

FORMA VRBIS



PROGETTO EAGLE. Le frontiere digitali dell'epigrafia
SPECIALE SIBARI. Nuova luce sulla città del VI secolo a.C.

ROMA



In copertina: Particolare base di statua presso la Biblioteca di Storia Moderna e Contemporanea MiBACT. Foto co-vincitrice del Premio speciale EAGLE per Wiki Loves Monuments 2013 (autore G. Anesin Fonte Wikimedia Commons CC-BY-SA)

FORMA VRBIS. Itinerari nascosti di Roma antica
Mensile Tecnico-Scientifico fondato da Luciano Pasquali

Pubblicazione registrata presso il Tribunale di Roma n°548/95 del 13/11/95

Direttore responsabile
Silvia Pasquali

Direttore scientifico
Claudio Mocchegiani Carpano

Direttore editoriale e curatore scientifico
Simona Sanchirico

Comitato scientifico d'onore

Silvia Aglietti DAI - Istituto Archeologico Germanico di Roma, Fondazione Dià Cultura; Giovanna Alvino già Soprintendenza Archeologia del Lazio e dell'Etruria Meridionale; Luca Attenni Museo Civico Lanuvino, Museo Civico di Alatri; Giovanni Attili "Sapienza" - Università di Roma, Fondazione Dià Cultura; Wouter Bracke Accademia Belgica; Elena Calandra Soprintendenza Archeologia dell'Umbria; Gianfranco De Rossi Espera Srl; Paola Di Manzano Soprintendenza Archeologica di Roma; Giuseppina Ghini Soprintendenza Archeologia del Lazio e dell'Etruria Meridionale; Dario Giorgetti Università degli Studi di Bologna; Michel Gras Accademia dei Lincei; Emanuele Greco Saia - Scuola Archeologica Italiana di Atene; Leonardo Guarnieri CoopCulture; Pier Giovanni Guzzo Istituto Nazionale di Archeologia e Storia dell'Arte; Claudio Honorati Fondazione Dià Cultura; Ettore Janulardo Università degli Studi di Bologna; Bruno La Corte già Gruppo Tutela Patrimonio Archeologico del Nucleo Polizia Tributaria di Roma della Guardia di Finanza; Eugenio La Rocca "Sapienza" - Università di Roma; Enzo Lipopolis "Sapienza" - Università di Roma; Daniele Manacorda Università degli Studi di Roma Tre; Raffaele Mancino Reparto Operativo del Comando dei Carabinieri Tutela Patrimonio Culturale; Federico Marazzi Università degli Studi "Suor Orsola Benincasa", Napoli; Paolo Moreno Università degli Studi di Roma Tre; Davide Nadali "Sapienza" - Università di Roma; Valentino Nizzo Soprintendenza Archeologia dell'Emilia Romagna, Fondazione Dià Cultura; Carlo Pavia già Direttore di Forma Urbis; Francesco Pignataro Fondazione Dià Cultura; Massimiliano Quagliarella Sezione Archeologia del Reparto Operativo del Comando dei Carabinieri Tutela Patrimonio Culturale; Silvana Rizzo Ministero dei Beni e delle Attività Culturali e del Turismo; Massimo Rossi II Sezione del Gruppo Tutela Patrimonio Archeologico del Nucleo Polizia Tributaria di Roma della Guardia di Finanza; Marco Santucci Università degli Studi di Urbino, Fondazione Dià Cultura; Vincenzo Scarano Ussani Università degli Studi di Ferrara; Giovanni Scichilone Loyola University of Chicago; Patrizia Serafin Petrillo II Università degli Studi di Roma Tor Vergata; Elizabeth J. Shepherd Istituto Centrale per il Catalogo e la Documentazione-Aerofototeca Nazionale; Christopher Smith British School at Rome; Mario Torelli Accademia dei Lincei; Catherine Virlouvet École française de Rome; Giuliano Volpe Università di Foggia

Editore

Laura Pasquali, amministratore unico della casa editrice
E.S.S. - Via di Torre S. Anastasia, 61 - 00134 Roma, tel. 06 710561

Linea editoriale, progetto scientifico e veste grafica

Fondazione Dià Cultura, www.diacultura.org; info@diacultura.org; via della Magliana 65 E/H, 00166 Roma, tel. 06 66990385; fax 06 66990422; Square Charles Maurice Wiser 19, 1040 Bruxelles; tel. +32 02 2304593; fax +32 02 2304593

Redazione

Chiara Leporati, Laura Pasquali, Francesco Pignataro, Simona Sanchirico

Impaginazione e grafica

Giancarlo Giovine per la Fondazione Dià Cultura

Documentazione fotografica

A cura degli Autori

Referenze fotografiche

Foto d'archivio privato e di Enti pubblici e privati

Pubblicità e diffusione

Laura Pasquali

Amministrazione e segreteria

E.S.S. - Via di Torre S. Anastasia, 61 - 00134 Roma, tel. 06 710561 - Fax 06 71056230

Abbonamenti: L'abbonamento partirà dal primo numero raggiungibile eccetto diversa indicazione

Italia: annuale 41,30 euro

Estero: annuale 77,50 euro

Arretrati: i numeri arretrati possono essere ordinati (previo riscontro della disponibilità via email, scrivendo a office@sysgraph.com) mediante versamento anticipato sul c.c. 58526005, intestato a ESS Srl Via di T.S. Anastasia, 61 - 00134 Roma, per un importo di 5,50 euro a copia; nella causale indicare la pubblicazione e il numero/anno desiderato. Le richieste saranno evase sino a esaurimento delle copie

Stampa

System Graphic Srl via di Torre Santa Anastasia 61, 00134 Roma - Telefono 06 710561

Distributore per l'Italia

Press-di Distribuzione Stampa e Multimedia Srl - 20090 Segrate (MI)

Nessuna parte della presente pubblicazione può essere riprodotta in alcun modo senza il consenso scritto dell'Editore

Finito di stampare nel mese di Gennaio 2016 © Copyright E.S.S. Editorial Service System



Questo periodico è associato all'Unione Stampa Periodica Italiana

Sommario

Editoriale: le frontiere digitali dell'epigrafia

di Simona Sanchirico

1

EAGLE: dalla nascita di un'idea alla realizzazione di un progetto

di Silvia Orlandi

4

EAGLE nella rete

di Raffaella Santucci

8

Il portale EAGLE

di Claudio Prandoni, Antonella Fresa, Tania Masi

12

EAGLE – L'infrastruttura di aggregazione dei dati e i servizi a supporto del portale e delle applicazioni

di Andrea Mannocci, Vittore Casarosa, Paolo Manghi, Franco Zoppi

18

Sistema di riconoscimento delle immagini e mobile app

di Giuseppe Amato, Paolo Bolettieri, Fabrizio Falchi, Lucia Vadicamo

22

SPECIALE SIBARI

di Pier Giovanni Guzzo, Alessandro D'Alessio, Simone Marino, Adolfo Tosti

26

La Flagship Storytelling Application di EAGLE

di Francesco Mambrini

36

EAGLE nella rete

di Raffaella Santucci

40

EAGLE e Wikimedia Italia

di Cristian Consanni, Pietro Liuzzo, Lorenzo Losa, Luca Martinelli, Andrea Zanni

44

Progetto e realizzazione del video promozionale per EAGLE

di Luca Giberti

48

Esperienze didattiche con Eagle

di Lucio Benedetti, Enrico Zuddas, Lavinio Del Monaco

52



1. CNR – Centro di calcolo

EAGLE – L'infrastruttura di aggregazione dei dati e i servizi a supporto del portale e delle applicazioni

di Andrea Mannocci*, Vittore Casarosa*, Paolo Manghi*, Franco Zoppi*

Introduzione

La lunga tradizione epigrafica, risalente al secolo XVI, epoca ben lontana dai concetti di globalizzazione, standardizzazione e interoperabilità, ha fatto sì che nel tempo si sedimentassero, nelle varie comunità di studiosi, *modus operandi* spesso contrastanti. All'inizio degli anni '30 il Sistema di Leida (B. A. VAN GRONINGEN, "Projet d'unification des systèmes de signes critiques", in *Chronique d'Égypte* 7, 1932, pp. 262-269) ha contribuito a ridurre notevolmente la frammentazione presente nei testi, ma una nuova deriva si è verificata dagli anni 90 in poi quando, con l'arrivo di Internet e del Web, gli archivi epigrafici hanno iniziato la loro conversione al digitale. Nonostante la definizione di uno standard per l'annotazione di documenti a carattere epigrafico (EpiDoc: <http://sourceforge.net/p/epidoc/wiki/Home/>) le comunità hanno per lo più operato in modo indipendente e senza nessuna linea guida condivisa, lasciando di fatto il panorama altamente frammentario.

Il progetto EAGLE mira proprio a riconciliare e riunificare sotto un'unica egida le varie comunità epigrafiche e rendere i loro contenuti ricercabili da un unico punto di accesso, e a questo scopo ha sviluppato un'infrastruttura che consente l'aggregazione di tali contenuti e la loro armonizzazione secondo un modello di dati condiviso, e permette infine di interrogare i dati sia attraverso il proprio portale che attraverso Europeana.

Panoramica del problema e della soluzione adottata

Il problema principale che il progetto EAGLE ha dovuto affrontare è stata l'alta eterogeneità degli archivi epigrafici partecipanti al progetto, in quanto i loro contenuti sono generalmente basati su modelli e schemi di metadati diversi, su vocabolari diversi per indicare gli stessi concetti, su standard diversi per nomi, date e periodi di tempo, ecc. Inoltre gli archivi epigrafici differiscono spesso anche nella tipologia dei metadati esportati. Parte di questi descrive oggetti reali (supporto fisico di un'iscrizione), un'altra parte descrive l'iscrizione stessa, considerata come una stringa di simboli espressa in un qualche alfabeto, mentre un'altra parte ancora descrive le relazioni che uniscono le diverse informazioni, eventualmente provenienti da altri archivi.

EAGLE ha affrontato questa eterogeneità sviluppando come prima cosa un modello di metadati comune, in cui sono stati trasformati tutti i metadati raccolti dai diversi archivi epigrafici. L'infrastruttura di aggregazione di EAGLE è basata sul sistema D-NET (<http://www.d-net.research-infrastructures.eu>), sviluppato e consolidato dal CNR-ISTI nel corso di numerosi progetti internazionali. Il software D-NET fornisce un insieme di servizi facilmente personalizzabili per la raccolta di metadati, loro trasformazione, omogeneizzazione e indicizzazione per un motore di ricerca. Dopo la definizione del modello di metadati comune, il processo di aggregazione si compone di quattro fasi:

1. *Definizione delle trasformazioni dei metadati.* In collaborazione con gli esperti degli archivi sorgente, vengono definite (nel formato richiesto da D-NET) le regole strutturali e semantiche da utilizzare per trasformare i metadati in entrata (basati sullo schema dell'archivio sorgente) nello schema di metadati EAGLE.

2. *Trasformazione e "pulizia" dei metadati.* I metadati vengono prelevati dagli archivi sorgente, di solito con protocolli di comunicazione standard, come FTP o OAI-PMH, e vengono quindi trasformati in "oggetti EAGLE" utilizzando le regole definite alla Fase 1.

3. *Controllo di qualità dei metadati.* I metadati in formato EAGLE sono controllati e validati per identificare errori di trasformazione e possibili altri errori. Questo processo di controllo di qualità può portare alla ridefinizione delle regole di trasformazione definite nella Fase 1 e alla ripetizione della fase di raccolta (Fase 2).

4. *Messa a disposizione dei metadati.* I metadati (in formato EAGLE) che superano la Fase 3 vengono indicizzati per permettere interrogazioni sul portale EAGLE (alla Google) e vengono resi disponibili per la raccolta da parte di Europea.

Sorgenti di dati EAGLE

EAGLE aggrega contenuti forniti da circa 20 diversi archivi di tutta Europa. Mentre la maggior parte di questi fornisce metadati basati sullo standard *EpiDoc*, alcuni archivi li forniscono in formati diversi. Oltre agli archivi epigrafici, EAGLE aggrega poi dati da altre due diverse fonti: *pagine MediaWiki*, contenenti traduzioni di iscrizioni e *Trismegistos Identifiers*, contenenti informazioni relative a metadati diversi che si riferiscono alla stessa iscrizione.

Qui di seguito descriviamo brevemente queste diverse sorgenti.

Archivi basati su EpiDoc

La maggior parte delle fonti di dati EAGLE fornisce edizioni, cioè documenti epigrafici annotati secondo lo standard TEI/EpiDoc, che consente una descrizione digitale olistica testo-centrica di un'iscrizione e di tutti i suoi aspetti correlati. Le edizioni EpiDoc vengono fornite come documenti in formato XML, un formato flessibile ed elaborabile automaticamente, che è diventato lo standard di fatto per lo scambio di informazioni nel Web.

Archivi non basati su EpiDoc

In alcuni casi, per ragioni storiche o tecniche, alcuni dei *data providers* di EAGLE esportano i propri dati come documenti XML, ma non seguendo lo standard EpiDoc. L'archivio Arachne, ad esempio, comprende una vasta collezione di fotografie, disegni, schizzi e le scansioni di volumi antichi, e i metadati sono modellati secondo CIDOC RM (<http://www.cidoc-crm.org/>); l'archivio UbiEratLupa è focalizzato sulla descrizione di monumenti e i dati sono organizzati secondo uno schema XML proprietario; l'archivio del Cyprus Institute, infine, si occupa principalmente di letteratura e archeologia cipriota antica, concentrandosi sulla modellazione 3D dei monumenti e dei siti archeologici ed esporta i metadati sfruttando un proprio schema (STARC) (<http://www.cyi.ac.cy/starc/research-information/starc-ongoing-projects-ri/item/310-starc-akgdc-archaia-kipriaki-grammateia-digital-corpus.html>).

Trismegistos

Trismegistos (<http://www.trismegistos.org>), iniziativa ospitata dall'Università di Lovanio, identifica edizioni distinte relative alla stessa iscrizione, per mezzo di sofisticate tecniche di disambiguazione. Queste informazioni vengono fornite a EAGLE per evitare che a seguito di una interrogazione vengano inserite nel risultato più "copie" della stessa iscrizione.

MediaWiki

EAGLE ha sviluppato anche una piattaforma MediaWiki (http://www.eagle-network.eu/wiki/index.php/Main_Page), che ospita traduzioni, in una o più lingue moderne, delle iscrizioni contenute in EAGLE. L'aggregatore EAGLE raccoglie anche tutte queste informazioni e le integra con le informazioni ricevute direttamente dagli archivi.

Il modello comune di dati

La necessità di supportare interrogazioni su un materiale così eterogeneo come quello raccolto dai vari archivi sorgente ha portato alla definizione di un modello dei dati in grado di mettere in relazione concetti e oggetti separati in modo semplice e intuitivo. Il modello EACM (*Epigraphy Aggregation Conceptual Model*) si basa su un numero limitato di entità base che permette comunque di esprimere tutte le diverse sfaccettature del materiale correlato alle epigrafi, quali supporti fisici, testi, traduzioni, immagini e altre informazioni di contesto, e consente all'utente di effettuare interrogazioni sofisticate per recuperare con precisione il materiale di interesse.

Il modello EACM è costituito da un'entità *root* (la radice) da cui è possibile creare istanze di quattro sotto-entità: (i) *Artefact* che contiene informazioni circa la natura fisica di un oggetto nel dominio epigrafico; (ii) *Inscription* che descrive la natura testuale e semantica di un testo eventualmente presente su un *Artefact*; (iii) *Visual representation* che raccoglie tutte le informazioni relative alla



2. CNR – Ricercatori del team EAGLE

natura visiva di un manufatto, che si tratti di una pietra, un monumento o un oggetto che fornisce un contesto ad altri oggetti epigrafici; (iv) *Documental manifestation* che contiene tutte le informazioni relative alla natura testuale di una iscrizione e ha due sotto-entità: la *Transcription*, che raccoglie le informazioni relative al testo dell'iscrizione, e la *Translation*, che raccoglie le traduzioni in lingue moderne del testo (interpretato) dell'iscrizione.

Tutte le informazioni da aggregare in EAGLE trovano il loro posto in una o più istanze di tali sotto-entità.

L'infrastruttura di aggregazione e i workflow

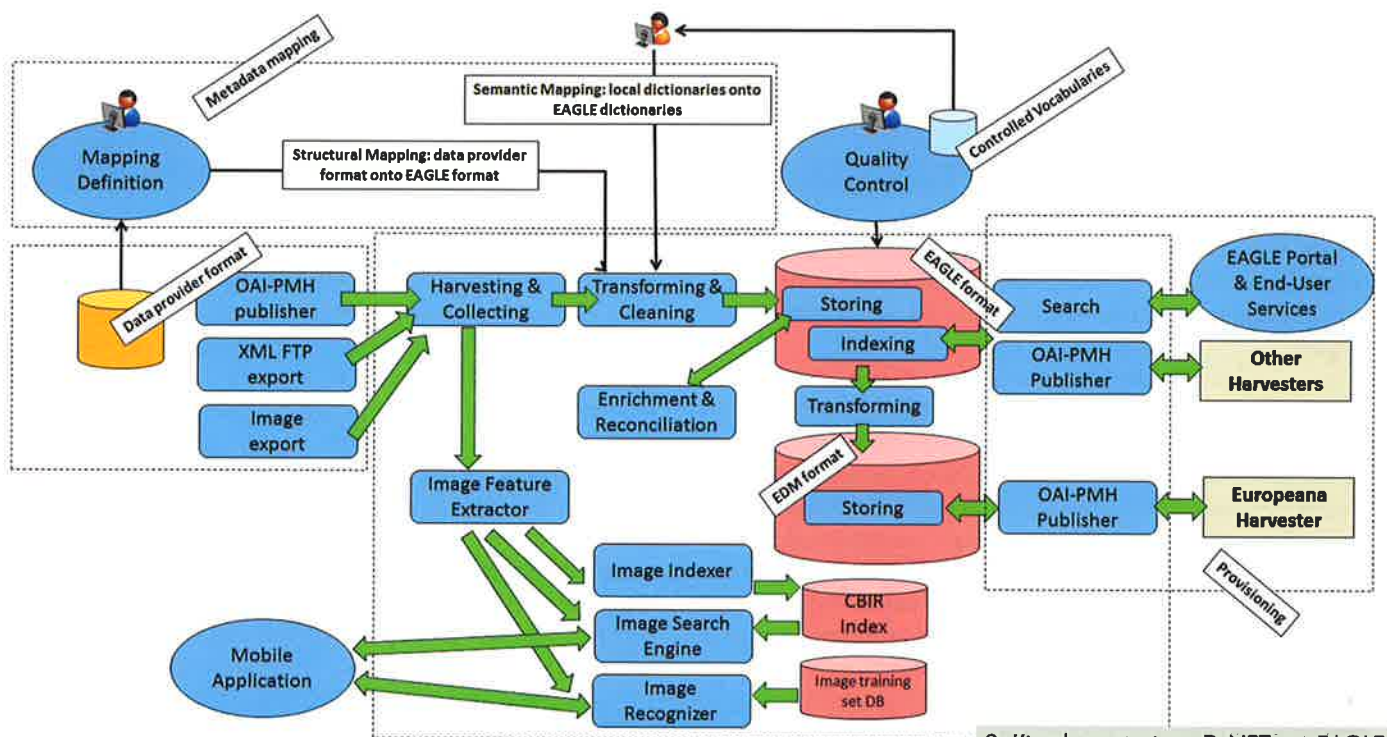
L'infrastruttura di aggregazione dati di EAGLE è basata sul software D-NET, personalizzato per soddisfare le esigenze di EAGLE (fig. 3). I servizi di *Collecting*, *Transforming* e

I workflow

In D-NET l'elaborazione dei dati viene specificata definendo i *workflow* (cioè sequenze di passi elementari) e *meta-workflow* (cioè sequenze di *workflow*). Entrambi possono essere facilmente configurati, programmati ed eseguiti attraverso l'interfaccia grafica di D-NET, mentre l'implementazione dei passi elementari avviene scrivendo un codice che realizza l'esecuzione del procedimento necessario. Ciascun *workflow* realizza un compito specifico e ben definito, come la raccolta di metadati, la trasformazione, l'indicizzazione, ecc.

Ad esempio, il *meta-workflow* per l'elaborazione dei metadati in formato EpiDoc consiste nella sequenza dei seguenti *workflow*:

- *Collect* (raccolta dei metadati originali da un archivio);
- *Transform* (trasformazione dei metadati nel formato EAGLE);



3. L'implementazione D-NET per EAGLE

Cleaning sono stati definiti in modo da soddisfare i requisiti degli archivi epigrafici sorgenti di EAGLE. Inoltre D-NET è stato ampliato con i servizi di elaborazione delle immagini per supportare il servizio di riconoscimento di immagini.

Il processo di aggregazione

Esperti degli archivi sorgente e amministratori dell'infrastruttura hanno definito insieme le regole per trasformare i metadati nativi nel formato EAGLE. Le regole di trasformazione coprono due aspetti: (i) un aspetto strutturale che definisce la trasformazione dello schema; (ii) un aspetto semantico che definisce la trasformazione dai vocabolari locali a quelli di EAGLE. Operativamente, queste trasformazioni avvengono tramite "programmi" XSLT che operano sui documenti XML in ingresso.

- *Clean* (verifiche di correttezza e allineamento semantico);
- *Index* (creazione dell'indice per supportare interrogazioni e navigazione);
- *OAI* (trasformazione dei metadati EAGLE nel formato richiesto da un *harvester* OAI-PMH).

Durante il passo di trasformazione i metadati raccolti vengono uniformati nel testo mediante markup EpiDoc conforme al modello EAGLE, vengono decomposti in molteplici oggetti EAGLE (*Artefact*, *Inscription*, *Transcription*, *Visual representation*) e, poi, arricchiti delle informazioni fornite da Trismegistos e delle traduzioni ottenute attraverso EAGLE MediaWiki. I metadati finali vengono, infine, indicizzati e resi disponibili per la ricerca e la navigazione da parte degli utenti EAGLE.

Nel *workflow* OAI i metadati finali sono sottoposti a un'ulteriore elaborazione per estrarre il sottoinsieme di infor-



4. L'Area della Ricerca CNR di Pisa

mazioni richieste da Europeana, che viene quindi trasformato nel formato richiesto dal servizio *harvester* OAI-PMH di Europeana.

Gli altri tipi di metadati (metadati non EpiDoc, dati da MediaWiki e Trismegistos) sono trattati con *meta-workflow* molto simili a quello descritto.

Il supporto del portale e delle applicazioni

Come mostrato in fig. 3, le funzionalità dell'aggregatore vengono rese disponibili all'esterno tramite interfacce specifiche. La principale interfaccia è *search*, che riceve le interrogazioni fatte da un utente attraverso l'interfaccia grafica del portale, e restituisce il risultato, ottenuto utilizzando il motore di ricerca interno (Solr e Lucene: <http://lucene.apache.org/solr/>).

L'interfaccia grafica del portale supporta una "ricerca semplice", alla Google, e una "ricerca avanzata" in cui l'utente può specificare il valore dei principali campi che descrivono l'epigrafe, in modo da effettuare una ricerca più selettiva. In entrambi i casi la richiesta dell'utente viene trasformata in una chiamata all'interfaccia *search*, fornendo gli opportuni valori per i parametri della chiamata.

L'interfaccia *search* viene usata anche dall'applicazione *Storytelling*, che mette a disposizione dell'utente strumenti che permettono di creare storie attorno alle epigrafi, in modo che il pubblico possa fruire in modo più facile e immediato del ricco materiale messo a disposizione da EAGLE. L'interfaccia *OAI-PMH Publisher* supporta le richieste di metadati ricevute tramite il protocollo OAI-PMH, come suggerisce il nome. In aggiunta a Europeana è quindi possibile per altre applicazioni esterne accedere al patrimonio epigrafico di EAGLE utilizzando una interfaccia standard.

L'interfaccia *Mobile Application* supporta le richieste dell'applicazione di riconoscimento immagini, che permette a un utente di ottenere informazioni riguardo a una iscrizione visibile (ad esempio in un sito archeologico o in un museo o in un libro). Tramite questa applicazione un utente può scattare una foto con un dispositivo mobile (Smartphone o Tablet) e inviarla al portale EAGLE specificando il tipo di riconoscimento desiderato. Nel modo *Exact Match* l'utente riceverà tutte le informazioni riguardo

all'iscrizione inviata (se presente nel database e riconosciuta). Nel modo *Similarity Search* l'utente riceverà una lista di iscrizioni, in ordine di somiglianza, e cliccando su un elemento della lista riceverà tutte le informazioni relative a quell'iscrizione.

Per supportare queste funzionalità, il software D-NET è stato arricchito con la parte relativa all'elaborazione e indicizzazione di immagini. Durante la fase di raccolta, tutte le immagini provenienti da un archivio sorgente (incluse quelle individuate da puntatori) vengono passate all'*Image Feature Extractor*, il quale estrae dalle immagini le principali caratteristiche visuali, per essere inserite nell'indice. Al momento della richiesta di riconoscimento (effettuata fornendo all'interfaccia l'immagine da riconoscere), il motore di ricerca (CBIR-Content Based Image Retrieval) accede all'indice ed estrae tutte quelle immagini che hanno caratteristiche visuali simili, mettendole in ordine di somiglianza.

*Andrea Mannocci, *Vittore Casarosa, *Paolo Manghi, *Franco Zoppi, Consiglio Nazionale delle Ricerche - Istituto di Scienza e Tecnologie dell'Informazione "A. Faedo", Pisa



5. La "nube" dei metadati



6. La Mobile Application

Bibliografia essenziale

V. CASAROSA, P. MANGHI, A. MANNOCCHI, E. RIVERO RUIZ, F. ZOPPI, "A Conceptual Model for Inscriptions: Harmonizing Digital Epigraphy Data Sources", in *EAGLE International Conference on Information Technologies for Epigraphy and Digital Cultural Heritage in the Ancient World*, Paris, September 29-30 and October 1 2014, pp. 23-40

P. MANGHI, M. ARTINI, C. ATZORI, A. BARDI, A. MANNOCCHI, S. LA BRUZZO, L. CANDELA, D. CASTELLI, P. PAGANO, "The D-NET Software Toolkit: A Framework for the Realization, Maintenance, and Operation of Aggregative Infrastructures", in <<http://www.emeraldinsight.com/journals.htm?issn=0033-0337>>