

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)

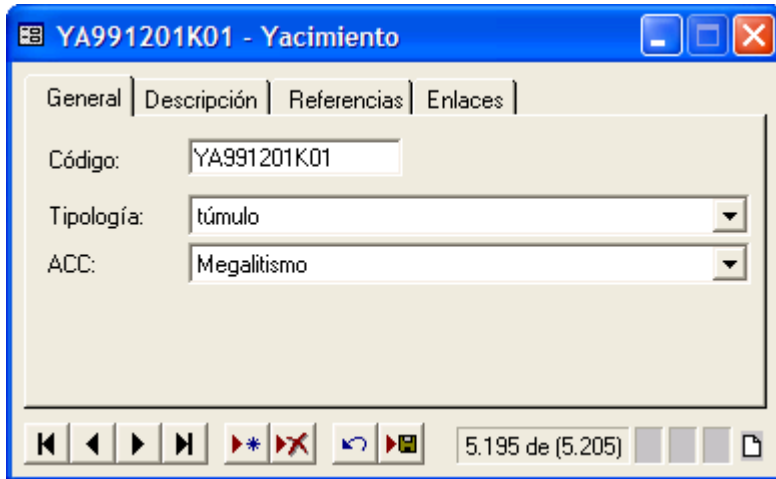
Problem

- 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



YA991201K01 - Yacimiento

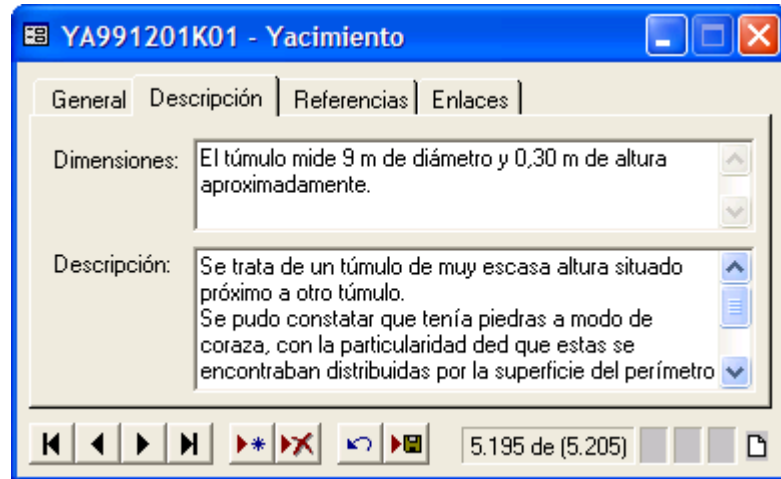
General | Descripción | Referencias | Enlaces

Código: YA991201K01

Tipología: túmulo

ACC: Megalitismo

5.195 de (5.205)



YA991201K01 - Yacimiento

General | Descripción | Referencias | Enlaces

Dimensiones: El túmulo mide 9 m de diámetro y 0,30 m de altura aproximadamente.

Descripción: Se trata de un túmulo de muy escasa altura situado próximo a otro túmulo. Se pudo constatar que tenía piedras a modo de coraza, con la particularidad de que estas se encontraban distribuidas por la superficie del perímetro

5.195 de (5.205)

Site
Code: 1 Text
Type: 0..1 Number
ACC: 0..1 Number
Dimensions: 0..1 Text
Description: 0..1 Text

Conventional Solutions

YA991201K01 - Yacimiento

General | Descripción | Referencias | Enlaces

Código: YA991201K01

Tipología: túmulo

ACC: Megalitismo

5.195 de (5.205)

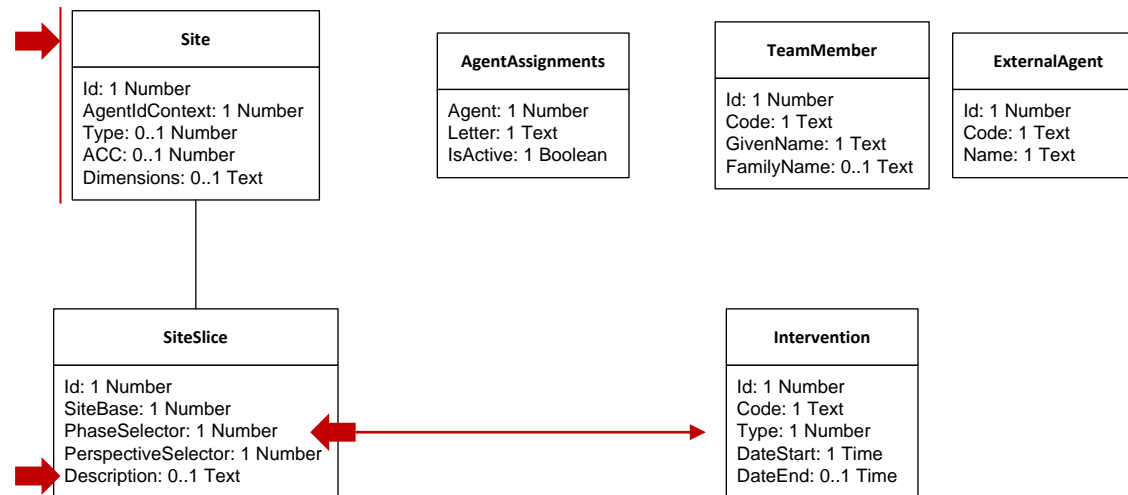
YA991201K01 - Yacimiento

General | Descripción | Referencias | Enlaces

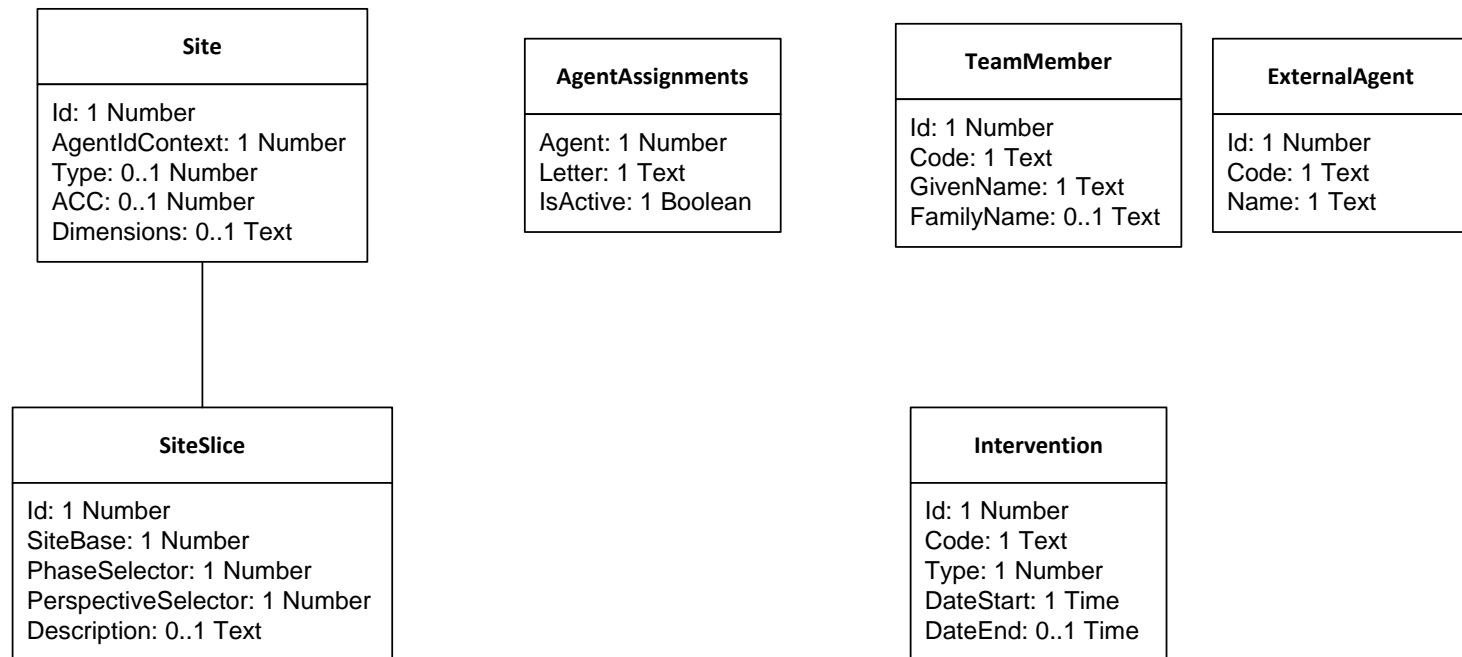
Dimensiones: El túmulo mide 9 m de diámetro y 0,30 m de altura aproximadamente.

Descripción: Se trata de un túmulo de muy escasa altura situado próximo a otro túmulo. Se pudo constatar que tenía piedras a modo de coraza, con la particularidad de que estas se encontraban distribuidas por la superficie del perímetro

5.195 de (5.205)



Conventional Solutions



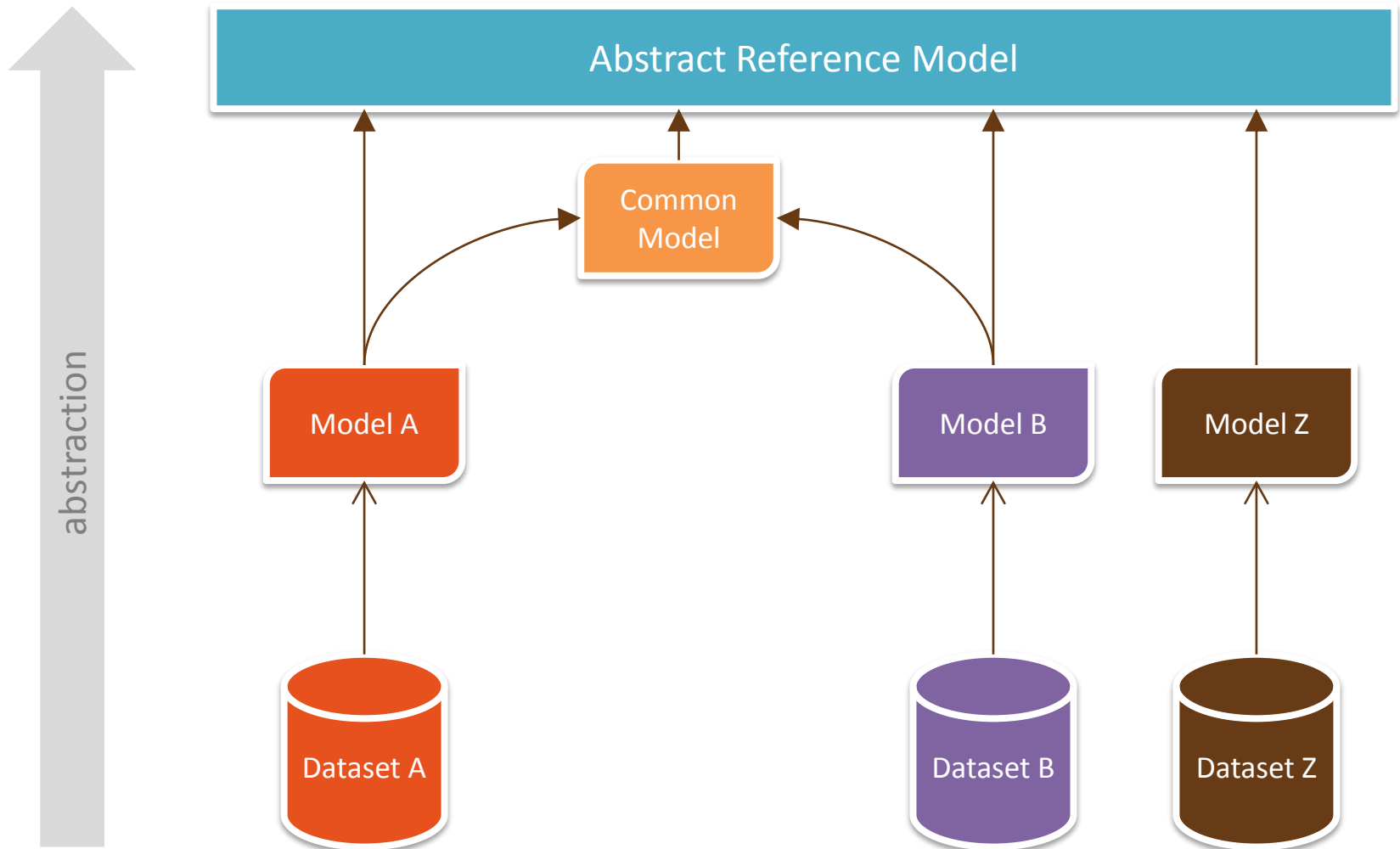
We need a better solution

- 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



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

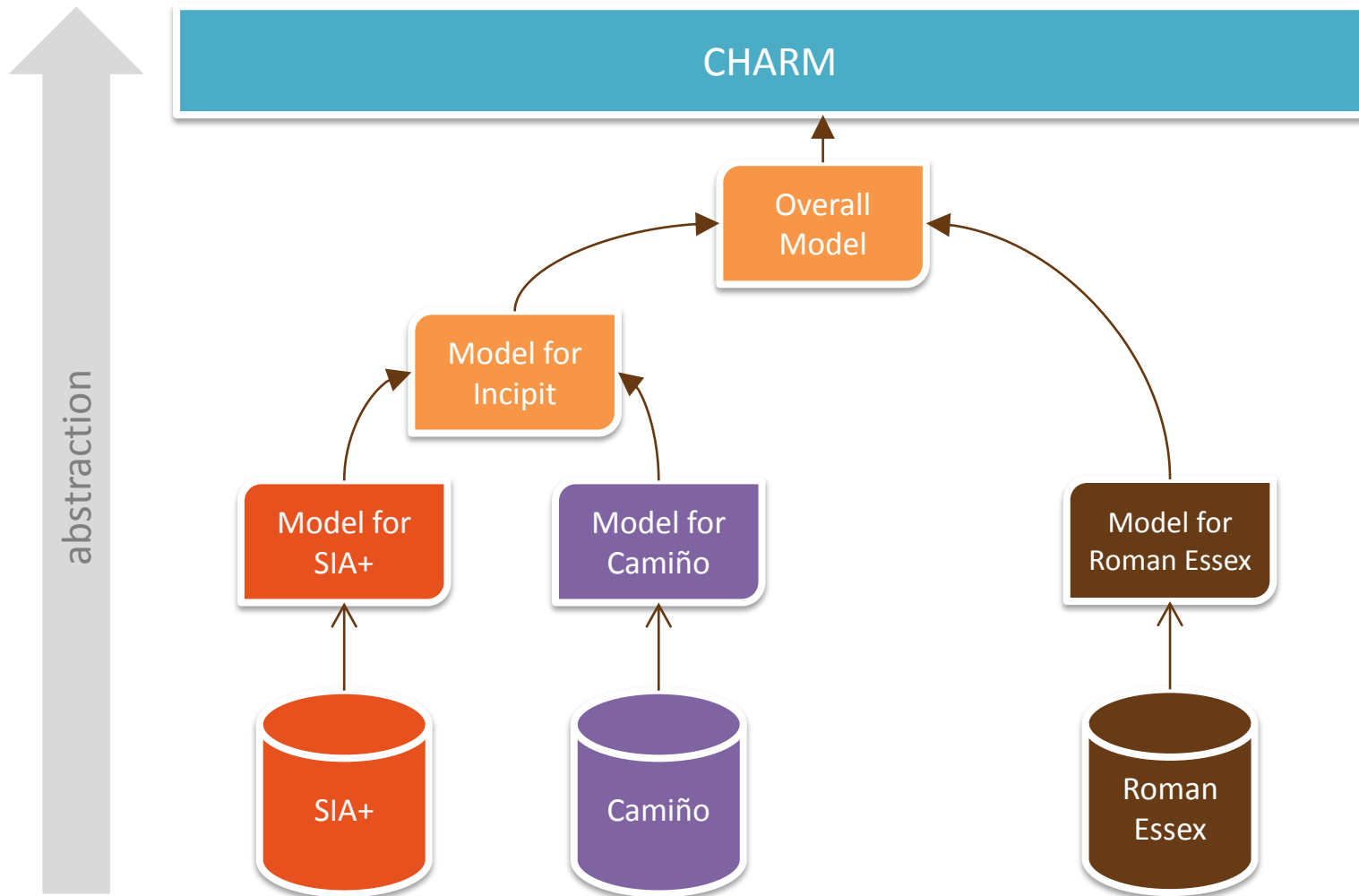
Camiño
Primitivo

- Project-specific information system
- Incipit

Roman
Essex

- Project-specific database
- Available from ADS

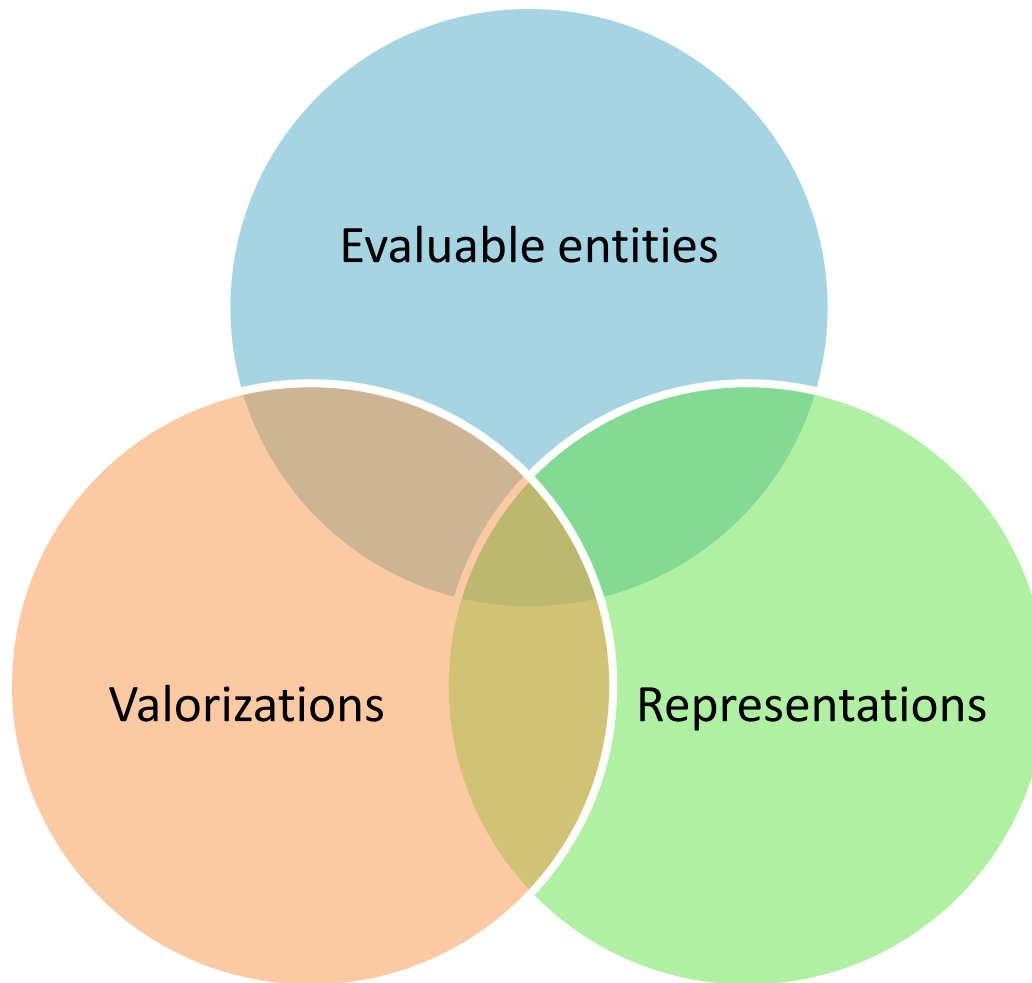
Case study



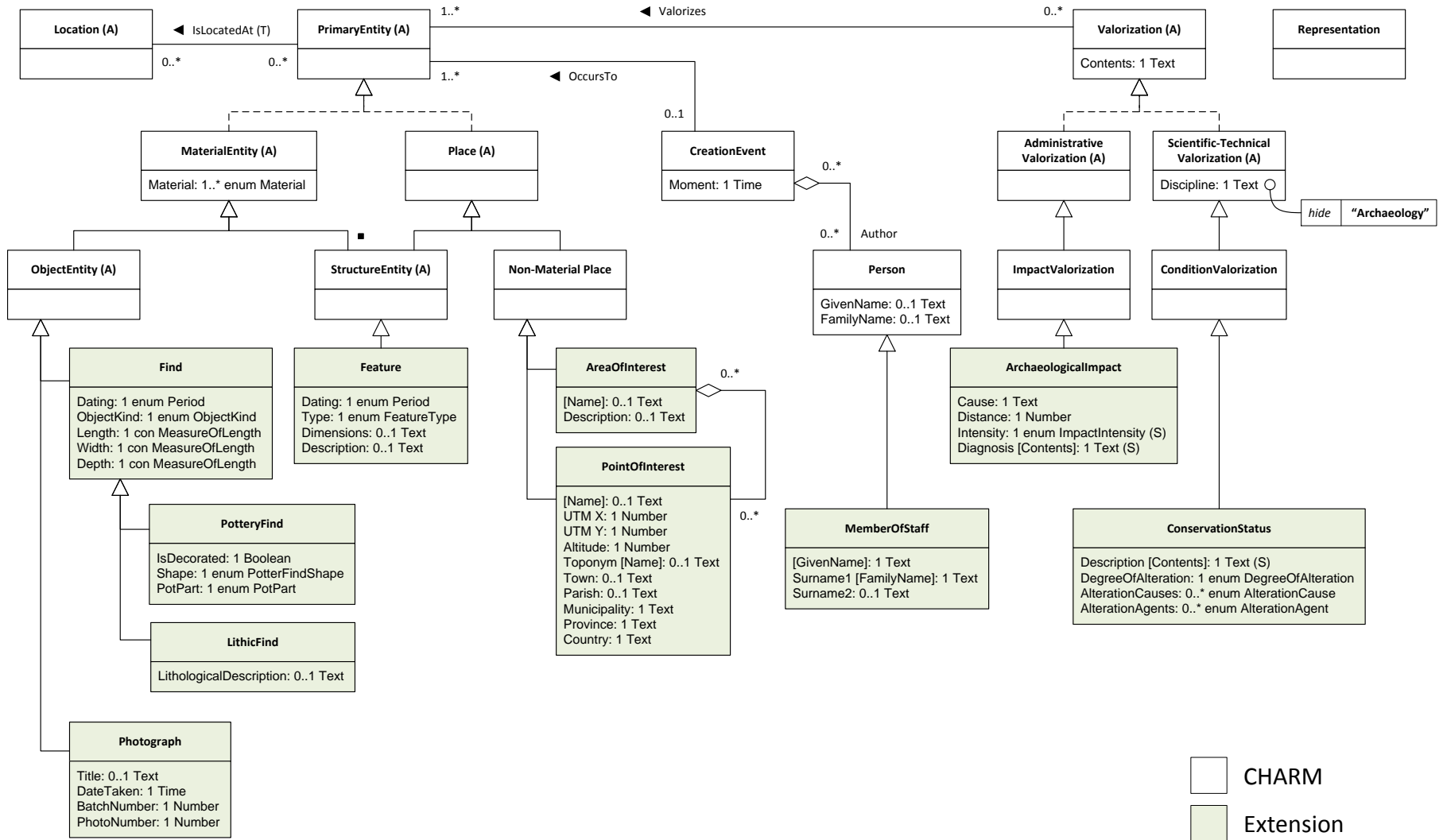
Case study: CHARM

- 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

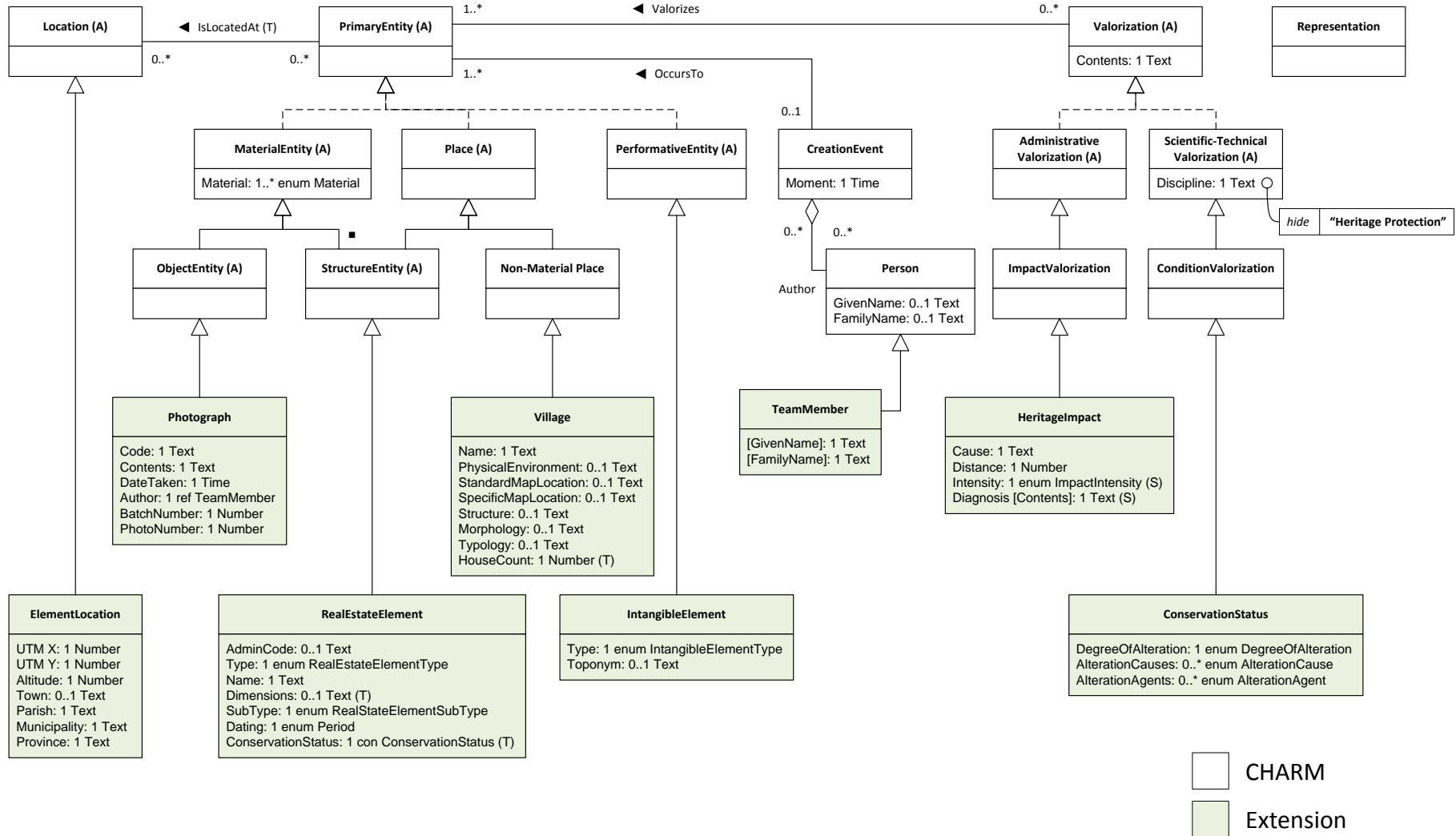
Case study: CHARM



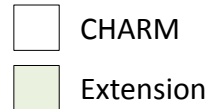
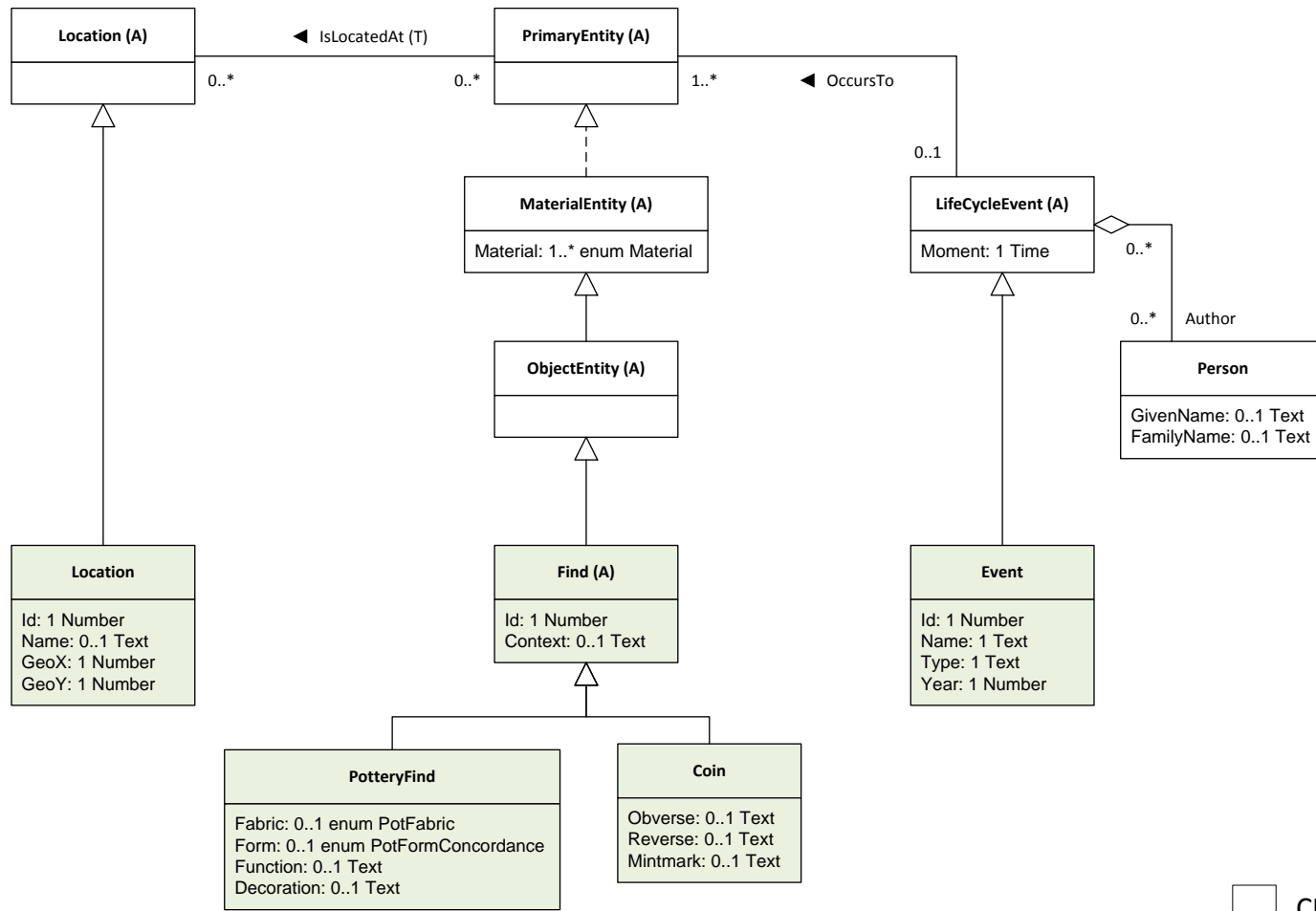
Case study: SIA+ model



Case study: Camiño model



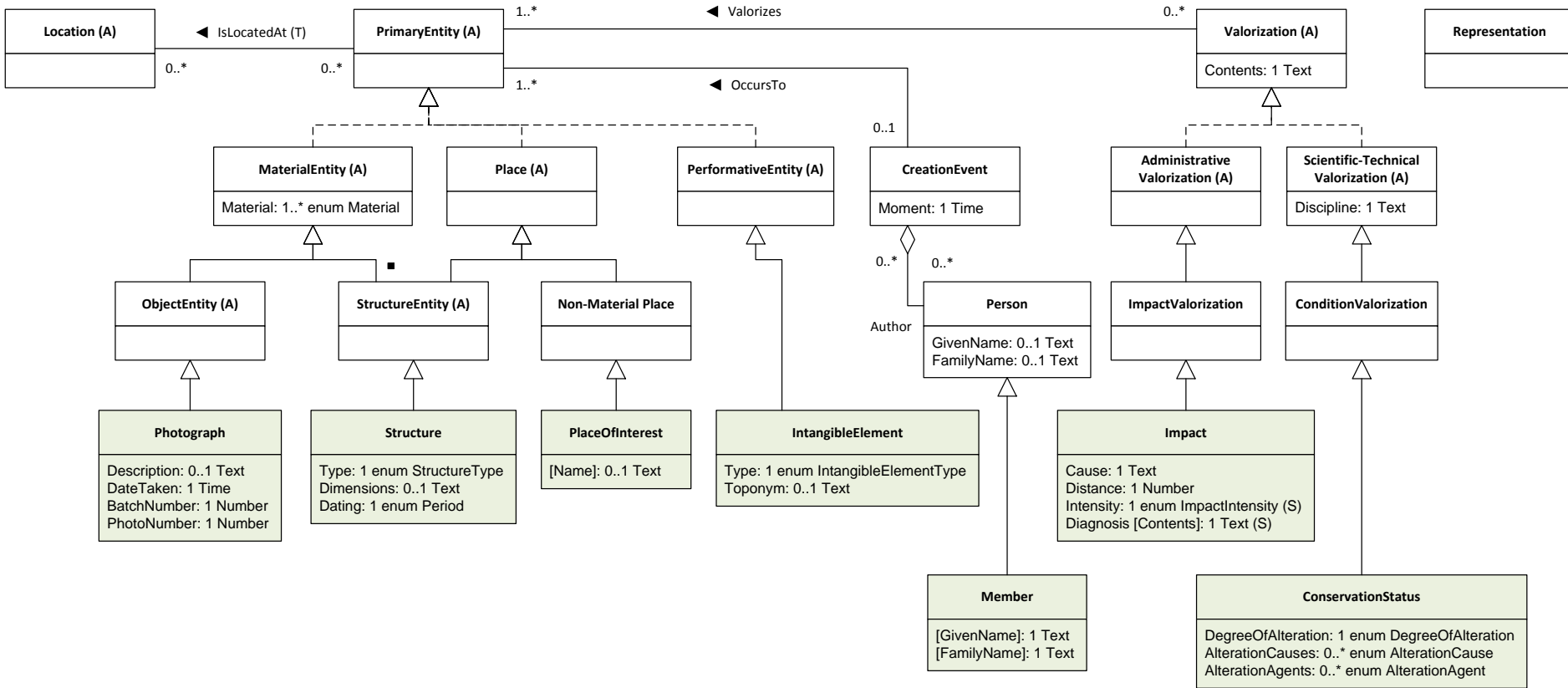
Case study: Roman Essex model



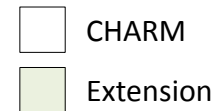
Integrating SIA+ and Camiño models

- 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

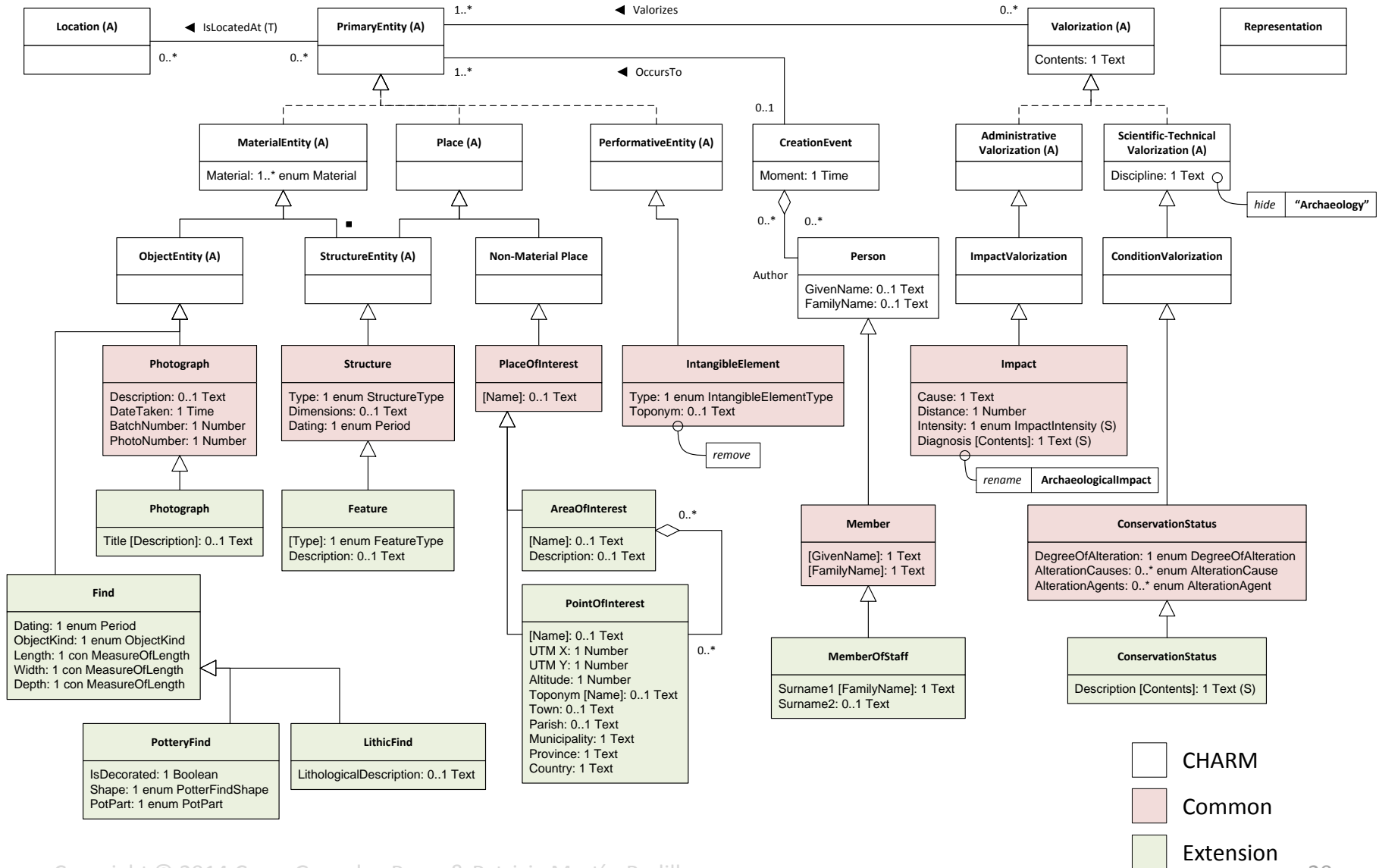
Integrating SIA+ and Camiño models



Shared, organization-wide model for Incipit.



SIA+ model as extension of common



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...*

Camiño

- All data are available, such as AdminCode, Subtype, or ConservationStatus.
- Enum data such as Type are expressed in maximum detail.
- No datasets are interoperable.

Incipit

- 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.

Overall

- Basic data are available, such as Materials and links to related objects.
- The object is still accessible.
- Any CHARM-based dataset is interoperable.

abstraction



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

CHARM^T
Cultural Heritage Abstract Reference Model
www.charminfo.org

ConML
www.conml.org