DELIVERABLE

Project Acronym:	**EAGLE**
Grant Agreement number:	**325122**
Project Title:	**Europeana network of Ancient Greek and Latin Epigraphy**

Report on the contributions to Europeana – first release

Deliverable 3.3.1

version: 1.0

Revision: final

Authors:

> **Valentina Vassallo (The Cyprus Institute)**

Contributors:

> **Mark Depauw (K.U. Leuven)**
> **Pietro Liuzzo (UHEI)**

Reviewers:

> **Claudio Prandoni (Promoter)**

Revision History

Revision	Date	Author	Organisation	Description
0.1	29/09/2014	Valentina Vassallo	The Cyprus Institute	First version
0.2	14/10/2014	Mark Depauw Pietro Liuzzo	K.U. LEUVEN UHEI	Contributions to the first version
1.0	15/10/2014	Claudio Prandoni	PROMOTER	Corrections, improvements and final check

Statement of originality:

This deliverable contains original unpublished work except where clearly indicated otherwise. Acknowledgement of previously published material and of the work of others has been made through appropriate citation, quotation or both.

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

ABBREVIATION	EXPLANATION
CP	Content Provider
DoW	Description of Work
EDM	Europeana Data Model
WP	Work Package
WG	Working Group
DEA	Data Exchange Agreement
TM number	Trismegistos number

EXECUTIVE SUMMARY

This Deliverable, created by the Cyprus Institute (CyI), with inputs opportunely cited, reports about the contribution to Europeana. Particularly focuses on the progress of the ingestion of EAGLE content towards Europeana, providing updates in the workflow and in the procedure of monitoring.

The document is structured in 6 Chapters: an introduction about the objectives of this Deliverable and its role in the project; a section dedicated to the content and their collections. The part dedicated to the ingestion plan and the importance of implementing and monitoring the ingestion through a structured workflow, different sources of information and the collaboration with others WPs and Tasks. Finally, the part dedicated to the effective report on the contributions from EAGLE to Europeana.

The Introduction of the Deliverable explains the aims of the Work Package and its role in the project. The first Chapter is an overview of the objectives of WP3: it describes the tasks of the WP3 and its overall objectives, focussing also on its Milestones.

To facilitate the lecture of the document, Chapter 2 reports the list of the Content Providers that are involved in the EAGLE Consortium plus the new affiliated partners.

Chapter 3 focuses on the importance of the ingestion plan form in order to carry out the work of monitoring the ingestion itself and on the status of the Data Exchange Agreement, fundamental step for the final publication into the European portal.

Chapter 4, together with Chapter 5, are the core of this Deliverable. Particularly Chapter 4 illustrates, according to the related Task, the implementation and monitoring of the ingestion through the update of the ingestion workflow, the establishment and monitoring of internal milestones, a structured feedback system and through the monitoring of all the Tasks of the WP3 as well as the work of the other EAGLE WPs.

Finally, Chapter 5 is dedicated to the report on the contributions for Europeana, with a detailed description of the work done by each Content Provider and the data ready to be harvested by Europeana.

Chapter 6 draws the conclusions and gives information about the next steps of the Work Package 3.

1. INTRODUCTION

This report provides an overview of the activities carried out by the Work Package 3 during the period 1st April – 30th September 2014. The work reported in this Deliverable is strictly connected with the Deliverable 3.2, since they are parts of the same work aimed at development of a methodology and of a workflow for the EAGLE content contribution into Europeana. The report illustrates all the steps of the workflow. Starting from the Content Providers planning of the content it will follow the progress of the ingestion, the feedback collected and the monitoring of the contributions activities towards Europeana. All the topics already treated in the previous issue (Deliverable 3.2 'Ingestion plan') will be cited but not repeated in this Deliverable unless they are useful for the general comprehension or there are noteworthy changes and updates in the structure.

1.1 OBJECTIVES OF THE WP3 'METADATA MODEL, MAPPING AND INGESTION'

WP3 'Metadata model, mapping and ingestion' is the core of the EAGLE project since it has to deal with the contribution of quality metadata and content from the individual institutional partners to the EAGLE Aggregation and Image Management (AIM) infrastructure implemented within WP4 ('EAGLE Aggregation and Image infrastructure'). In the WP3, the common EAGLE metadata structure to ensure full interoperability with Europeana has been defined (see Deliverable D3.1 'EAGLE metadata model specifications'). This is the basis for the local mappings. The work within the Work Package addresses all the specific activities that are needed locally to harmonize the content coming from the different sources and archives of the project partners and prepare it for the ingestion process. The work consists of the implementation of the agreed standards and protocols, the preparation of the necessary mappings for the metadata conversions, the ingestion plan itself and finally the publication of the content into Europeana portal, including appropriate disambiguation. The role of the Content Providers is that one of checking, enrich and contextualize their metadata thanks to the data curation developed within WP4.

The Cyprus Institute is the leader of the Work Package and leader of the tasks 'Planning and preparing the ingestion' and 'Implementing and monitoring the ingestion'. Other 3 Task leaders are associated: UAH ('Definition of the EAGLE metadata model' and 'Metadata mapping'); K.U.LEUVEN ('Duplicates identification') and UNIROMA1 ('Metadata enrichment and contextualization'). All the Content Providers contribute to the mapping and the ingestion of their content to the EAGLE Aggregator and to Europeana.

The overall objectives of the Work Package 3 are:

- To achieve the large scale implementation of agreed standards and best practices at the local and institutional level across the Best Practices Network (BPN)

- To define the common EAGLE metadata structure to enable BPN metadata harmonization

- To define mapping from local BPN metadata structures to the common EAGLE metadata structure

- To define mapping from EAGLE metadata structure to Europeana Data Model (EDM), to ensure full interoperability with Europeana.

- To ingest content into EAGLE and Europeana

- To check, enrich and contextualize the aggregated metadata

- To improve the quality of content and metadata across the BPN

The WP3 is in charge to organize the ingestion plan and the aggregation workflow from Content Providers to Europeana passing through EAGLE Aggregator, to monitor the work progresses and to assure that the Tasks and Milestones will be satisfactorily achieved.

The Milestones are:

- Month 18 (September 2014): 375.000 items available for Europeana (MS8)

- Month 24 (March 2015): 750.000 items available for Europeana (MS12)

- Month 30 (September 2015): 1.125.000 items available for Europeana (MS15)

- Month 36 (March 2016): 1.500.000 items available for Europeana (MS16)

The Deliverable 3.3.1. 'Report on the contributions to Europeana' illustrates the ongoing aggregation process, focussing on the implementation and monitoring of the ingestion and the effective data contribution to EAGLE and to Europeana. The cooperation between WP2 and WP4 is fundamental to carry out the WP3 tasks.

This document reports about the work carried out from month 13 (April 2014), to month 18 (September 2014).

1.2 ROLE OF THE DELIVERABLE 3.3.1 IN THE PROJECT

The aim of the Deliverable 3.3.1 is the report on the effective contribution from the EAGLE Content Providers to Europeana. This Deliverable is very important to assess the implementation and monitoring of the ingestion workflow. From now on, all the next Deliverables will report on the effective contribution of the content to the EAGLE infrastructure and to Europeana. They are strictly connected with the achievement of the ingestion procedure and of the Milestones: MS8 (Month 18) with the ingestion of 375.000 items available for Europeana; MS12 (Month 24) with the ingestion of 750.000 items available for Europeana; MS15 (Month 30) with the ingestion of 1.125.000 items available for Europeana; MS16 (Month 36) with the ingestion of 1.500.000 items available for Europeana. Table I reports the schedule of the WP3 Deliverables and Milestones (in light blue, the current report period):

MONTH	12	18	24	30	36
DELIVERABLE	D3.2	D3.3.1	D3.3.2	D3.3.3	D3.3.4
MILESTONE	-	MS8	MS12	MS15	MS16

Table I. The schedule of the WP3 Deliverables and Milestones.

2. CONTENT PROVIDERS AND DIGITAL COLLECTIONS

2.1 CONTENT PROVIDERS

The project Consortium, as described in the Description of Work, is composed of 19 partners from 12 different European countries. Within the EAGLE Consortium, 14 of them are Content Providers and they will supply Europeana with Ancient Greek and Latin inscriptions with related metadata: 1.5 M images and related metadata, (including translations of selected texts in Wikimedia) for the benefit of the public.

To facilitate the lecture of the Deliverable, the list of the EAGLE participating Partners with the specification of the Content Providers is reported in Table II.

N.	Organisation	Country	Content Provider
1	UNIVERSITA DEGLI STUDI DI ROMA LA SAPIENZA	ITALY	✓
2	PARIS-LODRON-UNIVERSITÄT SALZBURG	AUSTRIA	✓
3	KATHOLIEKE UNIVERSITEIT LEUVEN	BELGIUM	
4	THE CYPRUS RESEARCH AND EDUCATIONAL FOUNDATION	CYPRUS	✓
5	EUREVA SAS	FRANCE	
6	CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE	FRANCE	✓
7	RUPRECHT-KARLS-UNIVERSITAET HEIDELBERG	GERMANY	✓
8	DEUTSCHES ARCHAOLOGISCHES INSTITUT	GERMANY	✓
9	EÖTVÖS LORÁND TUDOMÁNYEGYETEM	HUNGARY	✓
10	UNIVERSITA DEGLI STUDI DI BARI "ALDO MORO"	ITALY	✓
11	THE BRITISH SCHOOL AT ROME	ITALY	✓
12	PROMOTER SRL	ITALY	
13	GOGATE SRL	ITALY	
14	CONSIGLIO NAZIONALE DELLE RICERCHE	ITALY	
15	SVEUCILISTE JURJA DOBRILE U PULI	REPUBLIC OF CROATIA	✓
16	ZNANSTVENORAZISKOVALNI CENTER SLOVENSKE	REPUBLIC OF	✓

	AKADEMIJE ZNANOSTI IN UMETNOSTI	SLOVENIA	
17	UNIVERSITATEA BABES BOLYAI	ROMANIA	✓
18	UNIVERSIDAD DE ALCALA	SPAIN	✓
19	THE CHANCELLOR, MASTERS AND SCHOLARS OF THE UNIVERSITY OF OXFORD	UNITED KINGDOM	✓

Table II. List of the EAGLE Partners and Content Providers.

As already reported in Deliverable 3.2 "Ingestion plan", the content is constituted of digital images (photographs, drawings, scanned notes, manuscripts and archive material, ancient books) and their related metadata. At present, there are no changes or updates regarding the type of collections that will be provided to the EAGLE project with respect to what we reported in D3.2.

2.1.1 New affiliated Partners

One of the aim of WP2 'Networking and best practices' is to attract new content providers to the EAGLE network, in order to increase the quantity and quality of the EAGLE digital resources and to ensure the wide array of scattered collections is covered appropriately. Content providers from outside the Consortium will be encouraged to join the Best Practice Network and to contribute their content. All the content providing partners fulfil an important role in awareness-raising and recruiting new content providers across Europe.

WP3 will assist WP2 to carry out these activities of consortium enlargement and to elaborate an action plan for the institutions that have shown interest in participating to the project activities. WP3 will monitor the involvement of the stakeholders that will contribute with content to Europeana[1].

With respect to what we reported in March 2014, one additional affiliated partner signed the agreement to provide content to the EAGLE project. Therefore, the full list of new content providers is the following:

- Pontificia Commissione di Archeologia Sacra
- The Israel Museum, Jerusalem
- Attic Inscriptions Online
- University of Beograd
- Città di Anagni

Moreover, during the present reporting period, different institutions have declared their interest in participate and contribute to the project in the occasion of several EAGLE dissemination events (e.g. CIDOC 2014 conference, Wikimania)[2] and contacts are ongoing to include them too as new EAGLE content providers.

---

[1] For the methodology and the activities about the enlargement of the network, please refer to the Deliverable D.2.1 'Networking infrastructure and terms of reference'.

[2] For a detailed list of the events and a report about the EAGLE consortium enlargement activity, please refer to the WP1 documents.

3. THE INGESTION PLAN

The tasks of the WP3 started at M1 (April 2013) and will last for the whole duration of the project (M36, March 2016)

The ingestion plan consists of all the activities on which is based the EAGLE aggregation procedure and the content delivery to Europeana: surveys, training sessions, partners support, procedures, methodologies, feedback and milestones. WP3 aims at the planning and preparing the ingestion and at its implementation and monitoring.

The task T3.3. ''Planning and preparing the ingestion'', was concluded in March 2014 and all the Content Providers participated in the WP carrying out work on selecting the content, harmonising the metadata, enriching and creating additional metadata, preparing the master object files for ingestion in the repository, preparing the surrogates, applying identifiers, etc. This task is connected with T3.4. "Implementing and monitoring the ingestion'', where the Content Providers will apply the best practices and implement locally the necessary workflows and procedures to ensure the continuous provision of metadata, surrogates and digital content through the EAGLE Aggregator (WP4) to Europeana.

The task of the WP3 leader is to ensure that the Content Providers make their content available according to the project schedule and that they provide the content declared. Therefore, WP3 leader, from month 12 (March 2014) started to operate the monitoring of the content ingestion into Europeana and its progress. From this Deliverable onwards, every 6 months WP3 will report about this monitoring activity and on the contribution to Europeana.

3.1 THE INGESTION PLAN FORM

As described in the Deliverable 3.2 "Ingestion plan", in order to perform the WP3 Tasks of keeping under control the content delivery to Europeana and to report about contributions as declared in the Description of Work (DoW), a working tool has been elaborated. The ingestion plan form is in fact useful to keep trace of the Content Providers information about the items they will provide to Europeana, along all the project period.

The tool is used to confirm or update the content declared in the EAGLE Description of Work, to describe the status and the availability of the collections to be ingested and to trace the ingestion progress. In order to maintain updated the monitoring activity during the months, the ingestion plan form has been provided with new fields for the related periods to be monitored (see Table III).

The ingestion plan form data are cross-checked together with the information gathered with other tools (reported in this Deliverable) and coming from the feedback of all the project's actors. The result is a synthesis of all the analytical reports of the Content Providers' contributions to EAGLE and Europeana.

At the beginning, the ingestion plan form was updated following the activities of launching and re-elaborating the information gathered through two Surveys (March 2013 and January 2014). Due to the importance to keep under control the general amount of the EAGLE contribution to Europeana, WP3 is planning to organise another Survey (a *tour de table*) for the Content Providers, in order to assess the total amount of each Provider and understand if the project is in line with the contribution foreseen in the DoW, or if any amendments is needed. The Survey will be launched during the next semester, around December 2013 - January 2015. This date has an important meaning because it is one year after the last Survey and just before the next Milestone (MS12, due by Month 24, i.e. March 2015).

No	Short Name	DEA	DoW	Second Survey	MS8 (September 2014)	MS12 (March 2015)
1	UNIROMA1	✓	60.000 images (photographs and drawings)	60.000	…	…
2	…	…	…	…	…	…
…	…	…	…	…	…	…
	TOTAL		1.472.198	1.441.924	…	…

Table III. The ingestion plan form

The *Tour de table* will be included in the activities of the Task 3.4 'Implementing and monitoring the ingestion' for the effective and practical ingestion procedure scope and the update of the information declared by the Content Providers in the Description of Work, as well as to monitor the content to be provided to Europeana. The aim of this activity is in fact to review and assess the content to be aggregated within EAGLE and to retrieve further and updated information about the digital collections that Providers are committed to give to the project.

The Providers will be asked to give detailed information about the collections amount to be provided and if there are any variations respect to the Description of Work; or if they are encountering any delays or problems for the content provision.

As usual, the information gathered will be put at disposal of all the WP3 Task Leaders and of the WP4.

On the basis of the direct contact with the Providers, at the moment (September 2014) we cannot envisage any other change in the total amount. The situation remains stable as reported in the Table IV of the Deliverable 3.2[3].

3.1.1 Data Exchange Agreement (DEA)

Within the WP3 tasks, an important requirement for the content provision and its monitoring is the sign of the Europeana Data Exchange Agreement by all the EAGLE Content Providers.

As already explained in the Deliverable 3.2 "Ingestion plan", every Content Provider, supplying content to Europeana, has to subscribe the Agreement and therefore the WP3 task (together with WP1) is to verify who signed and who is missing in order to keep under control the publication process in Europeana from the ingestion point of view.

During the last 6 months, also the last partner which was missing at M12, (Ausonius) signed the Europeana agreement. Therefore, at September 2014 all the EAGLE Content Providers signed the DEA and their contents are allowed to be transmitted to Europeana for the publication in the portal.

---

[3] The Providers almost confirmed the amount committed in the DoW and the contribution will be performed increasing the amount during all the project time-span.

4. IMPLEMENTING AND MONITORING THE INGESTION

In Task 3.4 'Implementing and monitoring the ingestion', each Content Provider applies the best practices and implements locally the necessary workflows and procedures to ensure the continuous provision of metadata, surrogates and digital content for ingestion to the EAGLE Aggregator (WP4) and to Europeana.

The WP3 leader ensures the monitoring of the progress of the ingestion of new content into Europeana as planned in T3.3. The aim of this task is:

- to ensure that the work of aggregation keeps pace with the releases of Europeana;
- to ensure that the work of metadata mapping keeps pace with the releases of Europeana;
- to ensure that the work of ingestion keeps pace with the releases of Europeana.

Work Package 3 is in charge of observing the content provision from the EAGLE partners to Europeana and of monitoring possible changes or delay, that might occur, controlling that the content will be maintained unchanged as declared in the Description of Work or as updated by the providers during the assessment (surveys, direct contact, etc.).

4.1 THE INGESTION WORKFLOW

This section briefly recaps the workflow implemented within the EAGLE project, the relations among the Work Packages involved (particularly WP3 and WP4), the Content Providers and Europeana team[4]. The workflow starts from the work of the Content Providers, passes through the activities of WP3 and WP4 and arrives to the Europeana ingestion team action, in a cyclic way in case of adjustments and update of the content ingested (Fig. 1).
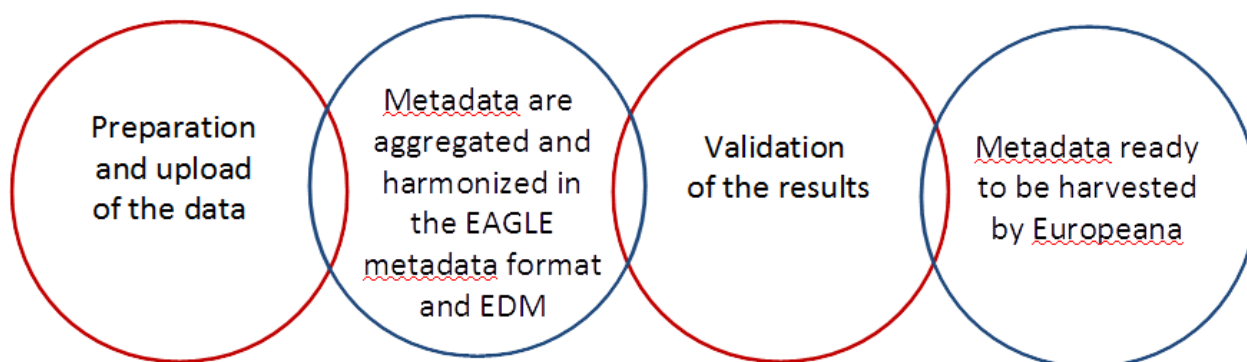

Fig. 1. The ingestion workflow

The workflow is useful to follow how the aggregation and monitoring process is being carried out within EAGLE and which cooperation mechanisms among the project WPs, Content Providers and Europeana is being set up. Some updates and specifications have been included.

The First step of this procedure consists of the work of the Content Providers regarding their collections, their preparation, export and upload of the data. Content Providers have to guarantee the online access to their collections and that they are compliant with the EAGLE and Europeana requirements. For this

---

[4] For the full description, see the Deliverable 3.2 'Ingestion plan'.

reason, the providers have to guarantee that these elements are supplied in their metadata schemas before the mapping into EAGLE and of course that all the resources are available online. Content Providers have to export their metadata from their collection management system in XML format and send them through the export system implemented by each provider (via OAI-PMH, FTP, HTTP) or through the FTP server implemented by WP4 for those they have no export methods (see further in this Chapter), to the EAGLE infrastructure.

The Second step is represented by the aggregation platform developed by CNR-ISTI, leader of the WP4. The platform is a customized version for EAGLE of the D-Net software. It is the web environment that allows WP4 to upload the project Content Providers data, automatically map to EAGLE metadata schema and to EDM in order to send the data to Europeana.

The ingestion service enables:

- the data uploading of the Content Providers' data;

- the mapping and transformations of the metadata records into EAGLE records and the aggregation in the repository;

- the validation of the content;

- the mapping and transformation of the EAGLE data into EDM and the final transmission to the Europeana ingestion via OAI-PMH.

The mapping for the transformation into the common metadata format, within the EAGLE infrastructure, is done on the base of the specifications given by the CPs[5]. After this step, the EAGLE metadata schema is transformed in EDM in order to be harvested by Europeana for the publication in its portal.

The Third step consists of the validation of the results. Content Providers in fact have to check the results of the transformation[6]. The Task 3.2 of preparing the necessary mappings for the metadata conversions in order to allow the alignment of proprietary provider's metadata schemata with the EAGLE schema, started in January 2014 and will last until March 2016. This means that Content Providers have to keep under control the results of their mapping and provide a continuous validation of the results and update of them in case of necessity. WP3, responsible for the metadata mapping together with WP4 and the Content Providers, will maintain updated the metadata mappings until the end of the project. During the period under report, in order to carry out the ingestion, WP3, WP4 and the Content Providers, have been carried out the work of control, validation and revision of the mappings and data for the publication in the EAGLE infrastructure. The feedback procedure (see the next Chapter) helps all the actors to achieve the results smoothly.

The Fourth step is the mapping and transformation of the EAGLE metadata schema into EDM and the consequent Europeana harvesting. After the transformation in fact the metadata are published by WP4 in the OAI-PMH server of EAGLE and made them available for the Europeana harvesting.

During the present reporting period, WP3 and WP4 finalized the definition of the mapping 'EAGLE2EDM',. The result of the elaborated mapping was sent to the Europeana Operation team in order to finalize it and to get feedback from them about the correctness of the result and in order to begin the work of content provision from EAGLE to Europeana. The Europeana feedback procedure is reported in the Feedback procedure section of the current Chapter.

---

[5] The specifications are published in the Deliverable D3.1 "EAGLE metadata model specification".

[6] The EAGLE infrastructure allows the Providers to check the results and allows to check that the files ready to be delivered to Europeana are correctly transformed and are compliant with Europeana technical requirements.

4.2 MONITORING OF THE MID-TERM INTERNAL MILESTONES

According to the ingestion workflow, WP3 asked to all the Content Providers to contribute to the September deadline (Milestone 8), distributing their work during the period April-September 2014[7]. All the Content Providers worked during this period to meet the deadline and sent the content to WP4 in order to upload and map/transform their data in the technical infrastructure.

To monitor the work, it has been asked the Providers to declare in advance the amount of data that they were going to ingest until September 2014. The use of internal intermediate deadlines is aimed at a continuous monitoring and assessment, in order to check the work done and plan the next steps. The comparison of the results achieved with the objectives and milestones foreseen is of great importance for the evaluation of possible bottlenecks, to prevent them and to look for the right solution on time.

As already described in D3.2 'Ingestion plan', the working tool elaborated for this scope will be always circulated among the Providers at predetermined intervals and according to the official Milestones. In this excel table are reported the amount declared by the Providers month by month (see Tab IV).

Mid-term internal deadlines and Milestones							
No.	Short Name	April 2014	May 2014	June 2014	July 2014	August 2014	September 2014
1	UNIROMA1	1.000	4.000	5.000	5.000	0	15.000
2	UNIBA	1.250	3.000	3.000	4.000	0	11.250
3	UHEI	1.000	3.812	3.812	3.812	3.814	16.250
4	UOXF	1	93	93	93	95	375
5	UAH	200	14.637	14.637	14.637	14.639	58.750
6	PLUS	35.805	0	0	0	0	35.805
7	UBB	800	50	50	50	50	1.000
8	ELTE	800	0	0	0	0	800
9	UNIPU	500	0	0	0	0	500
10	ZRC SAZU	108	0	0	0	0	108

---

[7] The strategy chosen of making all the Providers to contribute, is made with the purpose to avoid any postponing by the Providers and force them to contribute equally and well distributed during the project time, trying to avoid "last minute" contribution or delays.

11	**AUSONIUS**	10	310	310	310	310	1.250
12	**DAI**	50.000	50.000	50.000	50.000	50.000	250.000
13	**CYI**	10	25	25	25	25	110
14	**BSR**	1.016	0	0	0	0	1.016
	TOTAL	92.500	75.927	76.927	77.927	68.933	**392.214**

Table IV. Mid-term internal deadlines and milestones as planned by the CPs.

The monitoring of the provision is carried out thanks to several feedback typologies, further detailed in the following paragraphs. Furthermore, in Section 5 'Report on the contribution to Europeana' is reported the effective provision of all the Providers to the EAGLE infrastructure and ready to be harvested by Europeana for the final publication.

4.3. FEEDBACK

The collection of feedback is an important step of the workflow in the contributions to Europeana and in the monitoring of the ingestion procedure. It is in fact part of the monitoring and supporting activities of the Providers' work and of all the actors involved in the contribution to Europeana. This feedback activity consists of different sources of information in order to:

- collect information from the users to face their problems in the aggregation process;
- answer content providers' requests to overcome their difficulties;
- define an effective ingestion methodology and its monitoring.

The feedback collection methodology is part of the general methodology developed for the contribution of data to Europeana. Feedback is important because it helps WP3, WP4 and the Content Providers themselves to refine and improve the ingestion process and to support the contribution of this medium/large amount of content to Europeana.

The feedback is gathered making use of several tools and sources of information, particularly:

- direct contacts with the Content Providers
- training sessions and face-to-face meetings
- the EAGLE FTP
- the Metadata Record Inspector of the EAGLE infrastructure
- Europeana feedback

Thanks to these source of information, the progress of the state of the art of the ingestion is documented and recorded in the working tools elaborated by WP3 (and shared with WP4), where it is possible to trace the collected information. The 'Report on the contribution to Europeana' makes use of all these elements and of the crosscheck of the information from the different tools and actors participating in the project.

4.3.1 Direct contacts

As reported in the Deliverable 3.2 'Ingestion plan', WP3 implemented direct relations with the Providers. WP3 is periodically and constantly in contact with the Content Providers through e-mails and Skype calls in order to monitor the progress of their work, to inform about news and updates or discuss about any kind of issues that they may encounter, clarifications of doubts, etc. These kind of contacts effectively substitute the valuable face-to-face meetings, usually carried out during the plenary meetings of the project. Together with the direct communications, the EAGLE mailing list are used to inform about the general requirements, deadlines and get information about Content Providers general feedback.

All the results of these communications are reported in the working tool elaborated by WP3 to follow the Content Provider ingestion from the beginning of the project (uploading of the content in the EAGLE infrastructure) until the end (the publication in Europeana) in order to monitor the available content and the progress of the project. The working tool is the document that helps to maintain updated the information and the communication among WP3 and Content Providers, drawing a personal history in a way fast to read and to consult[8].

4.3.2 Training sessions and face-to-face meetings

In the occasion of the last EAGLE Plenary meetings (24-25 October 2013, Pisa; 18 February 2014, Ljubljana) WP3 and WP4 organised dedicated sessions in order to train the Content Providers on the EAGLE metadata schema, on the use of the ingestion platform and on the ingestion procedure and aggregation to Europeana. The aim is to keep informed and trained the Content Providers about the technical progress of the project and to give them direct assistance about the technical specifications needed for the delivery of the content to Europeana. Also in Paris, during the EAGLE International Conference and plenary meeting (29 September 2014), WP3, WP4 and the Content Providers could check together the status of the ingestion in the light of the fulfilment of the first milestone MS8.

Further training sessions will be planned, if needed, during the next EAGLE plenary meetings and public events.

All the training material produced for the Content Providers and all the technical presentations are always updated and shared in the EAGLE project repository.

4.3.3 The EAGLE File Transfer Protocol

A dedicated EAGLE File Transfer Protocol (FTP) was established to upload data from the Content Providers that do not have at disposal an OAI, FTP or web APIs to enable WP4 to collect data in a programmatic way. Therefore, pushing content to the EAGLE FTP server substituted the practice to send the content through e-mail or packages, and allowed to have a real time update of the content and of possible lack of data during the substitution of the content in the EAGLE infrastructure operated by WP4.

After a test with a couple of Providers, a guide to upload content to the EAGLE FTP server has been prepared by CNR-ISTI and shared in the EAGLE project repository together with all the needed instructions[9]. The Guide is divided in two sections: the first one is dedicated to the prerequisites needed

---

[8] For a detailed description of the monitoring file, please refer to Deliverable 3.2 'Ingestion plan'.

[9] The FTP Guide is available at the following address of the EAGLE Repository: http://goo.gl/LajFSy.

to upload content; the second one is a step-by-step guide, enriched with with instructions and images, to support the Providers during the installation of the tool and the upload and update of the data.

Through the access to the EAGLE FTP server, WP3 (as well as WP4 and the Content Providers themselves) is able to monitor the upload of the content carried out by each Content Provider (Fig. 2). Every Content Provider can upload its content in its folder; WP3 can visualize the amount of files sent and the last modification done. In this way, WP3 has a clear overview of the amount of data that the Providers prepared and keep trace of their work[10].
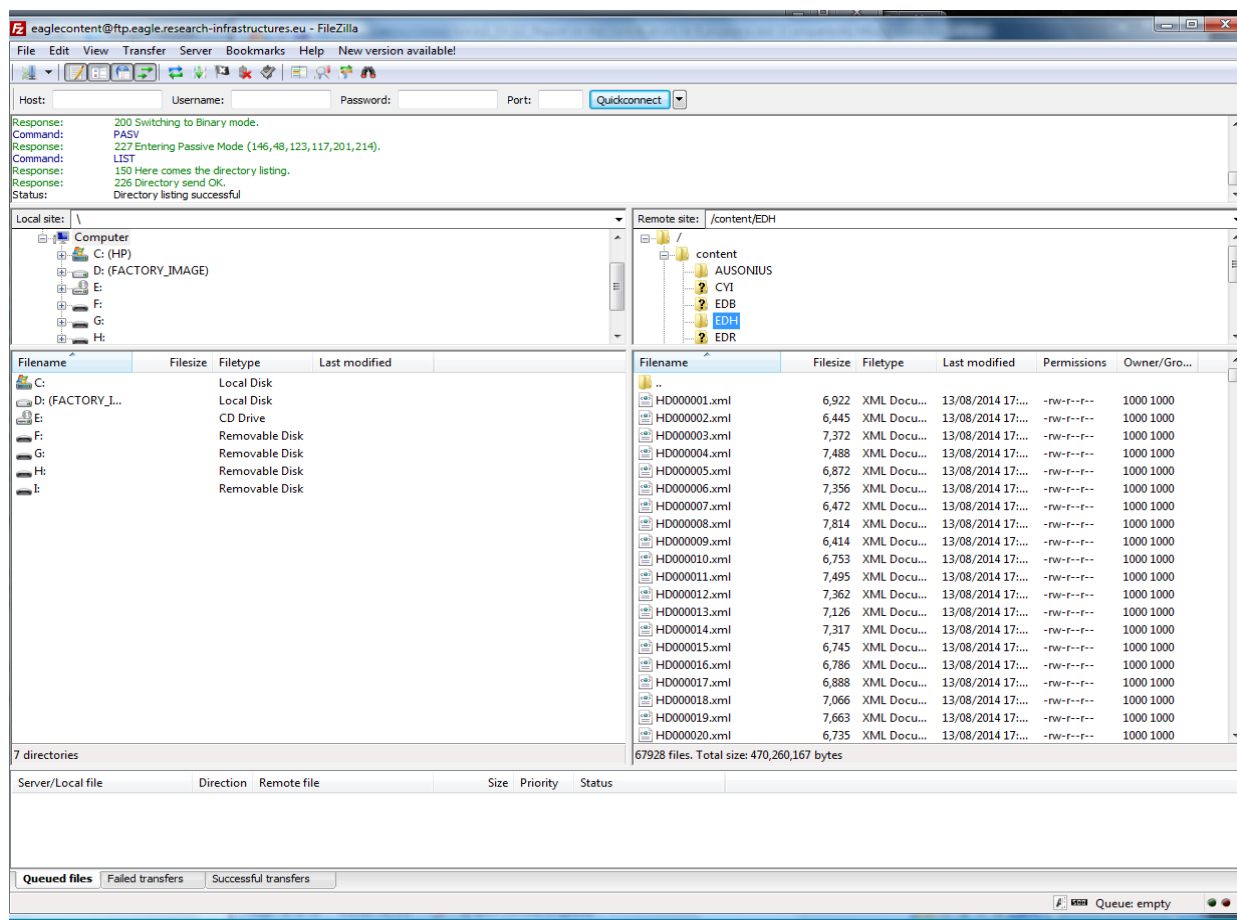
Fig. 2. The EAGLE FTP server
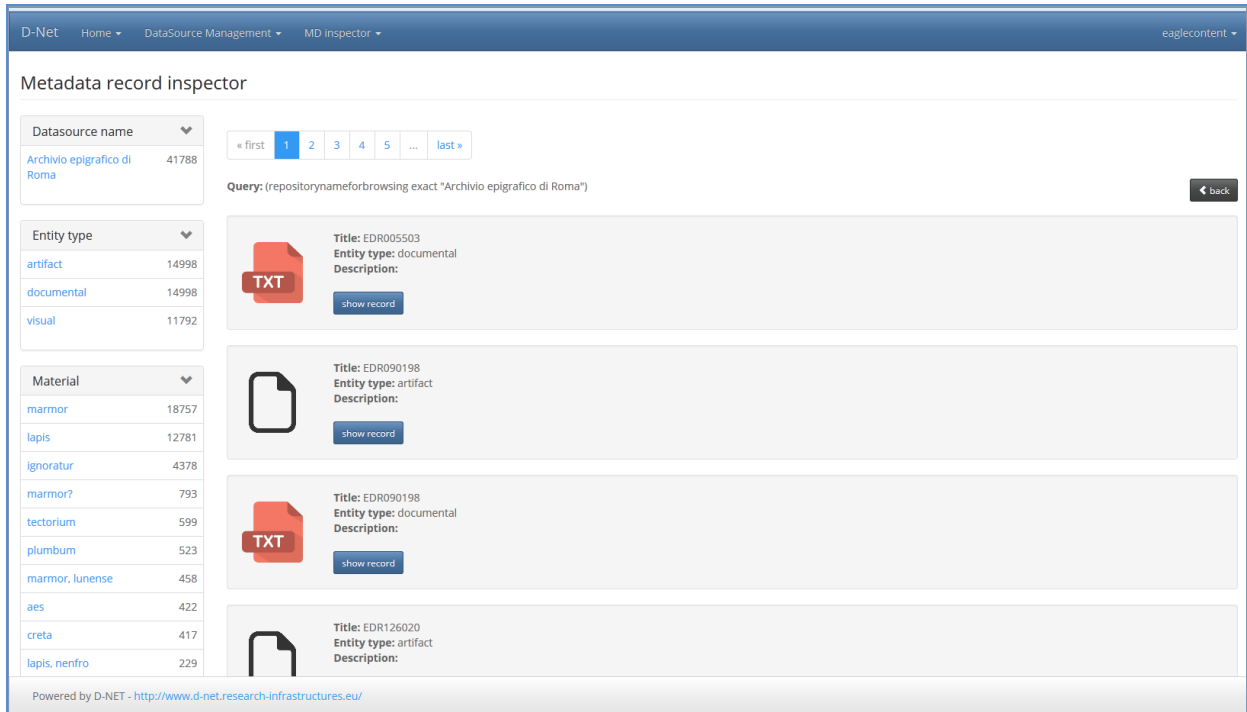
4.3.4 The D-Net metadata record inspector (EAGLE infrastructure)

This source of information is very useful to get an analytical overview of the effective content ingestion.

The D-Net data report functionalities provides features that allow to make detailed reports both for the single content providers and for WP3 and WP4 coordinators: the partners (after accessing with credentials) can have a quick view on the amount of metadata uploaded and transformed in the EAGLE infrastructure; the coordinators can supervise the work of all contributors. The data report offers to the

---

[10] The control of the data amount of the Providers that send their content through OAI server is carried out monitoring them directly.

users the breakdown of the items uploaded in the tool provider by provider and the consequent transformation operated after the mapping to the EAGLE metadata schema (Fig. 3).
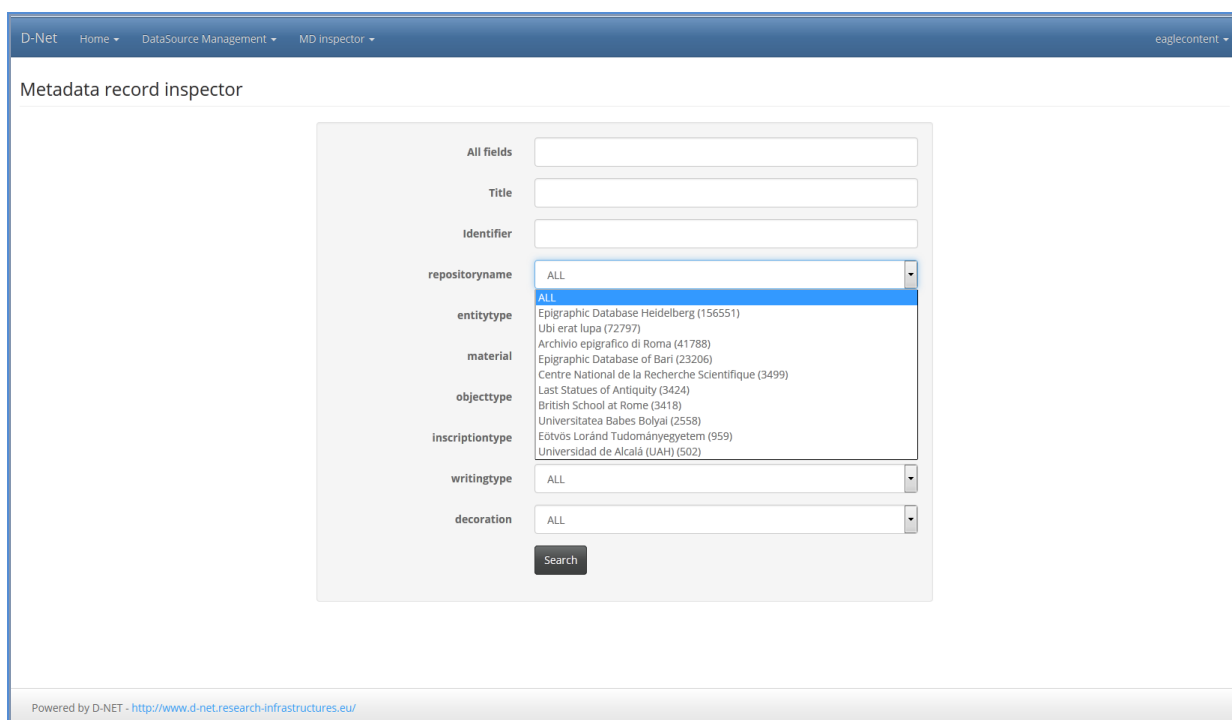

Fig. 3. The metadata record inspector: the amount of data of a Content Provider

WP3 and the Content Providers can also browse the index by using the following search fields (Fig. 4):

- All fields
- Title
- Identifier
- Repository name
- Entity type
- Material
- Object type
- Inscription type
- Writing type
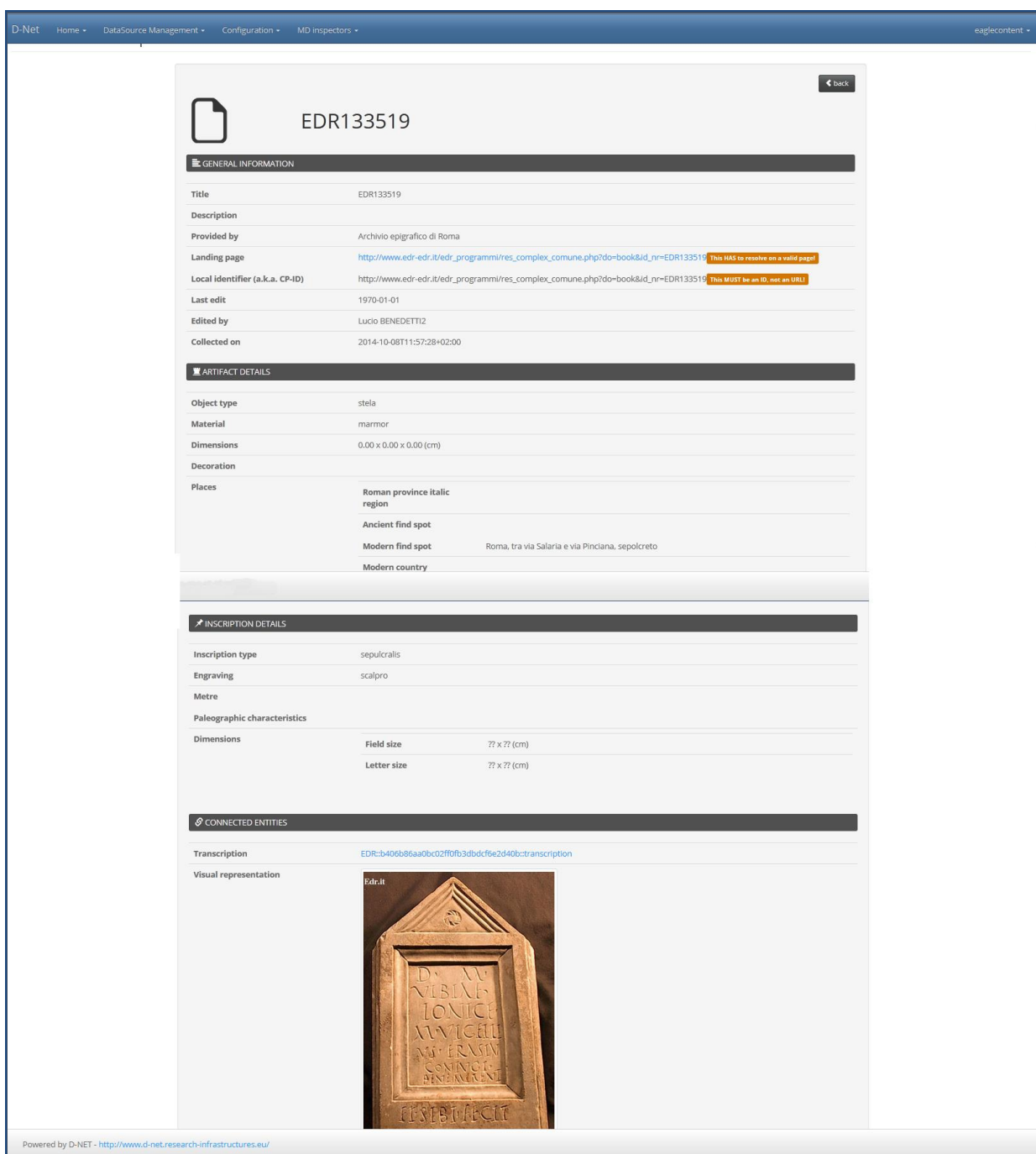- Decoration

Fig. 4. The metadata record inspector in D-Net: the list of the institutions that aggregate data for EAGLE Project.

Finally, the last functionality recently added to the metadata record inspector is the visualization of the metadata details, with the generation and visualization of all the element wrappers that constitute the metadata record (Fig. 5). In this way the CP can operate a quality control of the data and easily identify, through a user-friendly interface, the content provided and eventually adjust or enhance the data.

This feature include also the possibility to check the correlations between the various items (artefacts, documental manifestations and visual manifestations, allowing the Content Providers to verify that the data for the EAGLE portal and for Europeana are correct.

Fig. 5. The visualization of the metadata details and content

4.3.5 Europeana feedback

The Feedback from Europeana is the last important step to ensure that the content provided to the portal complies with Europeana's requirements.

During the present reporting period, a paramount feedback from Europeana Operations team was related to the control of the mapping result from EAGLE to EDM. WP3 and WP4, after the preparation of the mapping to EDM, had technical communications with the contact person for EAGLE project in Europeana. This technical hints and advices allowed to fine-tune and finalise the mapping and allowed us to proceed with the transformation in EDM of the EAGLE content and the harvesting by Europeana.

After the publication in the EAGLE infrastructure, WP3 is in charge to inform the Europeana Operations Team of the datasets that are ready for the harvesting. In agreement with the Europeana team and as already established in other Europeana sisters' projects, WP3 organises the process and the deadlines from the Content Providers side and then communicate to Europeana the datasets which are ready.

At the end of September-beginning of October, the first batch of content will be communicated to Europeana. The communication will be carried out after the Plenary on the 29th September in Paris, in order to fix eventual issues that might come out during the meeting.

Another important aspect of the feedback procedure is the validation report that Europeana sends once the datasets of the providers have been harvested. In fact, Europeana harvests the datasets published in the project OAI-PMH server and starts a process of quality control. After circa fifteen days Europeana sends to WP3 its validation report and WP3 informs the Content Providers and WP4 about the possible quality feedback and remedy actions to be performed on the datasets before the final publication in the portal. In this case, the procedure has to be repeated in order to adjust/fix the content. If no adjustments are needed, the dataset is finally published on the Europeana portal.

This feedback is also important to calculate the effective amount of data published in the Europeana portal after the ingestion workflow.

4.4. METADATA ENRICHMENT

As reported in the DoW, the EAGLE repository will be populated by metadata records following the EAGLE schema through the aggregation service. All the content-related activities aim to check and edit the aggregated metadata records through the web based editor developed in WP4, thus preparing them for external applications (e.g. the two Flagship Applications developed in WP5) to get and process them. Content providers will be able to browse their contributed items, filter them based on specific elements and values and group edit the resulted item set in order to perform:

- data cleansing (correct typographical errors, conform with typographical conventions etc.);
- data reconciliation (align elements with authority files, vocabularies or thesauri, link resources to external data sources and semantic binding);
- manual enrichment (link and explore external data sources to extract and reuse information about resources, i.e. locations, time-spans, agents, associated with provider's items;
- contextualisation (link the epigraphs to monuments, sites, ancient topography and any available related information in order to see them in their broader historical and cultural context and to reconstruct the history of collections and the transfer of inscriptions from one place to another).

Particularly, the metadata enrichment and contextualization is operated through the following steps:

- a check of the content through the D-Net metadata record inspector (see Section 4.3.4)[11];
- duplicates identification
- enrichment through the use of the EAGLE controlled vocabularies developed in WP2
- enrichment through the translations developed in WP2.

---

[11] The EAGLE infrastructure, beyond being the aggregator of the providers content and the WP3 working tool for the ingestion monitoring, through the metadata record inspector allows the CPs to check the correctness of their data and to further edit it.

4.4.1 Duplicates identification

One of the methodology to enrich the metadata that will provided to the EAGLE project is the duplicates identification through the Trismegistos platform. By standardizing publication references, collection information, material, provenance and date, the overlap of the various databases is mapped and doubles are identified. Afterwards a unique numeric identifier, the Trismegistos number, is assigned to each document and spread across the partner databases.

The task (T3.5) started at month 1 (April 2013) and officially concluded at month 12 (March 2014). At present, most of the work has been completed and only some minor refinement is still ongoing.

It has to be noted that, out of the total records, only the text on the basis of literature could be disambiguated. Total numbers are therefore different from the total amount of items provided, where the Content Provider gives both texts and artefacts without text (therefore not subject to disambiguation within the EAGLE project).

Duplicates identification/disambiguation is ready for the records provided by UNIROMA 1 (EDR), UNIBA (EDB), UHEI (EDH) and UAH (HispEpOl) since the beginning of the project: the correspondences of the projects' identifiers with TM numbers have been provided.

The only minor exception are a few EDH records from Italy which hitherto have not yet been matched with EDR records. These are high on the priority list. Also, for HispEpOl it has to be checked if what was sent to the Trismegistos platform was the complete dataset or if anything is missing. Some fine-tuning remains to be done in identifying doubles and one-to-many / many-to-one correspondences in the datasets. Many of these have been already tackled, but the final implementation remains to be done. For this step, however, it is preferable to wait until the (quasi-) exhaustivity per region is approached. This is already achieved for the continent Africa (to distinguish it from the Roman province), and it is being achieved for the Iberian peninsula. Afterwards the eastern provinces will follow, then Greece and the Balkan, Italy and Rome, and finally Western Europe (Britain is instead already completely covered)[12].

The fine-tuning with the one-to-many / many-to-one correspondences does not impede the use of the TM identifiers to pull items from the different databases together (it is usually a 1% matter anyway). For the 'minor' databases, such as Ubi erat Lupa, Last Statues of Antiquity and the other EAGLE databases, it will improve the existing publication lookup so that these projects can look up TM numbers themselves, although a first match on the basis of e.g. CIL or AE numbers can be provided for them too.

Partners aligning to one major content provider are not accounted for (UBB, ELTE, ZRC SAZU, UNIPU) as their items are affected by the work done on EDH and EDR. (For all: final one-to-many / many-to-one identification still need to be fine-tuned).

Table V presents the overview of the current status of the activity and its results.

Short name	Documental manifestations to be disambiguated	Documental manifestations with TM number	Comments
UNIROMA1	50.300	50.300	disambiguation completed
UNIBA	26.195	26.195	disambiguation completed

---

[12] This work will be concluded in Spring 2015.

UHEI	67.928	64.510	a few records from Italy still missing: high on priority list
UOXF	so far 1.707	666	disambiguation ongoing via correspondence with EDR
UAH	so far 23.579	23.579	disambiguation completed, to be checked if anything is missing
PLUS	so far 13.351	4.062	disambiguation ongoing via correspondence with EDH
UBB	already included in EDH disambiguation		
ELTE	already included in EDH disambiguation		
UNIPU	already included in EDR disambiguation		
ZRC SAZU	already included in EDR disambiguation		
AUSONIUS	so far 1.258	0	correspondences still to be checked
DAI	no documental manifestations to be disambiguated		
CYI	no documental manifestations to be disambiguated		
BSR	1.018	1.022	disambiguation completed via correspondence with EDH

Table V. The results of the duplicates identification activity

4.4.2 Vocabularies

Metadata enrichment is also performed through the use of the controlled vocabularies that have been developed by the WG1 in WP2.

The Working Group supported a best practice methodology which further stakeholder could use for the alignment of medium sized heterogeneous terminologies. The tool chosen for the development of the vocabularies is TemaTres to allow easy addition of terms, multiple export possible and continuous enlargement without ambiguities.

The EAGLE BPN followed the same approach that the Pleiades and Pelagios project choose for geographical data and adapt it to seven other possible ways of linking data for the international community.

The EAGLE Vocabularies are the following and are maintained to always contain the maximum information available from the current content providers:

- Type of Inscription: http://www.eagle-network.eu/voc/typeins/

- Object Type: http://www.eagle-network.eu/voc/objtyp/

- Material: http://www.eagle-network.eu/voc/material/

- Execution: http://www.eagle-network.eu/voc/writing/

- Decoration: http://www.eagle-network.eu/voc/decor/

- State of Preservation: http://www.eagle-network.eu/voc/statepreserv/

- Dating Criteria: http://www.eagle-network.eu/voc/dates/

The alignment of the vocabularies was carried out using simple Google spreadsheets following on a previous international effort by the EpiDoc Community agreed by WG1 members. This allowed to keep always track of relevant modification (e.g. new entry and related IDs), and to open up in an easy and familiar way to any interested parties the contribution to the alignment. New partners can align their vocabularies in this tables before the actual publication in the EAGLE TemaTres Vocabularies takes place, or directly add terms or links in TemaTres. Content enrichment has been more conveniently done with the use of the aligned vocabulary only in TemaTres.

Among the advantages of TemaTres there is the possibility to expand the search and perform searches of the term being viewed in the main Google search engines.

Exports from the TemaTres vocabularies have been enriched with missing information also following advice from Europeana, and a set of xslt trasnformations has been prepared to enrich the data in EpiDoc (delivery format for inscriptions in EAGLE) and EAGLE DATA MODEL[13].

This stylesheets allow the matching of contents in the skos vocabulary with the string content of the data provided.

For example the element <term> in EpiDoc or <inscriptionType> in the EAGLE data model

might contain:

<term>Akklamation</term>

<term>Acclamatio</term>

The stylesheet will look at the content and match the string using a set of conditions in order to always insert the correct URI once.

The following is the sample of the skos:

<skos:Concept rdf:about="http://www.eagle-network.eu/voc/typeins/lod/73">

    <skos:prefLabel xml:lang="de">Akklamation</skos:prefLabel>

    <skos:altLabel xml:lang="fr">Acclamation</skos:altLabel>

    <skos:altLabel xml:lang="es">Aclamación</skos:altLabel>

    <skos:scopeNote xml:lang="de"> Ausrufe; Anrufungen. Wahlslogans, Glkwsche, Sinnsprhe. Nicht

        beiDefixiooder Anrufungen an Gott (s.Gebet). </skos:scopeNote>

    <skos:historyNote xml:lang="en">Examples: vivas; spes in Deo salvo episcopo Marciano

        HD008112 HD018228</skos:historyNote>

    <skos:inScheme rdf:resource="http://www.eagle-network.eu/voc/typeins/"/>

    <skos:related rdf:resource="http://www.eagle-network.eu/voc/typeins/lod/75"/>

---

[13] All the documentation is available in the Git repository of the project: https://github.com/EAGLE-BPN.

```
<skos:related rdf:resource="http://www.eagle-network.eu/voc/typeins/lod/74"/>

<skos:exactMatch>

   <skos:Concept rdf:about="http://www.eagle-network.eu/voc/typeins/lod/243">

      <skos:prefLabel xml:lang="la">Acclamatio</skos:prefLabel>

   </skos:Concept>

</skos:exactMatch>

<skos:broadMatch>

   <skos:Concept rdf:about="http://nl.dbpedia.org/resource/Acclamatio"/>

</skos:broadMatch>

<skos:exactMatch>

   <skos:Concept rdf:about="http://edh-www.adw.uni-heidelberg.de/edh/inschrift/HD007473"/>

</skos:exactMatch>

<skos:exactMatch>

   <skos:Concept rdf:about="http://edh-www.adw.uni-heidelberg.de/edh/inschrift/HD058290"/>

</skos:exactMatch>

<dct:created>2013-08-01 12:27:53</dct:created>

<dct:modified>2013-08-20 14:17:18</dct:modified>

</skos:Concept>
```

The result of this operation, done twice will be the following:

```
<term ref="http://www.eagle-network.eu/voc/typeins/lod/73">Akklamation</term>

<term ref="http://www.eagle-network.eu/voc/typeins/lod/73">Acclamatio</term>
```

In this way the multilinguality is preserved seamlessly and a search for a term connected to the URI specified will return all items which have that URI in any languages.

For new content providers a process of extraction and enlargement of the vocabularies is carried by UHEI so that all new terms are aligned and new ones are aligned according to the guidelines in Deliverable 2.2.1.

4.4.3 Translations

A further methodology to perform the metadata enrichment is the use of the inscriptions' translations produced within the project (Task 2.2 in WP2). Translations in the Mediawiki are entered from sources outside the consortium and new partners who contribute that kind of content. The Mediawiki, available at http://www.eagle-network.eu/wiki), also allows to enter data:

- from registered independent users

- from users submitting translations via Perseids integration, a editorial vetting system developed in collaboration with Perseus Project

- from batch uploads with pywikipediabot (https://github.com/EAGLE-BPN/eagle-wiki)

All data are available to any users, to the EAGLE portal and to Content providers via a RESTful API native of Wikibase, the extension used for this purpose (http://www.eagle-network.eu/wiki/api.php)

To facilitate the alignment to the data in the EAGLE aggregator each translation has at least one identifier of the content providers and wherever possible a TM identifier.

For what concerns the work about the translations, one important result is connected with the content ingestion. At the moment, it is in fact under analysis the possibility to send this content to Europeana, improving the amount of new records and enriching the quality of the data already committed.

5. REPORT ON THE CONTRIBUTIONS TO EUROPEANA

The content provided by EAGLE matches the current strategy of Europeana of moving from quantity to quality. This is because the national institutions participating in the project, thanks to their authoritative and comprehensive knowledge of the provided materials, are in the best conditions to make EAGLE a trusted source for cultural heritage. The EAGLE Consortium is therefore working hard to reach high quality standards that will be visible in the final results of the Content Providers' efforts.

Aim of the WP3 is coordinating the content provision of the partners to Europeana through the EAGLE infrastructure. For this reason, WP3 operates a continuous monitoring about the data to be provided and checking of any possible delays or problems that may have a negative impact on the content provision.

During the period under report, all the Content Providers worked to meet the deadline and they prepared and exported the content to make it compliant with EAGLE and Europeana specifications. Apart from some Providers that have encountered technical issues and they are working to solve them, the majority of the Content Providers sent the planned amount (25% of what the total amount described in the DoW) through FTP or via OAI-PMH servers to EAGLE; some of them uploaded even more items than what was planned (meeting in some cases the final deadline of the project).

The content provided by the Consortium, for the effect of the transformation into the EAGLE data model, will produce different groups of metadata according to the digital objects they are connected to. So far in the EAGLE infrastructure there are ready 122.512 items related to inscriptions and 73.439 items related to visual representations.

For what concerns the CPs that have provided less content at the moment, the issues are all under control and they will be resolved before the Europeana harvesting, as promised by the involved partners.

The following table presents the current status of the ingestion, comparing what each content provider actually delivered with what it was planned for this first milestone (MS8), i.e. 25% of the total amount foreseen in the DoW. All the issues encountered by the content providers are reported in the last column.

No.	Short Name	Planned amount	Actual amount	Details	Comments
1	**UNIROMA1**	15.000	26.790	14.998 artefacts 11.792 visual repr.	
2	**UNIBA**	11.250	11.956	11.250 artefacts 706 visual repr.	
3	**UHEI**	16.250	88.623	67.928 artefacts 20.695 visual repr.	
4	**UOXF**	375	1.717	1.717 artefacts 0 visual repr.	The CP didn't provide yet the images but will fix the issue soon
5	**UAH**	58.750	302	200 artefacts 102 visual repr.	Delay due to the big effort to enrich the metadata, particularly of the images, and make them compliant with Europeana's guidelines. This issue

					will be solved by the time of the next milestone.
6	**PLUS**	35.805	59.595	22.953 artefacts 36.642 visual repr.	
7	**UBB**	1.000	1.657	901 artefacts 756 visual repr.	
8	**ELTE**	800	663	296 artefacts 367 visual repr.	Few items are missing from this CP
9	**UNIPU**	500	0	0	The 500 items have been sent by the CP to EDR, but for technical problems their ingestion is postponed. The amount will be substituted by 500 items from EDR
10	**ZRC SAZU**	108	0	0	The 108 items have been sent by the CP to EDR, but for technical problems their ingestion is postponed. The amount will be substituted by 108 items from EDR
11	**AUSONIUS**	1.250	2.248	1.251 artefacts 997 visual repr.	
12	**DAI**	250.000	0	0	Delay due to the big effort connected to the export and mapping of the complex Arachne metadata structure, based on CIDOC-CRM, to the EAGLE metadata schema. This issue will be solved by the time of the next milestone.
13	**CYI**	110	0	0	135 items have been sent but the transformation into EAGLE is still ongoing
14	**BSR**	1.016	2.400	1.018 artefacts 1.382 visual repr.	
	TOTAL	392.214	195.952	122.512 artefacts 73.439 visual repr.	

Table VI. Report on the contribution to Europeana for the MS8.

6. CONCLUSIONS

This Deliverable presented the work done by WP3, WP4 and the Content Providers for the implementation and monitoring of the ingestion phase and it outlined the status of the EAGLE contribution to Europeana.

In detail, it described the update of the ingestion workflow and the activities and sources of information elaborated and used to keep under control the ingestion itself.

From the use of internal milestones in order to monitor the content provision to the tools and techniques used to calculate the effective contributions' amount of each CP, WP3 is following on a daily basis the work of all the actors involved.

In this phase of the project, there are no particular problems to be reported but the main focus is on the delay in the achievement of the planned amount of items to be given to Europeana by the end of September 2014. In fact, due to technical issues encountered by some Content Providers, in particular DAI and UAH, more time was needed to finalise the ingestion of this first batch of content to Europeana. Both Content Providers confirmed during the last plenary meeting on the 29th of September in Paris that they have almost overcome their issues and that they will be able to start uploading their content by the end of October 2014.

All the Providers are aware of the deadlines and milestones and they are working to reach the project's targets.

These periodic reports about the contributions to Europeana will periodically updated every 6 months. The next issue is therefore planned at the end of March 2015 (MS12).