Making archaeological datasets
F-A-I-R
A panoramic view of the ARIADNE story

Julian Richards
Archaeology Data Service, University of York
Introducing the F-A-I-R principles

“One of the grand challenges of data-intensive science is to facilitate knowledge discovery by assisting humans and machines in their discovery of, access to, integration and analysis of scientific data.”

- **Findable**
  - Data are described with good metadata
  - Metadata are indexed in a searchable resource
  - Data are assigned a Permanent identifier

- **Accessible**
  - Data should be open and online

- **Interoperable**
  - Use a formal, open, shared language for knowledge representation

- **Re-usable**
  - Data should have clear data licenses
  - Metadata should meet domain-relevant standards
Challenges and opportunities

• There is lots of data in archaeology
  – Fragmented, distributed, heterogeneous, hidden, short-lived

• IT is there to enable
  – Infrastructure, integration, standards, mapping, access, preservation
Life before ARIADNE...
Project basics

- 4 year project (02/2013- 02/2017)
- FP7 Instrument “Integrating Activity”
- Funding 6.5m Euros
- Coordinators
  - Prof. Franco Niccolucci, University of Florence
  - Prof. Julian Richards, University of York
- Website: www.ariadne-infrastructure.eu
ARIADNE Community

- 23 partners in 18 European countries
- 9 ICT organisations
- 14 archaeological organisations
- 15 Associate partners
- Community building
  - Transnational access
  - Training events
  - Special interest groups
User Needs Research

- 94% of researchers agreed that it is important that datasets are available online in an uncomplicated way.
- 87% of researchers agreed that they often do not know what research data is available because it is stored in so many different places and databases.
- 74% of researchers consider it important to have easy access to international datasets.
- The additional effort required to prepare data (formatting, metadata, etc.) is considered to be a barrier to data sharing by 80% of researchers.
- The perceived lack of professional recognition and reward for sharing data is a barrier to data sharing for 72% of researchers.
- A lack of institutional or international repositories for archaeology data sets was a barrier to data sharing for 60% of researchers.
The ARIADNE roadmap

Dimensions of integration

1. Geographic extent and different languages
   - Global
   - Cross-country
   - Country
   - Region
   - 1 site

2. Metadata richness and standardisation
   - Legacy metadata, e.g., Dublin Core
   - General purpose, Application profiles
   - Domain ontologies, e.g., CIDCC-CRM

3. Addressing complexity
   - Extending the Conceptual Reference Model
   - Managing different knowledge models
   - Part-whole relations
   - Complex time-spatial relations
   - Richly integrated thesauri
   - Flat thesauri
   - Standardisation to high-level ontologies
The ARIADNE Architecture
Improving findability and accessibility

Carlo Meghini: ACDM
Dimitris Gavrilis: Registry
Achille Felicetti: TNA activity
Doug Tudhope
Sebastian Cuy
Data sharing

- Mobilize
- Integrate
- Make accessible

Archaeological Survey of Ireland, 141,000 records

ADS: 36,000 grey literature reports, over 1000 project archives; 1.3 million records

ARACHNE: 500,000 archaeol. objects / records and images

INRAP: 27,000 fieldwork reports

Dendrochronology data (DCCD) of 5200 objects

ZRC-SAZU (Slovenia), 11,000 site records

SIGECweb: 326,000 archaeo-logical finds records

FASTI online: 12,000 excavation reports of 3300 sites in 14 countries

Your data
The ACDM: ARIADNE Common Data Model
Achieving interoperability

- We have datasets in many languages and complying with many different standards
- ARIADNE uses the CIDOC CRM with extensions for archaeology to achieve integration
  - Existing datasets are mapped to the ARIADNE data model
  - Subject concepts are mapped to the Getty A&AT
Metadata enrichment in the Registry

* mono-lingual (prefLabel only)
** multi-lingual (prefLabel & altLabel)

Diagram:
- Provider Native Repository
- Excel Sheet XML Files
- ARIADNE Portal
- AAT
- Vocabulary Mapping Tool
- MORe
- Registry
- Elastic Search
- ACDM / Subjects (JSON)

Flow:
- Native Subjects to Vocabulary Mapping Tool
- Vocabulary Mapping Tool to MORe
- MORe to ACDM / Subjects (JSON)
- ACDM / Subjects (JSON) to ARIADNE Portal
- ACDM / Subjects (JSON) to Elastic Search

Note:
- * nativeSubjects
- providedSubjects
- nativeSubjects
- providedSubjects
- mappings
Interoperability Framework

ARIADNE e-Infrastructure and Integrated Services

ARIADNE Portal
Cross-search and other applications

ARIADNE Registry
Datasets, metadata schema, KOSs, mappings,..., CRM

Meta/data records => collect/transform2/RDF

Data Centers
(nat./int., archaeology or general)

Subject/Domain-based Repositories & Portals
(nat./international)

Institutional Repositories / DBs

Research Projects

Subject/Domain-based Referatories & Portals
(nat./international)

ARIADNE Registry
Datasets, metadata schema, KOSs, mappings,..., CRM

Several ARIADNE partners

ARIADNE aggregators
(nat./thematic )?

Legend:
dep=deposit, coll=collect, acc=access;
D= Data, M=Metadata

L4: e-infra & integrated services

L3: Higher-level data managers

L2: Institutions, institutional repositories/DBs

L1: Projects, research groups
ARIADNE brings together and integrates existing archaeological research data infrastructures so that researchers can use the various distributed datasets and new and powerful technologies as an integral component of the archaeological research methodology.
• Vocabulary concepts from the Getty Art & Architecture Thesaurus (AAT) are offered as users type a query
  – Enables query expansion
  – Supports multilingualism
Multilingual support

ring brooches

Getty AAT ID 300253352
Note Brooches in the shape of a ring, with a movable pin that may extend beyond the diameter of the ring; usually of metal and often ornamented. Used to fasten costume and worn especially during the Medieval period in Europe.
URI http://vocab.getty.edu/aat/300253352
Broader brooches

Terms

English ring brooches brooches, annular brooch, annular ring brooch annular brooches annular brooch brooches, ring ring-brooch

Spanish broches ring broche broche

Dutch ringbroches ringbroche

Provider mapping

ANNULAR BROOCH

Match URI http://www.w3.org/2004/02/skos/core#exactMatch
Source URI http://purl.org/heritagedata/schemes/MDA_OBJ/concepts/97114
An Archaeological Evaluation at 98-100 Bull Close Road, Norwich, Norfolk
Type: Site and monuments databases or inventories
Published: Archaeology Data Service
An Archaeological Evaluation at 98-100 Bull Close Road, Norwich, Norfolk

Archaeological Excavation Report: Norwich Road and Exeter Crescent Road alignment, RAF Lakenheath ERL 161
Type: Site and monuments databases or inventories
Published: Archaeology Data Service
Archaeological Excavation Report: Norwich Road and Exeter Crescent Road alignment, RAF Lakenheath ERL 161

Report on a Second Phase of Archaeological Evaluation at Land adjoining Swanton Morley Airfield, Beetley, Norfolk
Type: Site and monuments databases or inventories
Published: Archaeology Data Service
Report on a Second Phase of Archaeological Evaluation at Land adjoining Swanton Morley Airfield, Beetley, Norfolk

The Mucking Anglo-Saxon cemeteries
Type: Fieldwork archives
Published: Archaeology Data Service
The Anglo-Saxon cemeteries at Mucking, Essex, represent the burials of over 800 individuals from the 5th to early 7th centuries AD. The mixed rite Cemetery II is one of the largest and most complete Anglo-Saxon cemeteries yet excavated (282 inhumations, 463 cremation burials), while the pa...
Multilingual results via AAT mappings

ARIADNE Portal
Query on AAT subject: Settlements and Landscapes
shows results from IACA (Fasti), INRAP and DANS in multiple languages
ARIADNE and PeriodO

- PeriodO is a Linked Open Data gazetteer
  - Scholarly definitions of historical, art-historical and archaeological periods
  - Now includes period terms from ARIADNE partners
  - The terms have PeriodO URIs
- Allows for easier linking of datasets that define periods differently
  - (when is always linked with where in archaeology)

http://perio.do/
Transnational Activity

Training events

• 2D/3D documentation for archaeology
• Legacy data and dataset design
• Mapping existing datasets to CIDOC CRM
Improving interoperability and reusability
ARIADNE services

ARIADNE visual media service
Create your online showcase for 3d models, images and RTI.

Upload »  Browse »

3D models
3D representations produced with 3D scanners or photogrammetry are extremely high-resolution and hard to visualize at interactive rate. This service produces a webpage that supports interactive visualization of your data, after converting it into an efficient multiresolution encoding.

View details »  Demo

RTI images
Relightable images (called Reflection Transformation Images, RTI, or Polynomial Texture Maps, PTM) are becoming an increasingly used media. This service closes a current gap, giving support for easy publication on the web and interactive visualization of RTI images.

View details »  Demo

High-resolution images
High-resolution images are a commodity resource in archaeology. Unfortunately, they are most often disseminated and published on the web by using low-resolution versions (a single 40Mpixel images is 120MB in uncompressed format and around 10MB when lossy compressed).

View details »  Demo

http://visual.ariadne-infrastructure.eu/
Visual media service

Viewing the results 3D:

- 80.23 MB
- 3D model of the Arringatore produced by 3D laser scanning
- National Archaeological Museum and CNR-ISTI

http://visual.ariadne-infrastructure.eu/3d/arringatore
Visual media service

Viewing the results RTI:

- 199.4 mb
- Reverse, Medallion, Rome, Hadrian
- Palazzo Blu, ISTI-CNR, Simoneschi collection

http://visual.ariadne-infrastructure.eu/rti/medallion_hadrian_reverse
Landscapes Services for ARIADNE are a set of responsive web services that include large terrain datasets generation, 3D landscape composing and 3D model processing, leveraging on powerful open-source frameworks and toolkits such as GDAL, OSGjs, OpenSceneGraph and ownCloud. Here are a few examples of 3D datasets produced by the services:

http://landscape.ariadne-infrastructure.eu/
ARIADNE Reference Model

Few concepts, high recall

Special concepts, high precision
NLP and Data Mining

Information Extraction → ARIADNE Data store

Search/Query (ARIADNE portal)

Pattern(s)

Archaeological knowledge

NLP
Dendrochronology Case Study

- Extracts of 5 archaeological datasets, output from NLP on 25 grey literature reports
- Multilingual - English, Dutch and Swedish data
- Data integration via CIDOC CRM and Getty AAT
- 1.09 million RDF triples
- 23,594 records
- 37,935 objects
- Demonstration query builder for easier cross-search and browse of integrated datasets
Standards: Guides to Good Practice

Archaeology Data Service / Digital Antiquity
Guides to Good Practice

New Guides:

• Dendrochronology
• 3D Models in Archaeology
Preservation Services

- Care for preservation
- Developing trusted digital repositories:
  - ADS, DANS, IANUS, OAEW
- International collaboration on guidelines
5-10 year Innovation Agenda

- Research e-infrastructures and digital resources for archaeological research and the related domains
- Open sharing and re-use of data: promoting a culture of data sharing, re-use and citation, removing barriers to data sharing
- Data archives for the curation of archaeological research data: reliable and cost-effective community archives for long-term data curation and access
- Capacity building: guidance, training and support for data practices
- Providing services and enabling novel applications
Conclusion

Ariadne was the daughter of Minos, King of Crete. She gave a ball of thread to Theseus so that he could find his way out of the Minotaur's labyrinth.

Our ARIADNE is a research infrastructure. We give a portal so researchers can find archaeological datasets and tools to visualise and publish their results.
IMPROVING FINDABILITY AND ACCESSIBILITY

Introduction
The ARIADNE Common Data Model
The ARIADNE Registry
Metadata Mappings: What, When & Where
The ARIADNE Portal
Transnational Activity and training

Achille Felicetti, PIN
Carlo Meghini, CNR
Dimitris Gavrilis, ATHENA DCU
Doug Tudhope, Univ South Wales
Sebastian Cuy, DAI
Achille Felicetti, PIN
IMPROVING INTEROPERABILITY AND REUSABILITY

Introduction
Visual Media Services
Natural Language processing and Data Mining
Linked Data Experiments
Information modeling
Guides to Good Practice
Preservation services

Carlo Meghini, CNR
Roberto Scopingo, CNR
Hans Kamermans, Leiden Univ
Maria Theodoridou, FORTH
César Gonzalez-Perez, CSIC
Holly Wright, ADS, Univ of York
Hella Hollander, DANS
Thanks for your attention

email: info@ariadne-infrastructure.eu

http://www.ariadne-infrastructure.eu
http://portal.ariadne-infrastructure.eu

Twitter: @ariadne-infrastructure

ARIADNE is a project funded by the European Commission under the Community’s Seventh Framework Programme, contract no. FP7-INFRASTRUCTURES-2012-1-313193. The views and opinions expressed in this presentation are the sole responsibility of the authors and do not necessarily reflect the views of the European Commission.