if there is no persistent URI - generate a new one.

Once generated, the persistent identifiers will have to be recorded in the metadata of each entity. In CoGhent, the cultural heritage institutions will agree on a dedicated field in the CMS to store and maintain the URI of an object, agent, concept. When this field is non-empty, the adlib-backend service will reuse this value from this field when generating a Linked Data mapping of the object. Otherwise, the adlib-backend service will generate a URI using the URI template (described below).

Resolving

The persistent URI's must resolve to a html subject page providing specific information about this subject. Also here, the institutions need to agree on a dedicated field in the CMS to indicate the resource where to resolve. Coghent will make use of the resolver of Stad Gent where the persistent URI must have a "foaf:page" to the page the resolver needs to resolve. When the dedicated field in the CMS is empty, the adlib-backend service will create a foaf:page to the Coghent platform by default.

Syntax

⁵ <https://data.vlaanderen.be/doc/applicatieprofiel/DCAT-AP-VL/>

⁶

<https://joinup.ec.europa.eu/collection/semantic-interoperability-community-semic/solution/dcat-application-profile-data-portals-europe>

CoGhent uses following URI template based on the Flemish standard for URIs⁷:

<https://{{domain}}/{{type}}/{{concept}}/{{reference-basic}}+/{{reference-version}}?>

- CoGhent makes use of the **https://** protocol and will publish on the **stad.gent** domain.
- All data entities are given 'id' as type to clarify that the URIs are identifiers.
- Concept
- With the reference-basic, a namespace is given per collection
- For the reference-basic part of the URI, a hash is generated based on the database record in Adlib (preref) and the record's timestamp of creation.
- Reference-version is only used for the versioned entities in the event streams and is by default the timestamp of modification in Adlib, for example:
https://stad.gent/id/concept/dmg/570016285/2021-06-01T02:31:16.241Z.
- + means there can be one or more basic references? means the version reference is optional and can be maximum one

Label	Concept	Reference-basic 1	Reference basic 2	Description
catalogue	dcat	coghent	/	Catalogue of Coghent LDES datasets Only one URI is generated: https://stad.gent/id/dcat/coghent
dataset	dataset	uses the respective institution abbreviation: dmg, hva, industriemuseum, stam, archiefgent, adlib	hash(institution name, Adlib database name)	A dataset corresponds with one LDES Example: https://stad.gent/id/dataset/adlib/5d500058200ecaaf8b97bb1bed9fdb94
object (artifact)	mensgemaakt object	dmg	hash(preref, creation timestamp)	objects with provenance Design Museum Gent Example: https://stad.gent/id/mensgemaaktobject/dmg/530026423

⁷ <https://data.vlaanderen.be/standaarden/erkende-standaard/vlaamse-uri-standaard-voor-data.html>

object (artifact)	mensgemaakt object	hva	hash(priref, creation timestamp)	objects with provenance Huis van Alijn Example: https://stad.gent/id/mensgemaaktoobject/hva/470084104
object (artifact)	mensgemaakt object	industriemuseum	hash(priref, creation timestamp)	objects with provenance industriemuseum Example: https://stad.gent/id/mensgemaaktoobject/industriemuseum/570051115
object (artifact)	mensgemaakt object	stam	hash(priref, creation timestamp)	objects with provenance STAM Example: https://stad.gent/id/mensgemaaktoobject/stam/550027094
object (artifact)	mensgemaakt object	archiefgent	hash(priref, creation timestamp)	objects with provenance Archief Gent Example: https://stad.gent/id/mensgemaaktoobject/archiefgent/550027094
agent	agent	adlib	hash(priref, creation timestamp)	agent catalogue with provenance from all partners Example: https://stad.gent/id/agent/470000128
concept	concept	adlib	hash(priref, creation timestamp)	concept catalogue (thesaurus) with provenance from all partners Example: https://stad.gent/id/concept/570016285

F2: To increase the findability of data, the Data Intervention Team supports the institutions in optimizing the cultural heritage data by enriching the metadata.

- Enriching the metadata by saving the generated persistent URIs for entities in CMS. This way, the adlib-backend service knows which URI to use for the object.
- Enriching the metadata by linking with persistent URIs from other authorities (Getty, RKD, Wikidata...)

F3: URI's of cultural heritage datasets are mentioned in the exposed dataset catalogue.

To encourage the reuse and spreading of the persistent URIs for internal and external use cases:

- persistent URIs are a part of the metadata record of heritage objects
- persistent URIs are published in the online catalogue of the project

2.2.2 Making data openly accessible

All data exposed via the LDES will be offered for free and open access for both commercial and non-commercial purposes⁸ via a hypermedia web API. CoGhent will give access to the Linked Data via:

- a client to harvest and stay in sync with a collection, containing all triples in an entire dataset
- CoGhent will generate human-readable subject pages containing triples about a specific subject in a dataset.
- District09 will harvest the LDES and expose a specific SPARQL endpoint on top of the event streams. This will allow people to query for specific results.

CoGhent adopts the principle of Linked Data Fragments (LDF) to *minimize server resource usage* while allowing clients to have high query execution control, albeit that the way we fragment will be slower because they require retrieval of multiple indexes in order to answer certain questions.

Metadata is retrievable by their identifier using a standardised communications protocol

A1: CoGhent uses the Web as information space following the Linked Open Data principles. This means HTTP(S) is used as a standardized communications protocol. The protocol is open, free, and universally implementable and allows for an authentication and authorisation procedure, where necessary. Exact conditions under which data is accessible are provided (authentication).

By default: we work together with the city of Ghent to create resolvable URIs. Ghent harvests all triples from the event streams so it can create subject pages for the objects. If foaf:page is available, it will redirect.

If an institution wants to use URIs on a different domain, this is possible by updating the Adlib field with the URI. Then the resolving is responsibility of this other domain.

Metadata is accessible, even when the data is no longer available.

A2: End users of the project must **ensure open access to the deposited version** of their published material via the chosen repository. Open access should be provided as soon as possible.

⁸ unless the licensing mentions otherwise

CoGhent will publish data as Linked Data Event Streams (LDES). The implementation of this technology has the following advantages:

- LDES lowers the server-side cost
- LDES exposes harvestable immutable objects
- The retention policy can be defined on the LDES so third parties can know how long a certain version will be available and can anticipate on that knowledge graph to expose their own archive.

The LDES Linked Data Event Stream (LDES) will be archived so it will be retrievable even when the data is no longer available.

2.2.3 Making data interoperable

2.2.3.1 Descriptive metadata from CMS

As mentioned before, the descriptive metadata is optimized as close to the source as possible, meaning that the optimizations are made in the CMS itself. To increase the interoperability of the descriptive metadata, the thesauri of the participating cultural heritage institutions are linked to **external authorities**. The Data Intervention Team reconciles the thesauri to following external authorities, in this order: Getty Vocabularies (Art & Architecture Thesaurus (AAT), Getty Thesaurus of Geographic Names (TGN), Getty Union List of Artist Names (ULAN)), RKD (database of the Netherlands Institute for Art History) and Wikidata. The Data Intervention Team also scans the national cultural heritage landscape for other authorities with persistent identifiers lists, e.g. Onroerend Erfgoed.

Apart from the reconciliation with external authorities, the descriptive metadata is optimized based on the international **standard** for the description of cultural heritage, namely CIDOC-CRM (see 2.2.4 Making data reusable). Additionally, CoGhent makes use of **OSLO cultuur object** and **OSLO cultuur event** as an application profile for describing and modelling the cultural heritage data which is mainly based on CIDOC-CRM⁹ and Linked.art¹⁰. *Open Standards for Linking Organizations* (OSLO) wants to ensure more coherence, better understandability and findability of information and services and is a commitment by the Flemish Government to an unambiguous standard for the exchange of information. Because both OSLO cultuur standards are still in review mode and a pending candidate standard, the project can still make new suggestions for further expanding the model. The project will for example suggest the standard to refer to IIF resources.

Lastly, the descriptive metadata is allocated a **persistent identifier**. A policy for the use of PIDs is necessary in order to have an efficient digital infrastructure and to achieve optimal interoperability, leading to open data, open collections and efficient resource management (see 2.2.1 Making data findable).

2.2.3.2 Data from file servers cultural heritage institutions

⁹ <http://www.cidoc-crm.org/>

¹⁰ <https://linked.art/>

For access to image-based resources the **International Image Interoperability Framework (IIIF)** will be used. In adopting this framework we want to provide rich and uniform access to image-based resources around the world. Make use of a set of common APIs that support interoperability between image repositories. And make use of shared technologies that provide user experience in viewing, comparing, manipulating and annotating images.¹¹ The project will implement the **IIIF Presentation API 3.0**.¹²

2.2.3.3 Data input by end users

The descriptive metadata added by the end users is reconciled with external authorities. This is done manually by the Data Intervention Team.

IIIF is implemented at the digital reproductions added by the end users.

2.2.4 Making data reusable

To increase the reusability of data, a detailed **provenance** is necessary. In section 1.2 and 1.3 a detailed description has already been given. Also, the use of domain-relevant **standards** is crucial (see 2.2.3 Making data interoperable). And lastly, the data needs to be **cleared** from any **copyright**. In this section we focus on the right clearance of the data.

2.2.4.1 Descriptive metadata from CMS

There are no copyright restrictions in reusing existing metadata as **descriptive metadata** will always be published under a **CC0 license**.

2.2.4.2 Data from file servers cultural heritage institutions

First an analysis was made of the current workflow concerning right clearance at the participating cultural heritage institutions. This showed that two out of five institutions already used fixed licences for their digital reproductions: Public Domain Mark, CC0 and In Copyright - non-commercial use permitted. Based on this, CoGhent made a framework containing 6 licences that cover the needs of the institutions.

The participating cultural heritage institutions clear the rights on their digital reproductions before letting the data enter the data stream. The Data Intervention Team supports the cultural heritage institutions in clarifying and clearing the rights of the digital reproductions. In this context, CoGhent developed a workflow, together with the cultural heritage institutions and meemoo.

2.2.4.2.1 Workflow rights clearance

The physical object is the starting point for the determination of copyright.

¹¹ <https://iiif.io/about/>

¹² <https://iiif.io/api/presentation/3.0/#table-of-contents>

- physical objects in public domain: distinction between 2D and 3D objects. Reproductions of 2D objects in public domain hold a Public Domain Mark (based on art.14¹³ of the EU directive¹⁴ digital reproductions of physical objects in the public domain remain in the public domain.) Reproductions of 3D objects → new copyright is generated, creator of the reproduction can claim copyright and ascribe a licence to the reproduction.
- physical objects under copyright, when rights-holders are known and contactable: rights-holders are asked to renounce their property rights (for commercial use) by ascribing a Creative Commons licence.
- physical objects under copyright, rights-holders unknown or not contactable: Rights Statements

The selection of used licenses is made in cooperation with the participating cultural heritage institutions. Every institution has its own risk analysis and takes full responsibility in case of violation of copyright.

<p>Public Domain</p> <p>URI: <edm:rights rdf:resource="http://creativecommons.org/publicdomain/1.0/"></p>	<p>Data is public domain when there are no intellectual property rights applicable. Data can enter the public domain in different ways (expired copyrights, creators give up copyrights, non-existent copyright protection), so an investigation is necessary everytime there is an assumption of public domain data.</p>
<p>CC0</p> <p>URI: <edm:rights rdf:resource="http://creativecommons.org/publicdomain/zero/1.0/"></p>	<p><u>CC0</u> (aka CC Zero) is a public dedication tool, which allows creators to waive their copyright and put their works into the worldwide public domain. CC0 allows reusers to distribute, remix, adapt, and build upon the material in any medium or format, with no restrictions.</p>

¹³ 'Member States shall provide that, when the term of protection of a work of visual art has expired, any material resulting from an act of reproduction of that work is not subject to copyright or related rights, unless the material resulting from that act of reproduction is original in the sense that it is the author's own intellectual creation.'

¹⁴ <https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32019L0790&from=EN>

<p>CC BY-SA 4.0</p> <p>URI: <edm:rights rdf:resource="http://creativecommons.org/licenses/by-sa/4.0/"></p>	<p>You are free to share (copy and redistribute the material in any format or medium) and adapt (remix, transform, and build upon the material for any purpose, even commercially), under the following terms:</p> <ul style="list-style-type: none"> • You must give <u>appropriate credit</u>, provide a link to the license, and <u>indicate if changes were made</u>. You may do so in any reasonable manner, but not in any way that suggests the licensor endorses you or your use. • If you remix, transform, or build upon the material, you must distribute your contributions under the <u>same license</u> as the original. • You may not apply legal terms or <u>technological measures</u> that legally restrict others from doing anything the license permits.
<p>CC BY-NC-ND 4.0</p> <p>URI: <edm:rights rdf:resource="http://creativecommons.org/licenses/by-nc-nd/4.0/"></p>	<p>You are free to share (copy and redistribute the material in any format or medium), under the following terms:</p> <ul style="list-style-type: none"> • You must give <u>appropriate credit</u>, provide a link to the license, and <u>indicate if changes were made</u>. You may do so in any reasonable manner, but not in any way that suggests the licensor endorses you or your use. • You may not use the material for <u>commercial purposes</u>. • If you <u>remix, transform, or build upon</u> the material, you may not distribute the modified material. • You may not apply legal terms or <u>technological measures</u> that legally restrict others from doing anything the license permits.
<p>In copyright – non-commercial use permitted</p> <p>URI: <edm:rights rdf:resource="https://rightsstatements.org/page/InC-NC/1.0/?language=en"></p>	<p>You are free to use this item in any way that is permitted by the copyright and related rights legislation that applies to your use</p> <ul style="list-style-type: none"> • For <u>commercial uses</u>, permission from the rights-holder(s) needs to be obtained

In copyright - rights-holder(s) unlocatable or unidentifiable	onder voorbehoud (voor DMG)
--	-----------------------------

2.2.4.3 Data input by end users

For the data input by end users under copyright, a CC0 licence is proposed at all times. The end user is thoroughly informed about the requirements and consequences of a CC0 license via the proclaimer published on the web platform.

3. Privacy, ethics, trust and security

3.1 Privacy

The data (both descriptive metadata and digital reproductions) from the participating cultural heritage institutions are cleared and don't violate privacy or ethical issues. The challenge for CoGhent is the data from the end users. End users are well informed via the privacy policy and proclaimer, consultable via the CoGhent web platform. End users give **informed consent** to use and reuse their data when they create a profile via the web platform. In this chapter, the focus is on **personal data**¹⁵.

CoGhent takes into account *“the right to the protection of personal data and provides that such data is processed fairly for specified purposes and on the basis of the consent of the person concerned or some other legitimate basis laid down by law, and subject to control by an independent authority.”*¹⁶

CoGhent processes **first hand personal data** and **second hand personal data**.¹⁷

First hand personal data is collected and processed¹⁸ in the context of the project, therefore CoGhent will comply with the GDPR rules and legislation. Personal data is received **first hand** in the context of the crowdsourcing initiatives and the creation of a login for the CoGhent web platform. The data is collected only for the purposes of the CoGhent project. In compliance with the GDPR legislation, CoGhent (or the city of Ghent) will gain consent when required and give the mandatory information as defined in the GDPR legislation¹⁹.

Personal data is received **second hand** if it is included in the data input of the participating cultural heritage institutions. The legal entity responsible for CoGhent is the City of Ghent, therefore the City of Ghent will close a Data Processing Agreement with every individual partner when processing second hand personal data provided by the participating cultural heritage institutions.

In compliance with the GDPR legislation, CoGhent gives the mandatory information as defined in the GDPR legislation²⁰.

¹⁵ Personal data is defined as “any information relating to an identified or identifiable natural person (‘data subject’); an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person” (Art. 4.1 of the GDPR legislation).

¹⁶ Article 8 of the Open Data Charter

¹⁷ A register of all processing activities is attached to the Data Management Plan (art.30 of the GDPR)

¹⁸ Processing "means any operation or set of operations which is performed on personal data or sets of personal data, whether or not by automated means, such as collection, recording, organisation, structuring, storage, adaptation or alteration, retrieval, consultation, use, disclosure by transmission, dissemination or otherwise making available, alignment or combination, restriction, erasure or destruction" (Art. 4.2 of the GDPR).

¹⁹ Art.13 of the GDPR legislation

²⁰ Art.14 of the GDPR legislation

3.1.1 First hand personal data

Personal data will not be reused, only stored for the time needed. If personal data is provided by the user, the project will check and comply with the GDPR rules and legislation. Most personal data will only be processed within the project and thus not made public or reused unless there's legal ground to do so.

3.1.2 Second hand personal data

CoGhent processes personal data from professional providers. The personal data is second hand data, not collected by CoGhent but by the cultural heritage institutions. CoGhent closes an agreement to process the data, the institutions remain the administrators of the data.

When reusing existing data, the project will check and comply with the conditions of access and use. Terms and conditions of access and use depend on the nature of the data. At best these terms are made explicit in advance. If not the project will do the necessary research to trace the rights and permissions.

Institutions are data partners, who must make best efforts to ensure that reuse of existing data by CoGhent does not constitute an unlawful act towards a third party, including but not limited to:

- violation of intellectual property rights of a third party
- an infringement of personality, privacy, publicity or other rights; or
- an infringement of public order or morality (hate speech, obscenity, etc.).

3.2 Ethics

At this point, the main focus is on data sources and transformations. In the course of the project, when the data input by the end users is activated, ethical aspects can surface.

3.3 Trust

The success of the CoGhent project is dependent on the trust that both the cultural heritage institutions and the end users have in the management of their data. A responsible and correct disclosure of data is indispensable.

It's important to execute a correct rights clearance policy.

3.4 Security

The data managed in the CoGhent project doesn't require additional security measurements; the standard measures, as applied by District09, are sufficient. In the future, because of the bulk of the data, new security risks can occur but to date there are no indicators available.

CoGhent doesn't collect special categories of personal data, therefore standard security measures are sufficient.

Upon handover of the LDES and DAMS, District09 will take into account data security. Version 2 will further elaborate on the strategy and activities in place to ensure data security.

4. Roles and responsibilities

The determination of the roles is based on the model developed by the city of Ghent. For the CoGhent project and based on the current situation, there are 7 roles defined:

For cultural heritage data:

- **data owner:** a fixed role, refers to the cultural heritage institutions. Data ownership is only possible if it concerns cultural heritage data and is emphasized by its provenance. It's not possible for third parties to own personal data.
- **data responsible:** a variable role, during the project it refers to the WP5 lead. In project mode, the data responsible has the final responsibility of all data.
- **operational data steward:** a variable role, during the project it refers to both the data team and the data professionals of the cultural heritage institutions. The data steward is situated on the operational level, has a hands-on role and is responsible for (the optimization of) data sets.
- **coordinating data steward:** a variable role, during the project it refers to the coordinator of the data team.
- **data engineer:** during the project, it refers to Inuits and District09, who export, transform and import data.

For personal data:

- **data controller:** a variable role; during the project it refers to Stad Gent, AGB Erfgoed and AGB Kunsten en Design, who are joint data controllers when it comes to second hand personal data (see 3.1.2). For first hand personal data, Stad Gent is the data controller (see 3.1.1). Data controllers determine the purpose and the means of the processing of personal data.
- **data processor:** a variable role; in project mode, it refers to Inuits and District09. Data processors execute processings determined by the data controllers.

There are two levels where roles and responsibilities need to be determined: the **legal** level and the **operational** level. The legal level includes personal data, copyright issues and data ownership. The operational level includes the day-to-day handling and management of the data. Most roles exist in regular mode, in project mode and post-project mode. Who is taking up what roles and responsibilities may vary **before, during and after the CoGhent project**.

4.1 Legal

4.1.1 Descriptive metadata from CMS

Before, during and after the project, the cultural heritage institutions remain the **data owners** of the descriptive metadata, meaning that the institutions are in control of the original data sources, guard the integrity of the data and have the copyrights to the metadata. What does change, is the right of use of the data, once the cultural heritage institutions have the intention to join the project and the data needs optimization and transformation to meet the project's requirements. The descriptive metadata transforms into linked open data, usable to third parties without any legal restrictions.

4.1.2 Data from file servers cultural heritage institutions

Before, during and after the project, the cultural heritage institutions remain the **data owners**, meaning that the institutions are in control of the original data sources and guard the integrity of the data. What does change, is the right of use of the data, once the cultural heritage institutions have the intention to join the project and the data needs optimization and transformation to meet the project's requirements.

A major difference with the descriptive metadata is that the data from the file servers are digital reproductions of collection items where copyrights apply, either on the items or on the reproductions, and that the institutions don't always have these copyrights. Right clearance is discussed in detail under 2.2.4.2.

4.1.3 Data input by end users

Data input by end users is generated in project mode. For this data, the CoGhent project is the **data owner** and the work package 5 lead is **data responsible**. In 2023, when the project ends, there will be a transfer of the ownership and the responsibility.

4.1.4 LDES client

For data generated in project mode, the CoGhent project is the **data owner** and the work package 5 lead is **data responsible**. In 2023, when the project ends, there will be a transfer of the ownership and the responsibility.

4.1.5 Personal data

As mentioned above, the project makes a difference between first hand and second hand personal data (see 3.1). Stad Gent is the data **controller** of first hand personal data. Stad Gent, AGB Kunsten en Design and AGB Erfgoed are **joint data controllers** of second hand personal data. In case of joint data control, a protocol is drafted.

4.2 Operational

4.2.1 Descriptive metadata from CMS

During the project, when the data is optimized and transformed to meet the project's requirements and to enter the LDES, the data intervention team is the **operational data**

steward. The team is led by a **coordinating data steward**. The final responsibility lies with the work package 5 lead.

In project mode, Inuits is the **data engineer**. Inuits exports the data and transforms it into linked open data via the LDES before importing it into DAMS. After the project, District09 takes over from Inuits as data engineer.

4.2.2 Data from file servers cultural heritage institutions

During the project, when the data is optimized and transformed to meet the project's requirements and to enter the LDES, the data intervention team is the **operational data steward**. The team is led by a **coordinating data steward**. The final responsibility lies with the work package 5 lead.

The transformation of the data into IIIF standard is done by the **data engineer**; in project mode, this is Inuits. After the project, District09 takes over.

4.2.3 Data input by end users

Data input by the end users is directly in DAMS. In project mode, the necessary optimization is done by the Data Intervention Team, as **operational data stewards**. How the data input by the end users will be handled after the project is still in the research phase.

In project mode, the work package 5 lead is **data responsible** for the data input by the end users.

4.2.4 LDES client

On the operational level and in project mode, District09 provides storage for the LDES preservation.

Inuits, as **data engineer**, initiates the LDES.

4.2.5 Personal data

Inuits and District09 are **data processors** of personal data. In case of data processing, an agreement is concluded.