Integration of Archaeological Datasets Through the Gradual Refinement of Models

Cesar Gonzalez-Perez · Patricia Martín-Rodilla

Institute of Heritage Sciences (Incipit)
Spanish National Research Council (CSIC)
Problem

- Scattered archaeological information
- Different:
  - Technologies
  - Structures
  - Conceptualizations
- How do we integrate disparate sources?
Conventional Solutions

- Use a data-oriented approach
- Emphasize technology and structure
- Map syntactical elements (e.g. tables, fields) at the implementation level

But:
- Assumes relational, XML or similar
- Is oblivious to abstraction
Conventional Solutions

![Image of a computer interface with details about a site, including its code, type, ACC, and dimensions.]

**Site**

- **Code:** 1 Text
- **Type:** 0..1 Number
- **ACC:** 0..1 Number
- **Dimensions:** 0..1 Text
- **Description:** 0..1 Text
Conventional Solutions
We need a better solution

- Aware of abstraction issues
- Emphasis on the conceptual level
Gradual refinement of models

1. Agree on very abstract conceptual framework
2. Create particular model for each dataset
3. Reconcile those models into a common, shared model
4. Look at each dataset through the common model
Gradual refinement of models

Abstract Reference Model

Common Model

Model A

Dataset A

Model B

Dataset B

Model Z

Dataset Z
Consequences

- Abstraction means removal of details:
  - The more similar two datasets are, the fewer details a common model will need to remove
  - The more abstract a model is, the more datasets it will cover
- Models determine what the users see and how they interact with data
Case study

- **SIA+**
  - Integrated information system
  - Incipit

- **Camiño Primitivo**
  - Project-specific information system
  - Incipit

- **Roman Essex**
  - Project-specific database
  - Available from ADS
Case study

CHARM

Overall Model

Model for Incipit

Model for SIA+

Model for Camiño

Model for Roman Essex

SIA+

Camiño

Roman Essex

abstraction
Case study: CHARM

- Cultural Heritage Abstract Reference Model
- Very abstract, avoids concrete concepts
  - Includes e.g. Object, Place, Agent
  - Avoids e.g. Hillfort, Post hole, Excavation
- Integrates usual archaeological concepts with additional areas for multi-discipline work

Copyright © 2014 Cesar Gonzalez-Perez & Patricia Martín-Rodilla
Case study: CHARM
Case study: Camiño model
Case study: Roman Essex model
Integrating SIA+ and Camiño models

- SIA+ emphasizes finds
- Both use features
- Camiño adds performative entities
- Both use photographs to represent other entities
- Both use impact assessment, although from different perspectives
- Both use status assessments
Integrating SIA+ and Camiño models

Shared, organization-wide model for Incipit.
SIA+ model as extension of common

Location (A)  \( \xrightarrow{\text{IsLocatedAt (T)}} \)  PrimaryEntity (A)  \( \xrightarrow{\text{Valorizes}} \)  MaterialEntity (A)  \( \xrightarrow{\text{OccursTo}} \)  ObjectEntity (A)

Material: 0..1 Text

Place (A)  \( \xrightarrow{\text{OccursTo}} \)  PerformatveEntity (A)

Type: 0..1 Text

CreationEvent

Moment: 1 Time

Valueization (A)

Contents: 1 Text

Representation

ArchaeologicalImpact

Remove

Name: 0..1 Text

Description: 0..1 Text

DateTaken: 1 Time

BatchNumber: 1 Number

DocumentNumber: 1 Number

PotteryFind

IsDecorated: 1 Boolean

Shape: 0..1 Text

PotPart: 1 enum PotPart

LithicFind

LithologicalDescription: 0..1 Text

AreaOfInterest

Name: 0..1 Text

Description: 0..1 Text

PointOfInterest

Name: 0..1 Text

UTM X: 1 Number

UTM Y: 1 Number

Altitude: 1 Number

Toponym: 0..1 Text

Town: 0..1 Text

Parish: 0..1 Text

Municipality: 1 Text

Province: 1 Text

Country: 1 Text

ConservationStatus

Description [Contents]: 1 Text (S)

AdministrativeValorization (A)

Discipline: 1 Text

Valorizes 0..* 0..0

MemberOfStaff

Surname: 0..1 Text

FamilyName: 1 Text

Author

GivenName: 0..1 Text

FamilyName: 0..1 Text

Impact

Cause: 1 Text

Distance: 1 Number

Intensity: 1 enum ImpactIntensity (S)

AlterationCauses: 0..* enum AlterationCause

AlterationAgents: 0..* enum AlterationAgent

Archaeology

Remove

Renamed: ArchaeologicalImpact

Parentage

ConditionValorization

ScientificTechnicalValorization (A)

Discipline: 1 Text

Valorizes 0..* 0..0

Location (A)

IsLocatedAt (T)

0..1

0..1

0..*
Camiño model as extension of common
Integrating Incipit and Essex models

- Hypothesis: models within an organization would share a significant amount of conceptualizations, more than external models
- Not supported by this case study
  - Very small subsets of SIA+ and Essex models
  - Quite narrow project scopes for Essex and Camiño
Integrating Incipit and Essex models

- Location and Find concepts are similar between SIA+ and Roman Essex models
- Connecting Roman Essex to Camiño is harder

- CHARM still works as conceptual bridge
# Data integration consequences

*When we look at e.g. a real estate element in the Camiño dataset...*

<table>
<thead>
<tr>
<th>Camiño</th>
<th>Incipit</th>
<th>Overall</th>
</tr>
</thead>
</table>
| • All data are available, such as AdminCode, Subtype, or ConservationStatus.  
  • Enum data such as Type are expressed in maximum detail.  
  • No datasets are interoperable.  | • Only some data such as Dating and Dimensions are available. Others such as AdminCode are not.  
  • Enum data such as Type are expressed at a reduced level of detail.  
  • Datasets from Incipit are interoperable. | • Basic data are available, such as Materials and links to related objects.  
  • The object is still accessible.  
  • Any CHARM-based dataset is interoperable. |

---

*abstraction*

---

Copyright © 2014 Cesar Gonzalez-Perez & Patricia Martín-Rodilla
Conclusions

- Conceptual models allow us to:
  - Understand meaning of data
  - Relate datasets to each other
  - Look at data at the appropriate level of detail
- CHARM must be solid; your input is welcome
- Tool support is crucial; work in progress
Thank You

Cesar Gonzalez-Perez
cesar.gonzalez-perez@incipit.csic.es

CSIC  Incipit

www.charminfo.org  www.conml.org