Convegni

Studi umanistici – Antichistica

Information Technologies for Epigraphy and Cultural Heritage

Proceedings of the First EAGLE International Conference

Sapienza Università Editrice

Collana Convegni 26

Studi umanistici
Serie Antichistica

Information Technologies for Epigraphy and Cultural Heritage

Proceedings of the First EAGLE International Conference

a cura di
Silvia Orlandi, Raffaella Santucci,
Vittore Casarosa, Pietro Maria Liuzzo

SAPIENZA
UNIVERSITÀ EDITRICE
2014

Sapienza Università Editrice
Piazzale Aldo Moro 5 – 00185 Roma
www.editricesapienza.it
editrice.sapienza@uniroma1.it
Iscrizione Registro Operatori Comunicazione n. 11420
ISBN 978-88-98533-42-8
DOI 10.13133/978-88-98533-42-8

Distribuita su piattaforma digitale da:
Centro interdipartimentale di ricerca e servizi
Settore Publishing Digitale

In copertina: Logo EAGLE *The Europeana network of Ancient Greek and Latin Epigraphy*.
Divieto di riproduzione.

In den Flüssen nördlich der Zukunft / werf ich das Netz aus, das du / zögernd beschwerst / mit von Steinen geschriebenen / Schatten. (Paul Celan)

1. A Concpetual Model for Inscriptions

Harmonizing Digital Epigraphy Data Sources

Vittore Casarosa, Paolo Manghi, Andrea Mannocci, Eydel Rivero Ruiz, Franco Zoppi

Abstract

The TEI/EpiDoc encoding is considered the de facto standard in digital epigraphy as it enables a holistic digital description of an inscription and the semantic mark-up of its text, all of this in a flexible, machine-readable and exchangeable format. At the same time, an EpiDoc document consists of a monolithic, self-descriptive and self-standing information unit which hardly exposes an easy way for cross-linking different documents.

This drawback becomes particularly relevant when dealing with material of heterogeneous nature, collected from heterogeneous sources, as it happens in recent content aggregation projects aiming at the construction of a shared Information Space serving a federated community. This is exactly the case for the project EAGLE (Europeana network of Ancient Greek and Latin Epigraphy), whose main aim is to provide a single user-friendly portal to the inscriptions of the Ancient World, a massive resource for both the curious and the scholarly.

Modern search engines enable users to express a great variety of queries against heterogeneous material and provide a rich functionality for users to browse, explore and interlink the items found. To overcome the heterogeneity of the material collected by EAGLE from over 15 different Content Providers, a unifying conceptual model has been defined, and is presented in this paper.

The proposed conceptual model, which is based on a preliminary identification analysis based on the CIDOC-CRM ontology, consists of a few core entities which can: i) thoroughly express all the different facets of epigraphy-related content such as physical supports, texts, translations, images and other context-related information; ii) serve as the basis for the definition of a common metadata schema; iii) enable users to express sophisticated queries to accurately retrieve the material of interest.

Keywords

Epigraphy, EpiDoc, data model, metadata, portal service, search engine.

1.1. Introduction

The EpiDoc project [BODARD 2009] represents the most incisive innovation in the field of Epigraphy after the Leiden conventions were defined in 1932 [VAN GRONINGEN 1932], and it is considered the de facto standard in digital epigraphy. EpiDoc enables a holistic digital description of an inscription and the semantic mark-up of its text, all of this in a flexible, machine-readable and exchangeable format, satisfying many aspects of the requirements currently set by the Epigraphic community. At the same time, an EpiDoc file consists in a monolithic, self-descriptive and self-standing information unit. To make an analogy, an EpiDoc file is similar to a folder taken out from an archive. It is easy to make use of it as a whole, or to make searches within it, but it is not so easy to connect elements contained in the folder with elements contained in other folders of the archive or, even more difficult, with elements in other archives. Therefore, any "outer jump" leading to external information or evidence is left to the investigator's intuition. Similarly, an epigraphy practitioner navigating in an EpiDoc file faces the same situation. In 2009 Bodard et al. claimed: "We see EpiDoc as a small first step on the road to truly digitally-enabled epigraphy scholarship, in which not only will it be possible more efficiently to answer questions, but in addition it will be possible to ask and answer new types of question, and even to discover new questions to ask." [BODARD 2009].

However, the monolithic nature of EpiDoc becomes particularly evident when dealing with material of heterogeneous nature, collected from many different sources, as it happens in content aggregation projects such as EAGLE,[1] which is a Best Practice Network co-funded by the European Commission, under its Information and Communication Technologies Policy Support Programme. Its main aim is to bring together the most prominent European institutions and archives in the field of Classical Latin and Greek epigraphy, collecting inscriptions coming from 25 EU countries, with more than 1.5 M of images and related metadata, including translations of selected texts for the benefit of the general public. The collected material will serve two purposes: on one side it will be ingested to Europeana, providing a comprehensive collection of unique historical sources which constitute a veritable pillar of European culture (the collected material represent approximately

[1] http://www.eagle-network.eu/

80% of the total amount of inscriptions in the Mediterranean area). On the other side the collected material will be made available on the EAGLE portal to the epigraphers' community and to the general public, for queries and research. Two applications will be integrated into the portal, especially addressing the general public: a mobile application will enable tourists to get information about inscriptions they find in archaeological sites or in museums by sending their images to the EAGLE portal; a storytelling application will allow teachers and experts to assemble epigraphy-based narratives for the benefit of less experienced user.

Present search engines have accustomed users to express queries against heterogeneous material looking for a specific sets of entities, such as web pages, images, files, videos, etc., and enable the user to browse, drill-down and interlink the results found (e.g. search for an image and get all the web pages containing the same image or visually close images). Being able to transpose such functionality to the epigraphic world and being able to relate originally separate concepts and items will provide benefits both to the scholarly research and to the general public, which today are not so easy to achieve with the existing EpiDoc archives.

In this paper we present a conceptual model that starts with a preliminary identification of the main entities of interest in epigraphy, based on the CIDOC-CRM ontology [Doerr 2003]. In a second step a more concise conceptual model is defined, consisting in a few core entities which can adequately support: i) thoroughly expressing all the different facets of epigraphy-related content such as physical supports, texts, translations, images and other context information; ii) serve as the basis for the definition of a common metadata schema; and iii) enabling the user to express sophisticated queries to accurately retrieve the material of interest. This conceptual model is the basis for defining the EAGLE Common Metadata Model, unifying all the different data sets received from the EAGLE Content Providers, and underlying the query and search facilities provided by the EAGLE portal.

The remainder of this paper is organized as follows: in Section 2 we provide some additional considerations about mark-up languages and EpiDoc, in Section 3 we present the preliminary analysis based on the CIDOC-CRM ontology, in Section 4 we present the EAGLE conceptual model, in Section 5 we present its implementation based on the D-NET open source software,[2] and finally, in Section 6 we provide some final considerations.

1.2. EpiDoc and modern epigraphy

At the beginning of last century, the creation of large corpora of (printed) inscriptions made clear to the epigraphist community the importance of having a consistent, agreed upon standard for representing, as faithfully as possible, not only the text (the characters) appearing on an inscription, but also how the text appeared on the original carrier (e.g. abbreviations, erasures, unreadable or missing characters) and possibly also the editor's interpretation. In 1931-32 the so called Leiden Convention was proposed [VAN GRONINGEN 1932], basically specifying how features of an inscription, besides the text itself, should be represented in print, and it quickly became the de-facto standard for representing the transcription of inscriptions.

The arrival of computers and the increased use of digital (digitized) material in the last 20 years have almost completely transformed the large body of inscription printed material (articles, books, corpora), into "textual databases", that should support present-day tools for searching and retrieving information. From this point of view the digitized Leiden notation has severe limitations, as it was conceived to represent graphically (in print) all the features of an inscription (e.g. underlined text to indicate text previously known but presently disappeared, or a dot underneath a letter to indicate a dubious interpretation) and those "graphical features" disappear when the information is stored in a data base, and need to be represented with "special characters" to indicate the beginning and the end of the graphical feature. Other special characters, like the round and square brackets, are already present in the Leiden notation, to convey information of a more conceptual nature, like the (guessed) expansion of abbreviations, or the (guessed) insertion of missing characters. Modern search engines are essentially based on character string matching, and all those "extraneous" characters interspersed in the text of the inscription make it difficult and clumsy the retrieval of the desired information.

The arrival of Internet and the Web has led to the re-discovery and common usage of the so called Mark-up Languages, i.e. the possibility

2 http://www.d-net.research-infrastructures.eu

of defining opening and closing "tags" to be included in the data, providing additional information for the enclosed data. XML (eXtensible Mark-up Language) has become the de-facto standard for the exchange of information between computers, as it has the advantage of having a formal structure that can be understood (processed) by a computer program, and at the same time it can be easily defined (and understood) by a human.

At the end of the nineties, following a round table held in Rome, a manifesto from Prof. Panciera, recommending "the establishment of an online, free and unrestricted database of all surviving Greek and Latin epigraphical texts produced down to the end of Antiquity", prompted the publication of a proposal (EpiDoc, already under study at the University of North Carolina at Chapel Hill) for the encoding of epigraphic material based on TEI. The Text Encoding Initiative (TEI) was a standard already widely used in the Digital Humanities community for XML annotation of manuscripts and old documents [Bodard 2008]. EpiDoc consists of a group of recommendations and tools that provide a way of encoding scholarly and educational editions of inscriptions based on a subset of TEI. A full EpiDoc document may contain, in addition to the text itself, information about the history of the inscription, a description of the text and its support, commentary, findspot and current locations, links to photographs, translations, etc.

Although EpiDoc is a powerful tool for the scholarly annotation of inscriptions, as we have already pointed out before, it does not lend itself to an easy search and navigation through big epigraphic data bases, with hundreds of thousands of inscriptions. For example, given an EpiDoc document that includes pointers to two images of the described inscription, it is not easy to find another EpiDoc document describing the same inscription, but providing pointers to different images, or to retrieve detailed information about those images, that usually are not included in the EpiDoc documents. Unfortunately, modern search engines (a la Google) have accustomed both the general public and the scholars to believe that with few simple queries it is possible to get not only the items of direct interest, but also (most of) the information that in one way or the other is related to them.

1.3. Analysis in CIDOC-CRM

Usually, epigraphic databases use different information models to best fulfil their needs and purposes. In the frame of the EAGLE project, in order to try and provide a unified view, aimed at improving interoperability and exchange, a conceptual model of epigraphic entities was developed, based on CIDOC-CRM, to help describe in full detail common concepts and their relationships. CIDOC Conceptual Reference Model (CRM) provides a formal ontological model for describing the structure of Cultural Heritage objects and the relation between them, including event models for representing the life cycle of the objects.

1.3.1. High-Level view of the model

Epigraphic objects are represented as instances of the CIDOC concept E84 Information Carrier (in the following, CIDOC concepts and properties are in italics), which is a particular case of man-made object and provides the way for describing all the physical characteristics of monuments such as dimensions, materials, state of preservation and also for distinguishing the objects by names or any other identifiers.

If the monument bears an inscription of some kind, this can be represented through the use of an E34_Inscription object, which is related to its information carrier using property P128_carries.

Any other information related to the inscription (but not to the material object), such as transcription text, translation text, bibliography, critical apparatus, commentary and the different type of surrogates is represented by instances of E31_Document, related to the inscription by the property P70_Documents.

Fig. 1.1 summarizes this high level view of the conceptual model.

1.3.2. Low-Level expansion of the model

The many different concepts underlying the high-level model are divided in two different subgroups:

- Sub-model for physical characteristics of monuments and inscription, location, dating.
- Sub-model for documental information.

The second sub-model corresponds to the description of the characteristics of the textual information carried by the Physical Object.

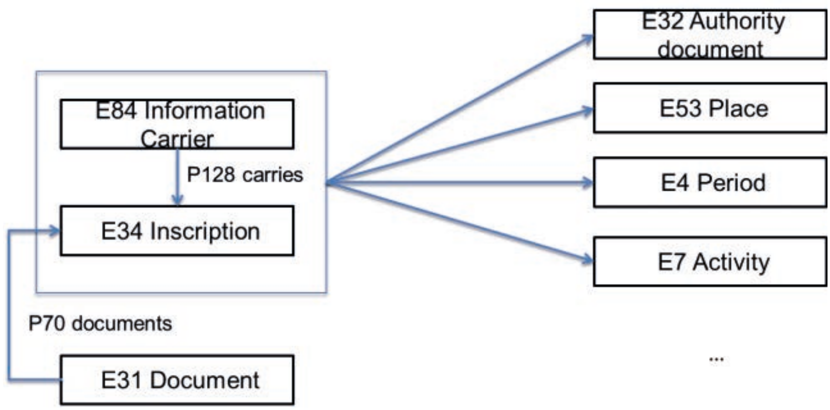

Fig. 1.1. Core Model

1.3.2.1. Sub-model for physical characteristics

Monuments are identified using E42_Identifier related via property P1_is_identified_by. In EAGLE, identifiers are usually the local identifiers assigned by the institution "owning" the object, qualified by the name of the institution, to avoid naming conflicts. Materials, dimensions, types and status of conservation are expressed by the concepts E57_Material, E54_Dimension, E55_Type and E3_Condition State respectively.

The life cycle of objects, including creation, finding and curation activities is represented using E4_Period, E7_Activity and its sub-concepts. Location of these events can be specified, when possible, using instances of E53_Place. Fig. 1.2 summarizes this sub-model.

1.3.2.2. Sub-model for Documental Information

CRM entity E31_Document is the way by which the CIDOC model allows the representation of non-physical elements that describe reality. It may be related with any CRM entity via property P70 documents and comprises several forms of expressing those descriptions about reality, such as texts, images, graphics, videos, including the special case of documentation databases. E31_Document is used in EAGLE for representing all the information not related to the physical characteristics of the information carrier object, which may include transcription text, translation text, images and graphics, bibliography, critical apparatus and commentary. Fig. 1.3 summarizes this sub-model.

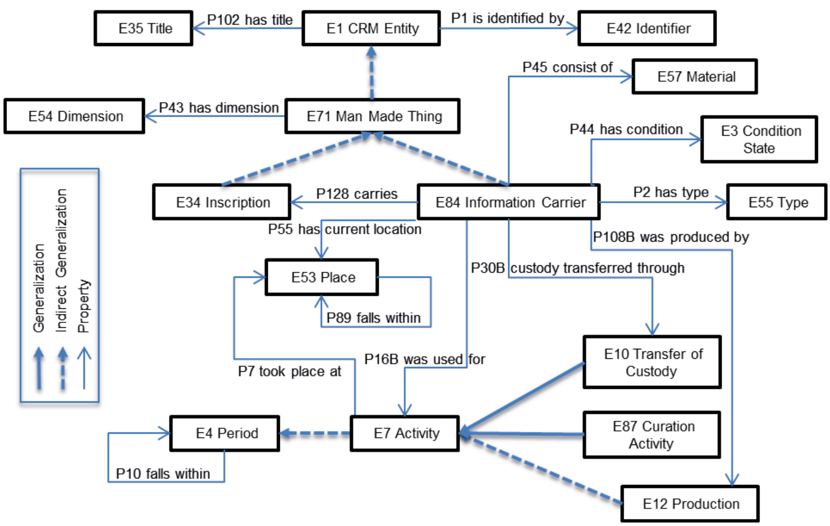

Fig. 1.2. Sub-model for physical characteristics, location, dating

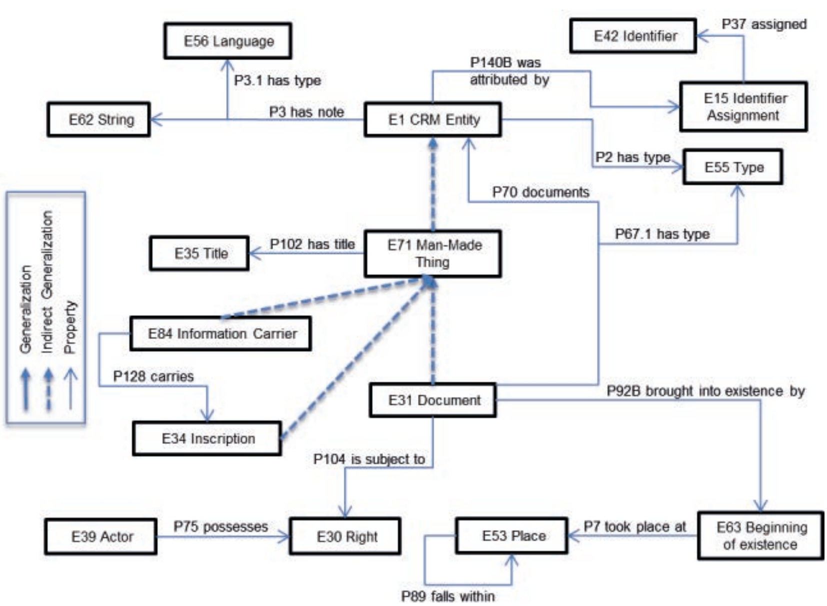

Fig. 1.3. Sub-model for documental information

1.4. The Aggregator Conceptual Model

1.4.1. Harmonizing Digital Epigraphy Data Sources

Since an EpiDoc file packs together several related (but distinct from an ontological perspective) entities, we have provided a preliminary analysis of the concepts and descriptive units conveyed through EpiDoc making use of the ontology defined by CIDOC-CRM [Doerr 2003].

Then, based on this analysis, we have derived a simple conceptual model able to map several inter-related entities which enables fine grained queries and exploration portal-wise.

1.4.2. Conceptual Model for the aggregation of heterogeneous epigraphic content

Very simply, the Epigraphy Aggregation Conceptual Model (EACM) consists of a root entity (the Main Object) from which four sub-entities can be derived, each one capturing some of the properties that have been identified thanks to the EpiDoc and CIDOC modelling efforts. The defined sub-entities are the following: *i)* Artefact, *ii)* Inscription, *iii)* Visual manifestation, *iv)* Documental manifestation.

All the information to be aggregated in EAGLE will find its place into one or multiple instances of the sub-entities mentioned above. In the following we briefly describe each entity, and provide a list of its main properties. It has to be noted that all of the concepts and properties here described have already been defined both in EpiDoc and in the CIDOC based conceptual model.

Fig. 1.4 shows a high-level view of the model, where solid lines represents a hierarchical relation between two entities, e.g. an Artefact (or any blue box) is a Main object, or a Translation/Transcription is a Documental manifestation. A dashed line instead represents a relationship between two entities where applied cardinalities can be expressed at the two ends of the line.

1.4.2.1. Main Object

The top entity in the conceptual model (the Main Object) is an abstract entity that in practice will be materialized into one or more instances of some or all of the sub-entities underlying it. These sub-entities share several common properties, namely a unique object identifier, source information, metadata editing/authoring information and intellectual property rights (IPR) of metadata, title and description.

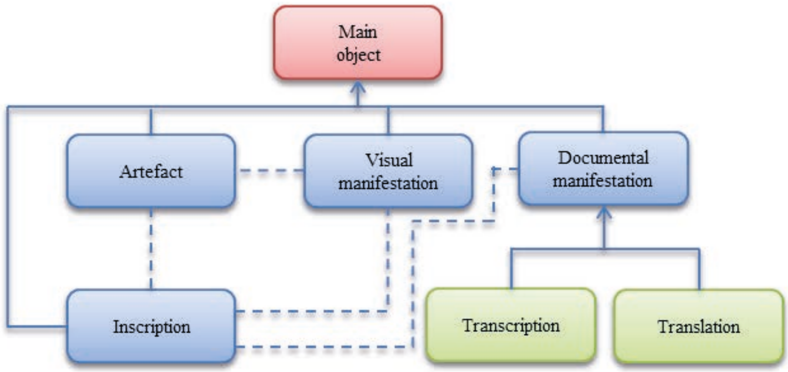

Fig. 1.4. Epigraphy Aggregation Conceptual Model

1.4.2.2. Artefact

The Artefact entity captures the physical nature of an object of study in the Epigraphic domain. From the analysis of properties identified in EpiDoc, we isolated a set of relevant properties such as the artefact type (i.e. the kind of monument), its material and dimensions, its decoration, the status of preservation and place of conservation, and relevant findspot and dating information.

In accordance with the CIDOC model, an Artefact can be related to one or more Visual manifestation (instances of images related to the artefact) and can be related to zero or more Inscription (instances of text regions possibly present on its surface).

1.4.2.3. Inscription

The Inscription entity is a collection of structured properties describing the textual and semantic nature of a text region possibly present on an artefact.

Common properties describing this concept consist in the inscription type, its engraving technique, the metric of the text, the sizes of the field and letters, its palaeographic characteristics, the author of the inscription, the honorand, the dedicator and other person names possibly cited, or referral to events.

Since an Inscription has both visual and textual characteristics, an Inscription instance can be put in relation with zero or one Artefact instance, zero or more Visual Manifestation instances, and zero or more

Documental manifestation instances (either Transcriptions or Translations or both).

1.4.2.4. Documental manifestation

The Documental manifestation entity contains all information related to the "textual nature" of an inscription. It has two sub-entities: the Transcription, which gathers the information related to the text of the inscription itself, and the Translation, which gathers the relevant properties of possible translations in modern languages of the (interpreted) text of the inscription.

Transcription

The Transcription entity describes the inscription text in its original (ancient) language. A Transcription is related to one and only one Inscription instance. Some of the main properties of a Transcription entity are (obviously) the transcription text, its critical apparatus, side commentaries and referenced bibliography.

Translation

The Translation entity captures all the aspects relevant to the translation of an ancient text. Some of the properties regarding a translation are the (modern) language used, the author(s) contributing to the translation, the text of the translation and possible annotations, its intellectual property right statement (IPR) and publishing information if available.

A Translation instance can be put in relation with one or more Inscription instances, of which it is a translation.

1.4.2.5. Visual manifestation

The Visual manifestation entity collects all the information related to the "visual nature" of a generic artefact, be it a stone, a monument, or an epigraphy-related object providing context to others epigraphic objects of interest. Visual manifestation instances can contain either born-digital material such as pictures from a digital camera, computer graphics, or digitized printed material, such as drawings, pictures, literature (e.g. CIL pages), or also digitized videos. .

The main properties of a Visual manifestation are the digital file, its location (URL, thumbnail), its general properties (e.g. dimensions, format, resolution, quality, etc.), and authoring information (e.g. author, date and copyright statement).

A Visual manifestation can be put in relation with zero or one Artefact instances (of which it is a picture) and zero or more Inscription instances.

1.5. The EAGLE implementation

1.5.1. Implementing Data Infrastructures

The increased usage of digital archives, which has taken place over the last twenty years in several communities, has stimulated the need for integrating and aggregating content from several different archives to make it available through a single access point. This tendency can be seen in several national initiatives (e.g., BASE,[3] DAREnet,[4] OAIster,[5]) and European projects (e.g., Europeana,[6] Bricks,[7] ScholNet,[8] DILIGENT,[9] D4Science,[10] DRIVER,[11] OpenAIRE,[12] CLARIN,[13] EFG,[14] HOPE).[15] In the last three Framework Programme calls, the European Union initiated the so called knowledge infrastructure vision, inspired by the same goal of unifying data resources of all kinds available in Europe. The idea was that of devising data infrastructures, which are environments through which several organizations can share, process, aggregate their data resources by adopting an economy of scale approach. Several technological solutions [MANGHI, MIKULICIC, CANDELA, ARTINI, et al. 2010] were devised in such projects, to offer functionality for collecting data from heterogeneous data sources (e.g. repository systems, archives, databases), curating such data to form a homogeneous Information Space, and offering customized portal services to operate over such space; e.g. search, inference of references between publications, citation calculation, etc.

---

3 http://www.base-search.net

4 http://www.darenet.nl/

5 http://www.oaister.org

6 http://www.europeana.eu

7 http://www.brickscommunity.org/

8 ftp://ftp.cordis.europa.eu/pub/ist/docs/rn/scholnet.pdf

9 http://diligent.e3rcim.eu/

10 http://www.d4science.eu/

11 http://www.driver-community.eu/

12 http://www.openaire.eu/

13 http://www.clarin.eu/

14 http://www.europeanfilmgateway.eu/

15 http://www.peoplesheritage.eu

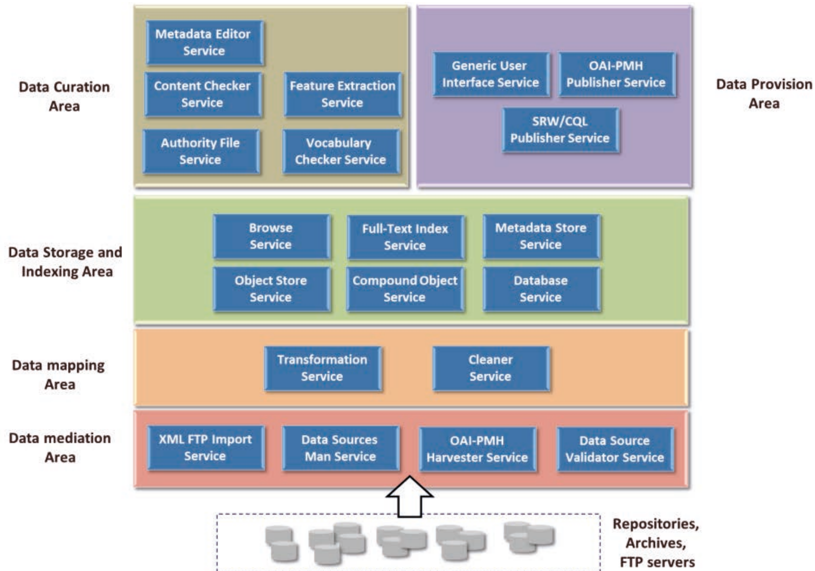

Fig. 1.5. D-NET service architecture

The D-NET software toolkit, resulting from the experience of ISTI-CNR through its participation in the OpenAIRE,[16] DRIVER, DRIVER-II,[17] and European projects, is an open source solution specifically devised for the construction and operation of customized data infrastructures. D-NET provides a service-oriented framework where data infrastructures can be constructed in a LEGO-like approach, by selecting and properly combining the required D-NET services [Manghi, Mikulicic, Candela, Castelli, et al. 2010]. The resulting infrastructures are customizable (e.g., transformation into common metadata formats can be configured to match community preferences), extensible (e.g. new services can be integrated, to offer functionality not yet supported by D-NET), and scalable (e.g., storage and index replicas can be maintained and deployed on remote nodes to tackle multiple concurrent accesses or very-large data size). D-NET[18] offers a rich set of services [Fig. 1.5] targeting aspects such as data collection (mediation area), data mappings from formats to formats (mapping area), and

16 http://www.openaire.eu/

17 http://www.driver-community.eu/

18 http://www.d-net.research-infrastructures.eu

data access (provision area). Services can be customized and combined to meet the data workflow requirements of a target user community. As proven by the several installations and adoption in a number of European projects (DRIVER,[19] OpenAIRE,[20] HOPE,[21] EFG[22]), D-NET represents an optimal and sustainable solution [MANGHI, BARDI, et al. 2014; MANGHI, ARTINI, et al. 2014] for the realization of the EAGLE infrastructure.

1.5.2. The EAGLE Infrastructure

The EAGLE infrastructure consists of the D-NET services shown in Fig. 1.5, appropriately combined to support the data ingestion workflow defined for the epigraphists' community. In particular, the services in the Data Mapping, Curation and Provision areas result from the project design activities. They were devised in order to meet the requirements of the EAGLE Content Providers, but engineered to support their functionalities when operating over arbitrary XML schemas.

1.5.2.1. Metadata Mapping Definition, Transformation, and Cleaning

Archives and their experts joining the EAGLE infrastructure are supported with a methodology that facilitates the definition of structural mappings from their archive schema onto the EAGLE common metadata schema and semantic mappings from their vocabularies onto the common vocabularies. A mapping consists in a set of rules, which serve as input to the infrastructure administrators to configure the services in the Data Mapping Area. Here, the Transformator Service and the Cleaner Service run programs which parse, validate and transform the source records into EAGLE records according to the defined rules.

The Transformator Service is responsible for the application of structural rules. Such rules define the correspondence among elements and attributes of the archive schema and elements and attributes of the EAGLE schema. Structural mapping is not as trivial as it may seem, due to the fact that input XML records are typically mapped onto several

[19] http://www.driver-community.eu/

[20] http://www.openaire.eu/

[21] http://www.peoplesheritage.eu

[22] http://www.europeanfilmgateway.eu/

interrelated EAGLE records, representing different EAGLE data model entities.

The Cleaner Service is responsible for the application of semantic rules. Such rules identify an element of the archive schema and the corresponding element of the EAGLE schema (i.e., source element and target element of structural rules), and define the correspondence between the terms of the respective vocabularies.

1.5.2.2. Metadata quality control and enrichment

The D-NET software toolkit customisation for the EAGLE infrastructure includes the following services, constituting the D-NET Data Curation Area.

Content Checker. The Content Checker is a validation tool that allows low-level searching and browsing the pre-production Information Space in order to check if metadata records have been correctly harvested and mapped.

Vocabulary Checker. The Vocabulary Checker gives access to the metadata records that do not satisfy the constraints imposed by the common metadata schema and vocabularies after the transformation and cleaning phases. The Vocabulary Checker displays the number, the types and the positions of errors in the records of the Information Space. Thanks to the browse by error typology functionality, curators can decide if an error can be solved directly in the Information Space via the Metadata Editor Tool or in the original source archive.

Metadata Editor Tool. The Metadata Editor Tool (MET) is a cataloguing tool for the enrichment of the Information Space. It allows data curators to add, edit and delete metadata records in the Information Space, as well as to establish relationships between existing (authority) records, even if coming from different sources. The MET is aware of controlled vocabularies, hence supports data curators while editing controlled elements by proposing a drop down list with all and only the terms defined by the associated controlled vocabulary.

1.5.2.3. Metadata Publishing

The EAGLE implementation will be shortly accessible via a dedicated portal. Facilities like advanced metadata search and browse (by collection, provider, etc.) and search results filtering enhance the user experience in the phases of search and access. Moreover, D-NET offers services to export metadata records through OAI-PMH, OAI-ORE, and SRW/CQL protocols. EAGLE will operate such services to automatically serve its information space to third-party consumers, above all the Europeana project[23], of which EAGLE is a direct feeder.

1.6. Conclusions

We have described the solution adopted in the EAGLE Best Practice Network to achieve a complete integration of different Ancient Greek and Latin Epigraphy archives, representing nearly the 80% of the existing assets in the area.

The solution is based on the creation of a conceptual model that has, at the same time, the power to preserve the metadata quality of the different archives' schema and the simplicity to enable simple mappings from all different archives. Metadata aggregation is based on the use of the D-NET open source software toolkit, a data infrastructure enabling software.

The need and rationale for moving from the TEI/EpiDoc encoding - the de facto standard in digital epigraphy - to the proposed conceptual model has been introduced and discussed.

[23] http://www.europeana.eu

Bibliography

Bodard, G. (2008). "The Inscriptions of Aphrodisias as Electronic Publication: a user's perspective and a proposed paradigm". In: *Digital Medievalist* 4. url: http://www.digitalmedievalist.org/journal/4/bodard/ (cit. on p. 27).

— (2009). "EpiDoc: Epigraphic documents in XML for publication and interchange". In: *Latin on Stone: Epigraphic Research and Electronic Archives*. Ed. by F. Feraudi-Gruenais, pp. 1–17 (cit. on p. 24).

Doerr, M. (2003). "The CIDOC conceptual reference module: an ontological approach to semantic interoperability of metadata". In: *AI magazine* 24.3, p. 75. url: http://www.aaai.org/ojs/index.php/aimagazine/article/viewArticle/1720 (visited on 07/29/2014) (cit. on pp. 25, 31).

Manghi, P., M. Artini, C. Atzori, A. Bardi, A. Mannocci, S. La Bruzzo, L. Candela, D. Castelli, and P. Pagano (2014). "The D-NET Software Toolkit: A Framework for the Realization, Maintenance, and Operation of Aggregative Infrastructures". In: *Program* 48 (4). url: http://www.emeraldinsight.com/doi/abs/10.1108/PROG-08-2013-0045 (cit. on p. 36).

Manghi, P., A. Bardi, and F. Zoppi (2014). "Coping with interoperability and sustainability in cultural heritage aggregative data infrastructures". In: *International Journal of Metadata, Semantics and Ontologies* 9 (2), pp. 138–154 (cit. on p. 36).

Manghi, P., M. Mikulicic, L. Candela, M. Artini, and A. Bardi (2010). "General-Purpose Digital Library Content Laboratory Systems". In: *Research and Advanced Technology for Digital Libraries*. Ed. by M. Lalmas, J. Jose, A. Rauber, F. Sebastiani, and I. Frommholz. Lecture Notes in Computer Science 6273. Springer Berlin Heidelberg, pp. 14–21. isbn: 9783642154638. url: http://link.springer.com/chapter/10.1007/978-3-642-15464-5%5C_3 (visited on 07/29/2014) (cit. on p. 34).

Manghi, P., M. Mikulicic, L. Candela, D. Castelli, and P. Pagano (2010). "Realizing and maintaining aggregative digital library systems: D-net software toolkit and oaister system". In: *D-Lib Magazine* 16.3, p. 1. url: http://dialnet.unirioja.es/servlet/articulo?codigo=3742742 (visited on 07/29/2014) (cit. on p. 35).

Van Groningen, B. A. (1932). "Projet d'unification des systèmes de signes critiques". In: *Chronique d'Egypte* 7.13, pp. 262–269. url: http://brepols.metapress.com/index/92w111270m365748.pdf (visited on 07/29/2014) (cit. on pp. 24, 26).

Acknowledgments

This work is partly funded by the EU EAGLE Best Practice Network project: Grant Agreement CIP 325122, call CIP-ICT-PSP-2012-6.