

# Standard e interoperabilità in ARIADNE

Paola Ronzino

<sup>1</sup> PIN, Prato, Italia  
paola.ronzino@pin.unifi.it

**Abstract.** The paper describes the activities undertaken by the ARIADNE project to achieve interoperability of archaeological resources at the datasets and item level. Moreover, the archaeological extension of the CIDOC CRM, CRMarchaeo, developed within the project, to enable the representation of existing archaeological datasets in a formal way is presented.

**Keywords:** CIDOC CRM, interoperability, standards, CRMarchaeo.

## 1 Introduzione

Negli ultimi anni è stata accertata una crescente consapevolezza riguardo l'importanza di creare delle reti di dati che permettano di effettuare un accesso integrato alla documentazione e agli archivi di risorse archeologiche digitali creati e gestiti da diverse organizzazioni e centri di ricerca. Un'importante condizione al fine dello sviluppo di tali reti di dati risiede nella definizione di standard e linee guida che stabiliscano un grado di compatibilità tra le basi di dati che costituiscono tali reti.

Generalmente questi dati sono archiviati in database individuali, non standardizzati e pertanto con una limitata possibilità di integrazione. Le diverse comunità di settore che hanno l'esigenza di archiviare i propri dati strutturandoli secondo i propri modelli, hanno definito negli anni degli standard per asservire alle necessità del proprio settore di ricerca. Quando però le diverse comunità decidono di condividere i propri dati con una comunità più estesa e per scopi più ampi, è allora che si pone il problema legato all'interoperabilità.

Molte istituzioni hanno sviluppato degli standard e linee guida per rendere la propria documentazione uniforme, per promuovere la condivisione dei dati, facendo ricorso all'uso di tesauri e vocabolari controllati, strumenti che consentono di migliorare l'accesso alle informazioni archeologiche e di rendere i propri dati interoperabili.

Alcuni di questi standard sono stati definiti in ambito nazionale, come per esempio le normative sviluppate dall'ICCD in Italia [1, 2], o le linee guida dettate dall'English Heritage attraverso il MIDAS [3]. Altri sono stati definiti all'interno di progetti europei il cui scopo era quello di facilitare l'interoperabilità dei dataset disponibili, come ad esempio, gli schemi di LIDO [4], CARARE [5] e 3D ICONS [6].

Databases differenti sono definiti interoperabili quando le operazioni di manipolazione effettuate su di essi restituiscono risultati coerenti indipendentemente dallo schema, dal modello o dal formato in cui i singoli dati sono archiviati. Rendere interoperabili delle basi di dati è un compito piuttosto impegnativo da raggiungere nell'ambito di database archeologici.

Questa è la sfida che si pone ARIADNE [7], l'infrastruttura di ricerca europea per l'integrazione dei dataset archeologici il cui obiettivo principale è quello di fornire un accesso integrato e garantire l'interoperabilità semantica di dataset archeologici distribuiti in tutta Europa. ARIADNE ha come obiettivo quello di riunire e integrare diversi insiemi di dati archeologici in modo che i ricercatori possano fare uso di queste risorse e beneficiare di esse attraverso l'uso di tecnologie e servizi messi a disposizione dalla infrastruttura stessa.

Il primo passo verso l'integrazione di dati archeologici in ARIADNE consiste nella possibilità di scoprire quali risorse sono disponibili, attraverso dei meccanismi di ricerca che permettono di rilevare l'esistenza di dati relativi a un determinato argomento, scoprendo il luogo in cui sono archiviati e ciò che rappresentano.

Questo servizio è garantito dall'ARIADNE Registry, grazie al quale tutte le risorse di dati archeologici disponibili nella rete sono state catalogate e le informazioni relative ai dataset disponibili modellate secondo l'ARIADNE Catalog Data Model (ACDM).

Un successivo livello di interoperabilità, il quale sarà condotto a livello sperimentale, si basa sulla semantica, permettendo l'integrazione di dati a livello del singolo record, e sarà basato sul CIDOC CRM [8].

Il CIDOC CRM è un'ontologia formale che permette di estrapolare un'enorme quantità di informazioni e di produrre nuova conoscenza dai dati disponibili, grazie alla sua capacità di descrivere le relazioni implicite esistenti tra persone, eventi, luoghi e documenti.

All'interno di ARIADNE, il CIDOC CRM è stato esteso per coprire le esigenze specifiche legate all'indagine archeologica, comprese le informazioni provenienti da scienze archeologiche, nonché da metodologie digitali utilizzate per il rilievo dei manufatti, dei monumenti e dei siti, dando vita all'ARIADNE Reference Model [9] (figura 1).

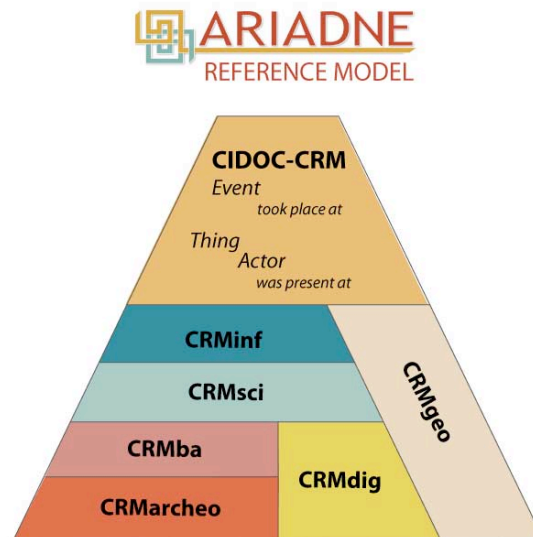


Fig. 1 L'ARIADNE Global Model

## 2 Standard archeologici e i dataset di ARIADNE

Durante l'autunno del 2013, all'interno del progetto ARIADNE, è stata effettuata una vasta indagine sugli standard utilizzati in ambito archeologico comprendendo anche quelli sviluppati da altri domini rilevanti per l'archeologia o che consentono l'interoperabilità con essi. L'indagine è stata condotta attraverso un'analisi approfondita della letteratura disponibile e contattando, in alcuni casi, le istituzioni che utilizzano tali normative per ulteriori delucidazioni sulla loro applicazione [10].

Parallelamente, un'ulteriore indagine ha coinvolto i partner del consorzio di ARIADNE i quali sono stati invitati a fornire informazioni relative ai dataset in loro possesso e che sarebbero stati messi a disposizione della rete per l'integrazione. Un questionario redatto ad hoc è stato distribuito ai *content providers* al fine di raccogliere le informazioni sul contenuto, sul formato dei dati, sulle dimensioni, sui metadati disponibili, sugli standard, gli schemi e sui vocabolari controllati utilizzati nella loro ricerca [11].

L'indagine ha rivelato che il Consorzio si avvale di diverse tipologie di dataset tra cui:

- banche dati archeologiche che comprendono siti, insediamenti, sepolture, reperti, oggetti appartenenti a specifiche regioni e periodi archeologici;
- database dei manufatti dell'industria litica: asce di pietra, strumenti litici;
- set di dati etno-archeologici;
- database di dati scientifici tra cui legno, carbone, frammenti paleo-ambientali, datazioni C14 e isotopi, dendrocronologia, archeozoologia, petrologia, dati geofisici, analisi chimiche di ceramiche e argille;

- dati di scavi archeologici e relazioni con siti e manufatti;
- set di dati di telerilevamento: lidar, fotografia aerea e geofisica;
- GIS archeologici comprendenti cartografia di siti archeologici, monumenti, indagini sul campo;
- report di scavo non pubblicati e altra letteratura (la cosiddetta letteratura grigia);
- immagini, PDF e altri documenti di testo, fogli di calcolo Excel e CSV, filmati, file audio e modelli 3D in una vasta gamma di formati.

Dall'indagine si è pertanto evinto che il Consorzio dispone di una grande quantità di dati disponibili per l'integrazione, e che gran parte di questi dati ha una natura estremamente eterogenea. Il risultato dell'indagine ha inoltre evidenziato che diversi partner dispongono di importanti collezioni di dati di diversa struttura e formato, mentre altri partner detengono singoli set di dati con struttura identica.

Un altro importante passo compiuto verso l'integrazione dei dati è stato quello di individuare delle aree comuni all'interno dell'ampio bacino di dati messi a disposizione dai partner di ARIADNE con lo scopo di identificare quei dataset che avrebbero permesso l'integrazione dei dati a livello del singolo record.

### **3 Integrazione dei dataset archeologici**

I dataset che sono stati identificati in seguito all'analisi dei dati disponibili, hanno fornito le informazioni necessarie per procedere alla pianificazione dei processi d'integrazione che in ARIADNE si eseguiranno su due livelli: (i) l'integrazione dei metadati, attraverso l'ARIADNE Registry, e (ii) l'integrazione dei dati, che permetterà di integrare le risorse selezionate da alcuni partner e di eseguire ricerche incrociate attraverso meccanismi di accesso alle risorse integrate.

L'ARIADNE Registry è un catalogo che permette di registrare le informazioni relative ai dataset e alle collezioni messe a disposizione dal Consorzio. Il catalogo è predisposto in modo da fornire la possibilità di aggiungere, in un secondo momento, ulteriori dataset detenuti da partner esterni associati al progetto. Tale catalogo fornirà uno strumento prezioso per la scoperta di risorse da parte di ricercatori e utenti europei e non solo.

Il Registry fornirà una piattaforma di interoperabilità per la comunità di ricerca che permetterà l'integrazione dei metadati raccolti nell'ARIADNE Catalog in modo da sostenere la ricerca incrociata rispondendo alle domande: (i) What, per recuperare informazioni per soggetto (ad esempio tipo di monumento, collezioni, etc.), (ii) Where, per consentire la scoperta di risorse digitali secondo criteri spaziali, (iii) When, per consentire la scoperta di risorse secondo criteri temporali.

L'integrazione di dati e metadati delle risorse archeologiche, a livello del singolo record si avvarrà invece di strumenti di mappatura basati sul CIDOC CRM e sulle sue estensioni archeologiche.

## **4 CIDOC CRM e la sua estensione archeologica**

Come anticipato, una delle sfide che ARIADNE si propone è quella di porre le basi per l'integrazione di informazioni complesse, provenienti da fonti eterogenee, superando l'attuale modello delle biblioteche digitali. Lo scopo principale è quello di ottenere una rappresentazione consistente dei dati che prescindano dalla modalità di creazione e dal loro uso, tale che tutte le informazioni direttamente o indirettamente correlate possano essere estratte dall'enorme quantità dei dati ed essere utilizzate a supporto di future interpretazioni.

Pertanto, la sfida non è quella di adattare i vari modelli esistenti ad un unico modello, ma di sviluppare uno schema globale sotto forma di una ontologia formale, che possa essere adattabile alle varie necessità e permettere quindi l'integrazione di dati eterogenei senza alcuna perdita di significato.

Al fine di affrontare la complessità dell'integrazione dei dati archeologici, ARIADNE ha scelto di adottare il CIDOC CRM e di estenderlo secondo le necessità ed i problemi posti dalla ricerca archeologica [9]. I dataset messi a disposizione dal Consorzio di ARIADNE sono stati analizzati con lo scopo di definire una base empirica di strutture di dati archeologiche identificando delle categorie distinte che avrebbero richiesto delle estensioni. Al fine di rappresentare le diverse metodologie di documentazione e la diversa struttura dei dati, sono state definite delle priorità concordate con i partner sulla base della quantità di dati a disposizione, della rilevanza nella catena di ragionamento e delle aspettative da parte degli esperti coinvolti. Per ogni categoria individuata sono stati selezionati e raccolti dei dati che sono stati mappati sul CIDOC CRM al fine di identificare eventuali concetti mancanti e di sviluppare le rispettive estensioni. Sulla base di queste premesse, alcuni esperti del settore sono stati coinvolti nelle attività del progetto, partecipando ai sistematici workshop interdisciplinari organizzati al fine di definire, in maniera collaborativa, la semantica dei dati e definire i nuovi concetti necessari per documentare al meglio i processi coinvolti nella ricerca archeologica. Questo lavoro ha portato allo sviluppo del modello ontologico CRMarchaeo [12].

### **4.1 CRMarchaeo: un'estensione del CIDOC CRM per modellare i dati archeologici**

Il motore trainante che ha spinto alla ricerca e alla definizione di uno standard ontologico a servizio della ricerca archeologica è stata la necessità, particolarmente forte tra la comunità di archeologi, di definire degli standard per la documentazione dello scavo archeologico. Finora le norme nazionali esistenti in Europa sono in parte riuscite a fornire una sistematizzazione coerente alla documentazione di scavo. Nonostante grandi siano stati gli sforzi compiuti in questa direzione dalle diverse comunità di ricercatori ed istituzioni archeologiche, notiamo una notevole differenza tra gli standard utilizzati a livello nazionale dalle varie scuole archeologiche ed istituzioni governative. La creazione di uno strumento in grado di identificare gli

elementi chiave coinvolti nel processo della ricerca archeologica e di metterli in relazione tra loro utilizzando una cornice semantica adeguata era solo auspicabile fino ad ora. Da questa necessità è nato il modello ontologico CRMarchaeo, un'estensione archeologica del CIDOC CRM in grado di fornire agli archeologi uno strumento per modellare in maniera semantica le informazioni e la documentazione prodotta dalle indagini archeologiche.

CRMarchaeo è stato creato mettendo a confronto standard e modelli utilizzati per la documentazione del processo di scavo archeologico già in uso da istituzioni nazionali e internazionali. Il modello si è sviluppato in seguito ad una approfondita analisi dei metadati esistente nella reale documentazione archeologica ed è stato arricchito dalla continua collaborazione con varie comunità di archeologi da diversi paesi e scuole.

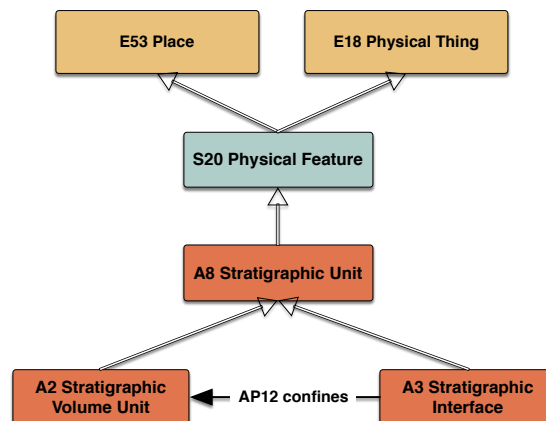
Come anticipato, uno degli obiettivi più importanti del modello è quello di superare le differenze risultanti dall'applicazione di varie tecniche di scavo e di procedure derivanti dalle diverse tradizioni archeologiche e scuole, per identificare i comuni metodi legati all'attività dello scavo archeologico. Questo servirà a fornire una visione unificata in grado di esprimere i concetti comuni senza imporre modelli specifici o particolari tecniche d'indagine, fornendo quindi una solida base per l'integrazione della documentazione della ricerca archeologica.

#### **4.1.1 CRMarchaeo: il modello concettuale**

L'interpretazione delle sequenze cronologiche, basate sull'analisi spaziale e temporale di un sito specifico, offre tutti gli elementi necessari alla ricostruzione dell'identità di un gruppo in uno luogo specifico ed in un determinato lasso temporale.

Uno degli elementi fondamentali del modello CRMarchaeo è infatti quello di offrire la possibilità di registrare le informazioni relative alla disposizione fisica della stratificazione archeologica e degli eventi che hanno condotto alla sua formazione, permettendo di documentare e successivamente interpretare la stratificazione e gli oggetti rinvenuti nel corso dello scavo, offrendo agli archeologi uno strumento di supporto nella determinazione di una cronologia relativa degli strati.

Il CRMarchaeo si sviluppa a partire dalla definizione delle principali entità coinvolte nell'indagine archeologica che sono le Unità Stratigrafiche (SU) e l'Interfaccia Archeologica (SI) (fig.2).

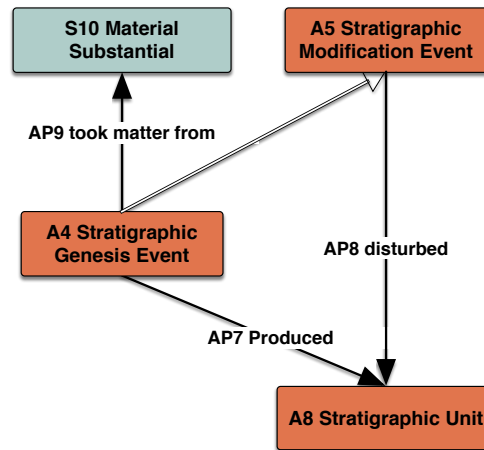


**Fig.2** Il grafico rappresenta il modo in cui una US è modellata in CRMarchaeo

L'unità Stratigrafica, intesa come la minima unità di informazione nella ricerca archeologica, è caratterizzata dalla sua natura fisica e spazio-temporale, essendo collocata in un determinato luogo per un determinato periodo ed è collegata con altre unità stratigrafiche. L'Unità Stratigrafica viene rappresentata in CRMarcheo dalla classe *A8 Stratigraphic Unit*, sottoclasse di *S20 Physical Feature*, entità sviluppata all'interno di un'altra estensione del CIDOC CRM, CRMsci [13], definito in origine per descrivere le entità relative alla stratigrafia geologica. S20 è una sottoclasse di *E18 Physical Thing* e *E53 Place*, entrambe classi del modello CIDOC CRM, pertanto la classe *A8 Stratigraphic Unit* di CRMarchaeo, ne eredita le caratteristiche fisiche e spaziali.

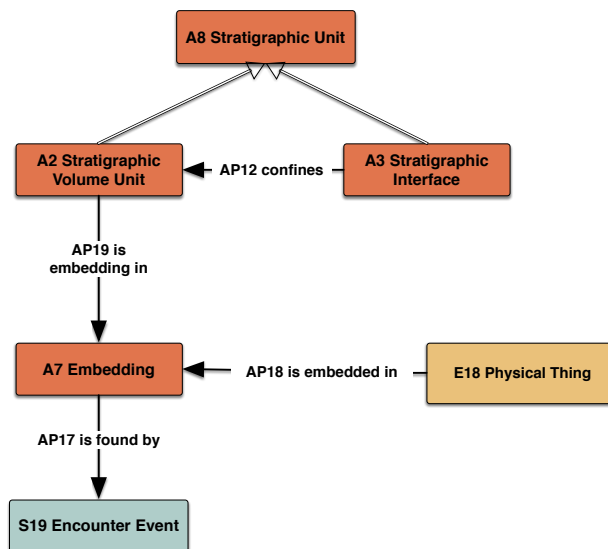
L'unità stratigrafica A8 è composta da una parte volumetrica definita con la classe *A2 Stratigraphic Volume Unit* e dalla interfaccia stratigrafica, rappresentata dalla classe *A3 Stratigraphic Interface*.

Le attività e i processi che hanno portato alla creazione di una unità stratigrafica sono rappresentate in CRMarcheo dalla classe *A4 Stratigraphic Genesis*, mentre i processi che ne hanno successivamente modificato la forma e la posizione, sono codificati tramite la classe *A5 Stratigraphic Modification* (figura 3).



**Fig.3** Il grafico riporta le relazioni che collegano gli eventi di genesi e modifica di una US. La doppia freccia rappresenta la relazione isA tra le classi

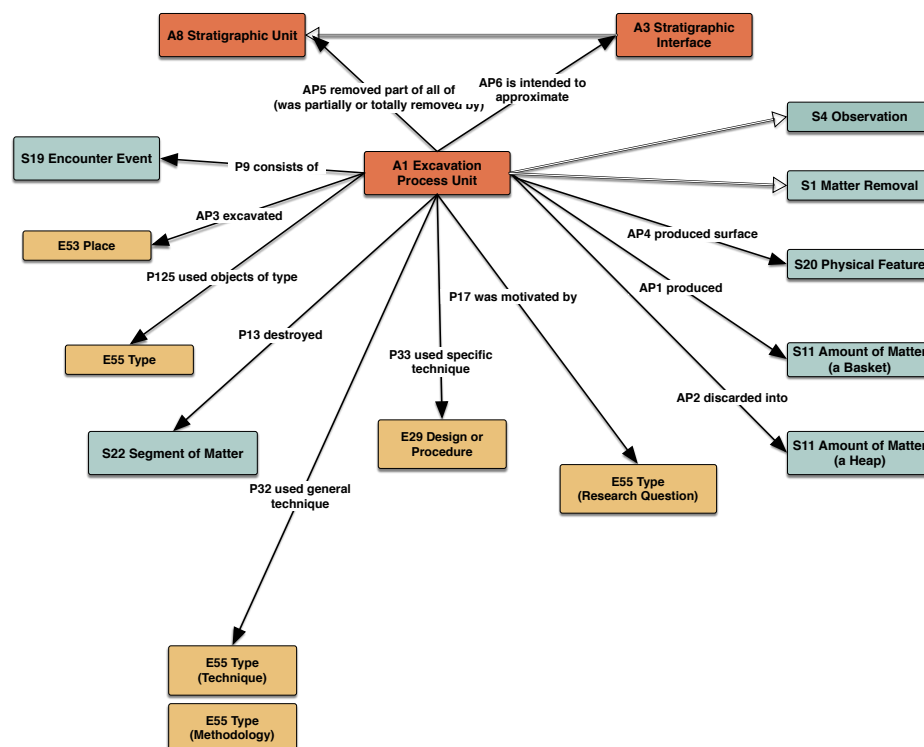
Un altro importante fattore a supporto dello studio sulla formazione dello strato archeologico riguarda l'analisi dei resti organici e inorganici rinvenuti all'interno dello stesso. Gli oggetti rinvenuti all'interno dell'US sono codificati attraverso la classe del CIDOC CRM *E18 Physical Thing*. La loro condizione di essere "contenuti" all'interno di una unità stratigrafica viene codificato attraverso la tripla *E18 Physical Thing* (oggetto fisico) -> AP18 is embedded (è contenuto) -> A7 Embedding (contenitore) (figura 4).



**Fig.4** Lo stato di "embedding" di un oggetto rinvenuto in una US



Le attività di documentazione e rimozione del terreno durante lo scavo archeologico, sono codificate attraverso la classe *A1 Excavation Process Unit*. Tutte le attività di scavo che si susseguono all'interno della stessa campagna di scavo, sono invece identificate attività complementari di una attività più ampia, esemplificata dalla classe *A9 Archaeological Excavation*.



**Fig.5** L'attività di scavo è modellata attraverso le classi di CRMarchaeo con l'adozione di classi e proprietà del CIDOC CRM core e dell'estensione CRMsci

Le relazioni stratigrafiche tra diverse US sono espresse attraverso la proprietà *AP11 has physical relation*, la quale, attraverso la proprietà *AP11.1 has type* descrive i rapporti tra US facendo riferimento ad un vocabolario che contiene i termini comuni utilizzati per documentare i rapporti tra US osservati e registrati durante uno scavo. I termini utilizzati da CRMarchaeo sono:

- fills
- is filled by
- cuts
- is cut by
- is bonded with
- butted

- jointed
- above
- below

Le relazioni stratigrafiche tra US sono invece codificate con la proprietà *AP13 has stratigraphic relation*, anch'essa popolata da vocabolario comune per rappresentare questi tipi di rapporti:

- before
- after
- same as

Il modello ontologico CRMarchaeo è stato approvato dai membri dello Special Interest Group (SIG) del CIDOC CRM come modello formalmente e metodologicamente compatibile con il CIDOC CRM. Il modello è in continuo aggiornamento ed è sempre aperto ad ogni tipo di integrazione o aggiunta che si ritenga necessaria, soprattutto in seguito a particolari necessità che possano emergere dalla sua applicazione ad uno specifico caso di studio. E' intenzione del CIDOC CRM SIG di gestire e promuovere il modello come standard internazionale.

## Ringraziamenti

Il presente lavoro è stato supportato dal progetto ARIADNE, finanziato dalla Commissione Europea (contratto 313193) nel contesto del VII Programma Quadro di Ricerca e Sviluppo, bando INFRA-2012-1.1.3. Le opinioni dell'autore non riflettono necessariamente quelle della Commissione Europea.

## Bibliografia

- [1] Mancinelli, M.L., 2004a. Nuovi strumenti per la catalogazione dei beni archeologici. In E. Plances & A. Leon, eds. La catalogazione statale censimento ed elementi di analisi. Roma: ICCD, pp. 35–45.
- [2] Mancinelli, M.L., 2004b. Sistema Informativo Generale del Catalogo: nuovi strumenti per la gestione integrata delle conoscenze sui beni archeologici. Archeologia e Calcolatori, 15, pp.115–128.
- [3] Lee, E. ed., 2003. MIDAS: A Manual and Data Standard for Monument Inventories 3rd ed., Swindon: English Heritage.
- [4] LIDO: <http://www.lido-schema.org/schema/v1.0/lido-v1.0-schema-listing.html>
- [5] CARARE project: <http://www.carare.eu/>
- [6] 3D ICONS: <http://www.3dicons-project.eu/>
- [7] ARIADNE infrastructure: <http://www.ariadne-infrastructure.eu>
- [8] CIDOC CRM (2015). Current Official Version of the CIDOC Conceptual Reference Model. La versione 6.2 del CRM è disponibile al sito <http://www.cidoc-crm.org/docs/cidoc-crm-version-6.2.pdf>

- [9] ARIADNE, 2016, D14.1 “Extended CRM”, 2016, <http://ariadne-infrastructure.eu/Resources/D14.1-Extended-CRM>
  - [10] ARIADNE, 2013, D 3.1 “Initial *report standards and on the project registry*”, 2013, <http://www.ariadne-infrastructure.eu/Resources/D3.1-Initial-Report-on-the-project-registry>
  - [11] ARIADNE, 2013, D 3.2 “*Report on project standards*”, 2013, <http://ariadne-infrastructure.eu/Resources/D3.2-Report-on-project-standards>
  - [12] CRMarchaeo, 2016. CRMarchaeo: the Excavation Model, version 1.4., <http://www.ics.forth.gr/isl/CRMext/CRMarchaeo/docs/CRMarchaeo1.2.1.pdf>
  - [13] CRMsci, 2014. CRMsci: the Scientific Observation Model, version 1.2.2. Available at <http://www.ics.forth.gr/isl/CRMext/CRMsci/docs/CRMsci1.2.2.pdf>
- L’accesso a tutti i siti risale a Luglio 2016