Towards Semantic Interoperability in Historical Research: Documenting Research Data and Knowledge with *Synthesis*

Pavlos Fafalios, Konstantina Konsolaki, Lida Charami, Kostas Petrakis, Manos Paterakis, Dimitris Angelakis, Yannis Tzitzikas, Chrysoula Bekiari, Martin Doerr

Centre for Cultural Informatics (CCI) & Information Systems Laboratory (ISL),
Institute of Computer Science (ICS), Foundation for Research and Technology (FORTH)

Heraklion, Greece

fafalios@ics.forth.gr

Outline

❑ Introduction

❑ Context

❑ Data Documentation with *Synthesis*

❑ User Interface and Usage Statistics

❑ Conclusion and Lessons Learned

Introduction

Historical Science and Computing

❑ Historical Science

  ➢ A vast area of research concerns the collection, organization, **documentation** and interpretation of information about historical objects, areas or events

❑ Computing in the field has developed enormously over the last years

  ➢ Semantic Web technologies have started playing a significant role
  ➢ Nevertheless, **data management problems** still exist and are still vast and varied

Current practice and related problems

❑ Current practice mostly uses **spreadsheets** or simple **relational databases**

❑ Common **problems**:

➢ Difficulty in collaborative but controlled documentation

➢ Difficulty in representing the details from which the documented relations are inferred

➢ Difficulty in extending the existing data structures on demand

➢ Difficulty of third parties to understand and re-use the data

The *Synthesis* system and its application

❑ Web-based and collaborative system for the **documentation** of data and knowledge in **cultural heritage** and the **humanities**

➢ Can be easily **configured** for different fields

❑ Focus on **semantic interoperability**

➢ Making use of **standards** for data modelling and storage (CIDOC-CRM, RDF)

➢ Aiming at the production of data of **high value** and **longevity**

❑ Application in the **History of Art** field

➢ In the context of a European research project called **RICONTRANS** (ERC)

Context

Context: the **RICONTRANS** project

https://ricontrans-project.eu/

❑ RICONTRANS *Visual Culture, Piety and Propaganda: Transfer and Reception of Russian Religious Art in the Balkans and the Eastern Mediterranean (16th – early 20th c.)*

- ➢ ERC Consolidator Grant (**May 2019 – April 2024**)
- ➢ Field: **Art History**
- ➢ Principal Investigator: **Dr. Yuliana Boycheva** (Institute of Mediterranean studies, FORTH)
- ➢ Research teams in **Greece**, **Bulgaria**, **Serbia**, **Romania**, **Russia**

❑ The Research Focus

- ➢ The **Russian religious artefacts** brought from Russia to the Balkans (16th – early 20th c.)

RICONTRANS – The data

❑ Information and data about:

➢ Art objects (icons, triptychs, crosses, …)

➢ Object transfers (from/to, purpose, …)

➢ Historical figures (archbishops, priests, saints, …)

➢ Events (archbishop ordination, church erection, …)

➢ Locations (cities, villages, monasteries, churches, museums, ...)

❑ Information Sources

➢ Archival sources

➢ Oral history sources

➢ Old books / newspapers

➢ Web sources

❑ Research data

➢ Findings, comments, …

❑ Digital files

➢ Images, scans, docs, …

Data Documentation with *Synthesis*

System overview, entity types, data model, data transformation

Synthesis – System overview

❑ Web-based system for the **collaborative documentation** of data and knowledge in cultural heritage and (digital) humanities
  ➢ Multilingual, supports versioning

❑ Utilizes **XML technology** and a multi-layer architecture
  ➢ **Flexibility** and **extensibility** (in terms of data structures and data types)
  ➢ **Sustainability** (XML documents readable by both humans and machines)
  ➢ Database: **eXist-db** (native XML database)

Synthesis – Entity Types

❑ Users create and document **entities** organized in **entity types**

➢ Objects
➢ Object Transfers
➢ Routes

Objects and Transfers

➢ Source Passages
➢ Collection of Source Passages
➢ Researcher Comments

Passages and Comments

➢ Archival Sources
➢ Books
➢ Newspapers and Periodicals / Reviews
➢ Oral History Sources
➢ Web Sources

➢ Bibliography

Sources

➢ Historical Figures
➢ Collections
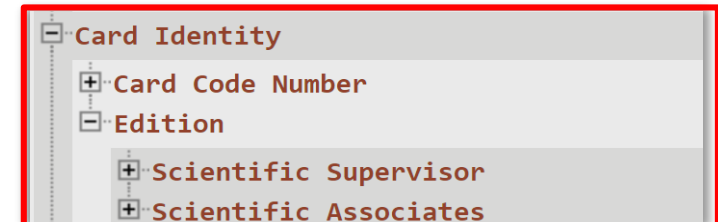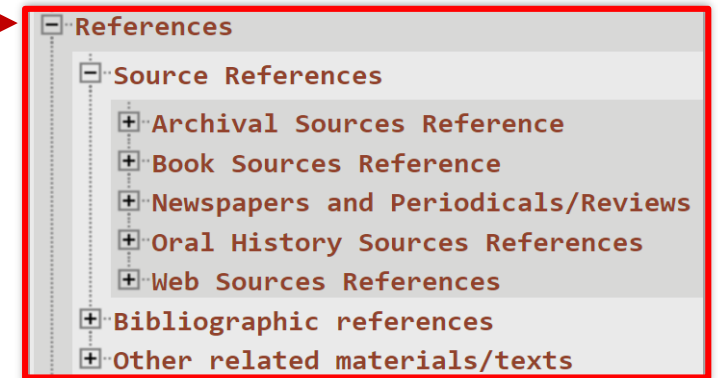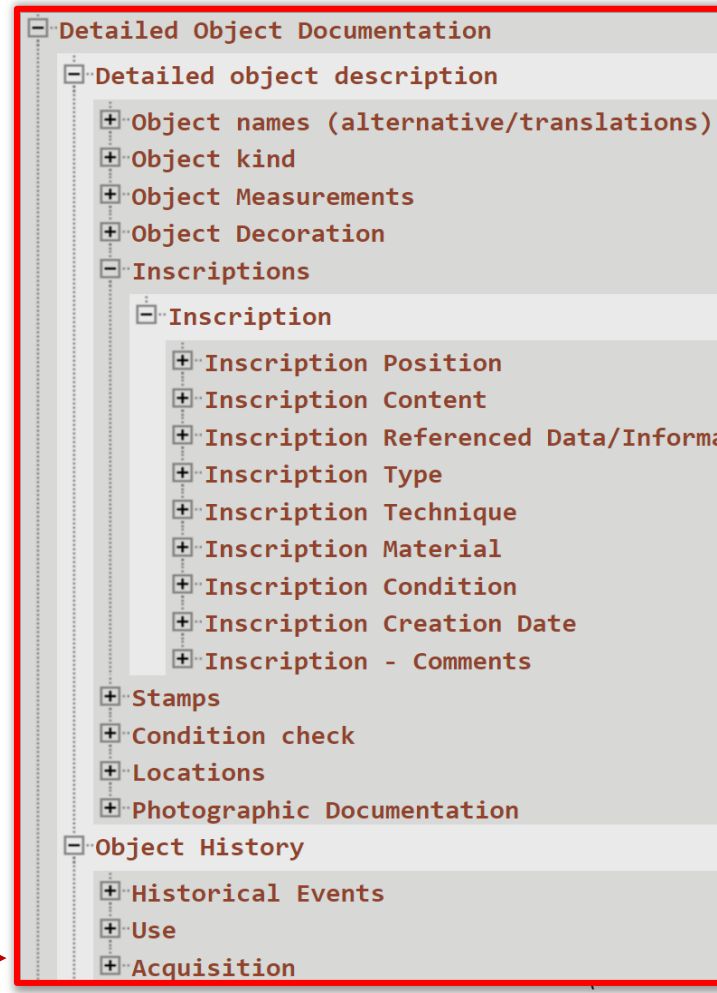➢ Events
➢ Locations
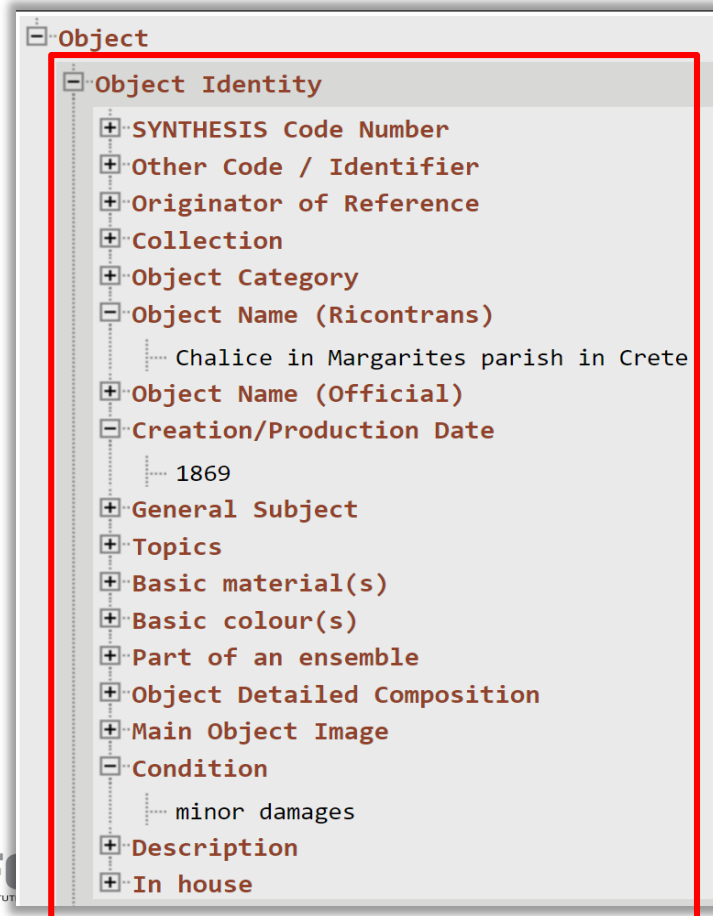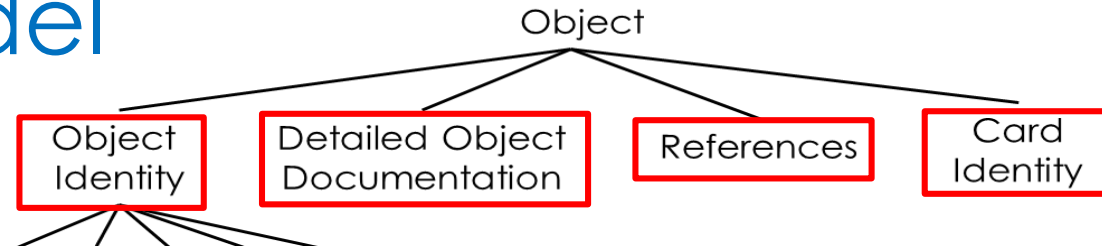➢ Organizations
➢ Digital Objects

Related entities

Synthesis – Data Model

❑ Specially-designed for the domain of **History of Art** and the particular needs of the **RICONTRANS** project

➢ Focus on **semantic interoperability**

❖ Linking each element of the data model to a **target (domain) ontology** (*more later*)

❖ Linking terms to controlled (shared) **vocabularies** or **thesauri of terms**

❖ Enabling the inclusion of rich **metadata** about the documented data

❑ Each **entity type** has its own data structure (**documentation schema**)

➢ A documentation schema is **XML-based**, containing a set of fields organized in an **hierarchical (tree-like) structure**

➢ The leaves in this tree-like structure are the **documentation fields** that are to be filled by the users
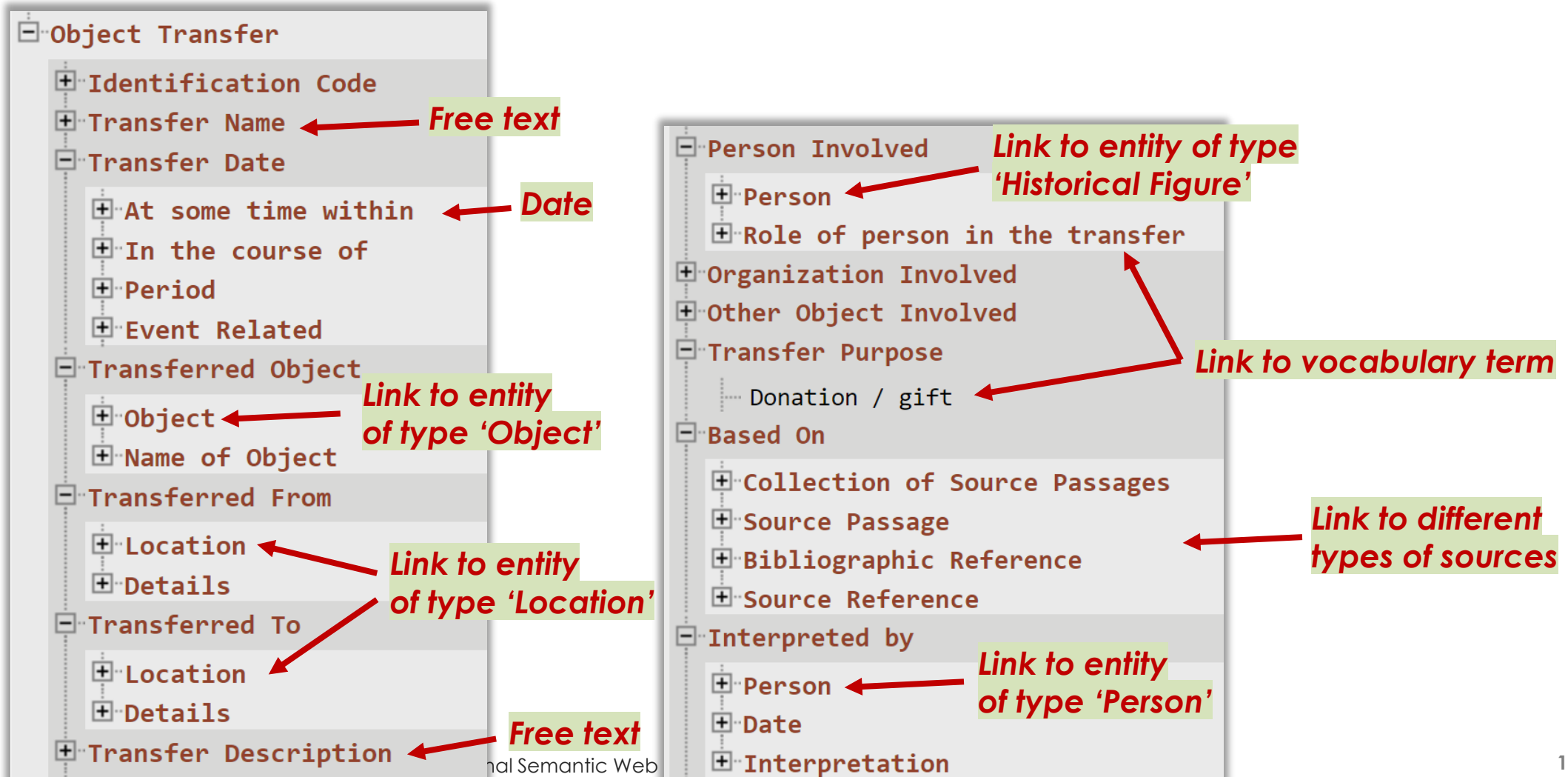
Synthesis – Data Model

Entity type: Object

Object
- Object Identity
- Detailed Object Documentation
- References
- Card Identity

Object
- **Object Identity**
  - SYNTHESIS Code Number
  - Other Code / Identifier
  - Originator of Reference
  - Collection
  - Object Category
  - Object Name (Ricontrans)
    - Chalice in Margarites parish in Crete
  - Object Name (Official)
  - Creation/Production Date
    - 1869
  - General Subject
  - Topics
  - Basic material(s)
  - Basic colour(s)
  - Part of an ensemble
  - Object Detailed Composition
  - Main Object Image
  - Condition
    - minor damages
  - Description
  - In house

Detailed Object Documentation
- **Detailed object description**
  - Object names (alternative/translations)
  - Object kind
  - Object Measurements
  - Object Decoration
  - Inscriptions
    - Inscription
      - Inscription Position
      - Inscription Content
      - Inscription Referenced Data/Information
      - Inscription Type
      - Inscription Technique
      - Inscription Material
      - Inscription Condition
      - Inscription Creation Date
      - Inscription - Comments
  - Stamps
  - Condition check
  - Locations
  - Photographic Documentation
- **Object History**
  - Historical Events
  - Use
  - Acquisition

References
- Source References
  - Archival Sources Reference
  - Book Sources Reference
  - Newspapers and Periodicals/Reviews
  - Oral History Sources References
  - Web Sources References
- Bibliographic references
- Other related materials/texts

Card Identity
- Card Code Number
- Edition
  - Scientific Supervisor
  - Scientific Associates

14

Synthesis – Data Model

Entity type: Object Transfer

- Object Transfer
  - Identification Code
  - Transfer Name ← **Free text**
  - Transfer Date
    - At some time within ← **Date**
    - In the course of
    - Period
    - Event Related
  - Transferred Object
    - Object ← **Link to entity of type 'Object'**
    - Name of Object
  - Transferred From
    - Location ← **Link to entity of type 'Location'**
    - Details
  - Transferred To
    - Location
    - Details
  - Transfer Description ← **Free text**

- Person Involved
  - Person ← **Link to entity of type 'Historical Figure'**
  - Role of person in the transfer ← **Link to vocabulary term**
- Organization Involved
- Other Object Involved
- Transfer Purpose
  - Donation / gift
- Based On
  - Collection of Source Passages
  - Source Passage
  - Bibliographic Reference ← **Link to different types of sources**
  - Source Reference
- Interpreted by
  - Person ← **Link to entity of type 'Person'**
  - Date
  - Interpretation

Synthesis – Types of documentation fields

- ❑ Link to entity
- ❑ Link to (**static** or **dynamic**) vocabulary term
- ❑ Link to **thesaurus of terms** (managed through **THEMAS**[1])

- ❑ Unformatted free text
- ❑ Formatted free text
- ❑ Number

- ❑ Time expression (date range in an accepted format)

- ❑ Location coordinates (**point** or **polygon**)
- ❑ Location ID (**TGN** or **Geonames**)

- ❑ Digital file(s)

[1]THEMAS: https://www.ics.forth.gr/isl/themas-thesaurus-management-system

Synthesis – Data transformation

❑ *Synthesis* has embedded processes for transforming the data stored in the XML documents to an **ontology-based RDF dataset** (a **Knowledge Base**)

➤ It **decouples** data entry *(made by the research team)* from the ontology-based integration and creation of the KB *(supported by data engineers)*

Documentation

Modeling

CIDOC CRM
ISO 21127:2014

Transformation

Exploration

Domain expert

Domain expert   Data engineer

Domain expert
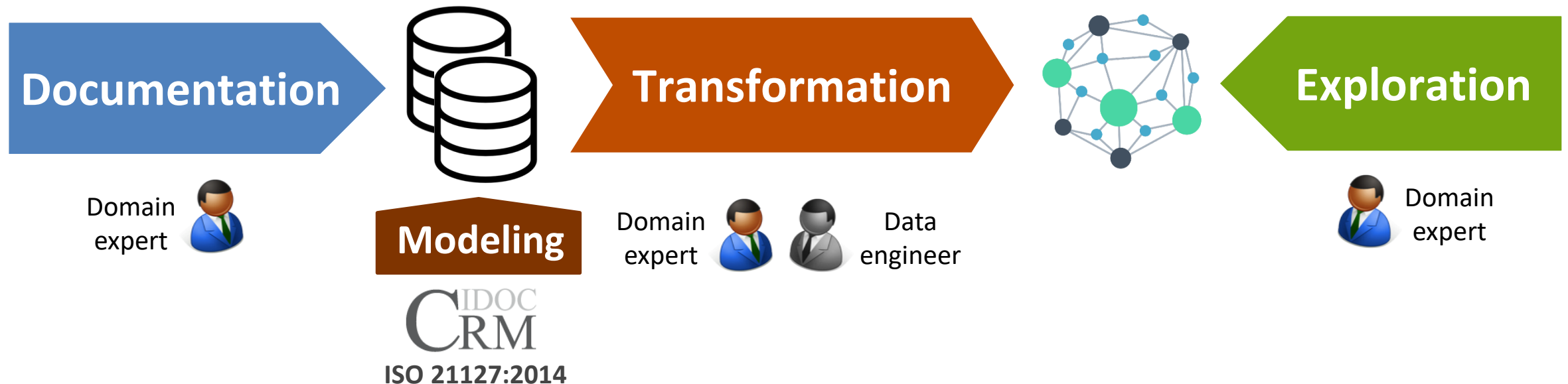
Synthesis – Data transformation

Part of object's documentation schema
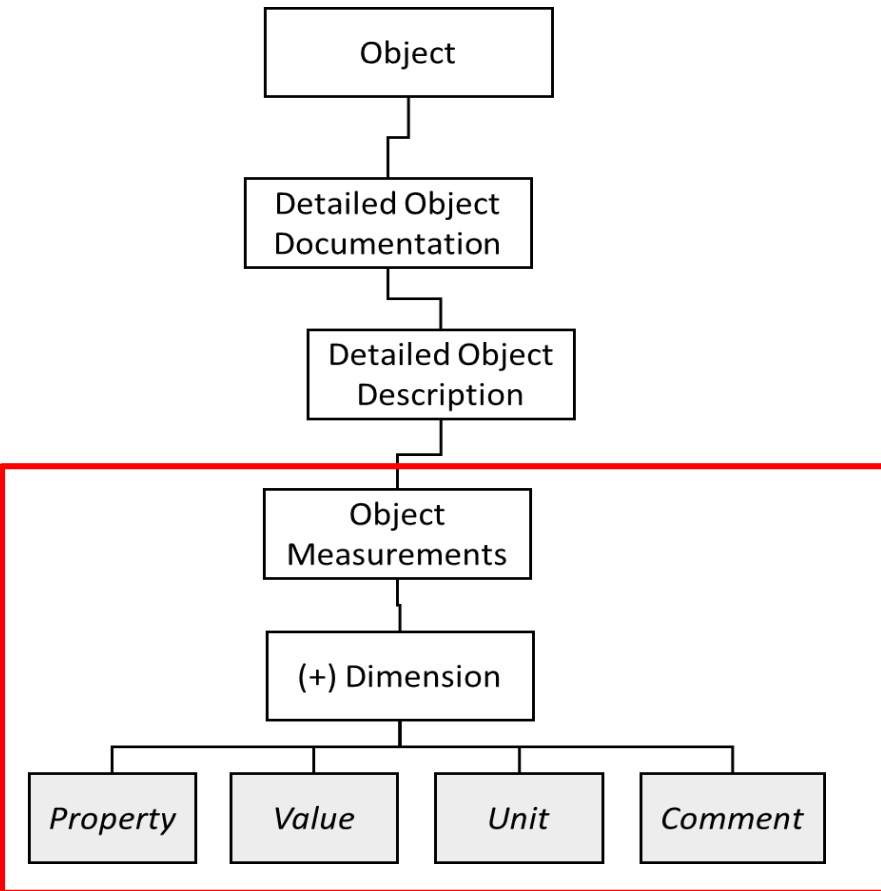
Modelling of object measurements in CIDOC-CRM
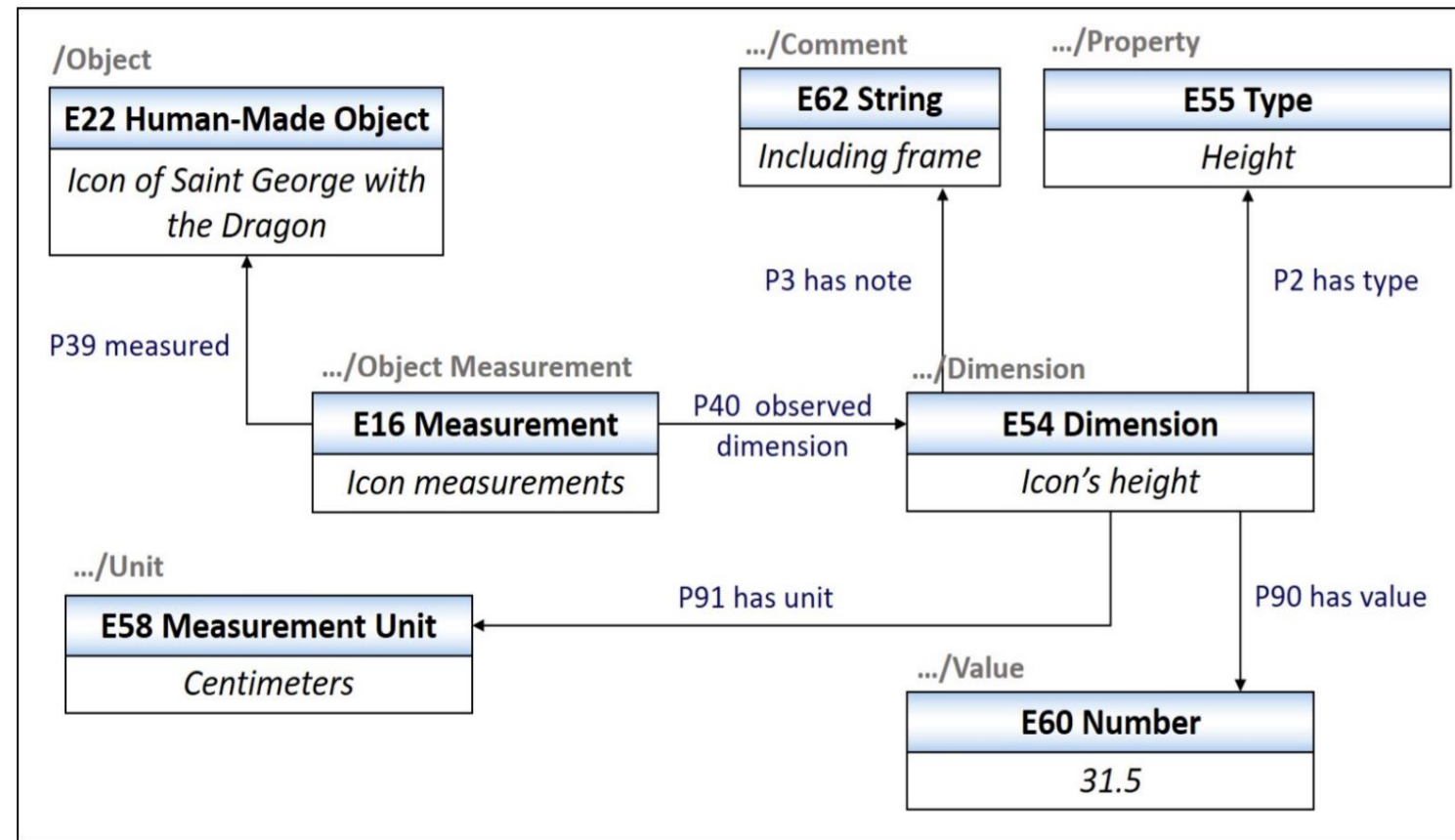
Synthesis – Overall philosophy

❑ Why NOT creating a Knowledge Base from the beginning?

1. We regard as very different a KB of facts believed together as true, versus **managing and consolidating the knowledge acquisition process** of a large research team

2. We consider a KB as an ideal tool for integrating the **latest stage of knowledge** acquired through diverse processes

3. It allows the straightforward production of different KB versions for **different ontologies**, or **different versions** of the same ontology

   ❖ We just need to create and maintain the corresponding **schema mappings**

User Interface and Usage Statistics

The Web Interface of Synthesis

Entities management

Entity types

Table of documented entities of selected entity type

VISUAL CULTURE, PIETY
AND PROPAG[...]
TRANSFER AN[...]
RUSSIAN REL[...]
AND THE EAS[...]
(16TH – EARL[...]

Pavlos Fafalios

RICONTRANS

Objects and Transfers

Objects
Object Transfers
Routes

Sources

Archival Sources
Books
Newspapers and
Periodicals/Reviews
Oral History Sources
Web Sources

Related Bibliography

Bibliography

Passages and Comments

Source Passages
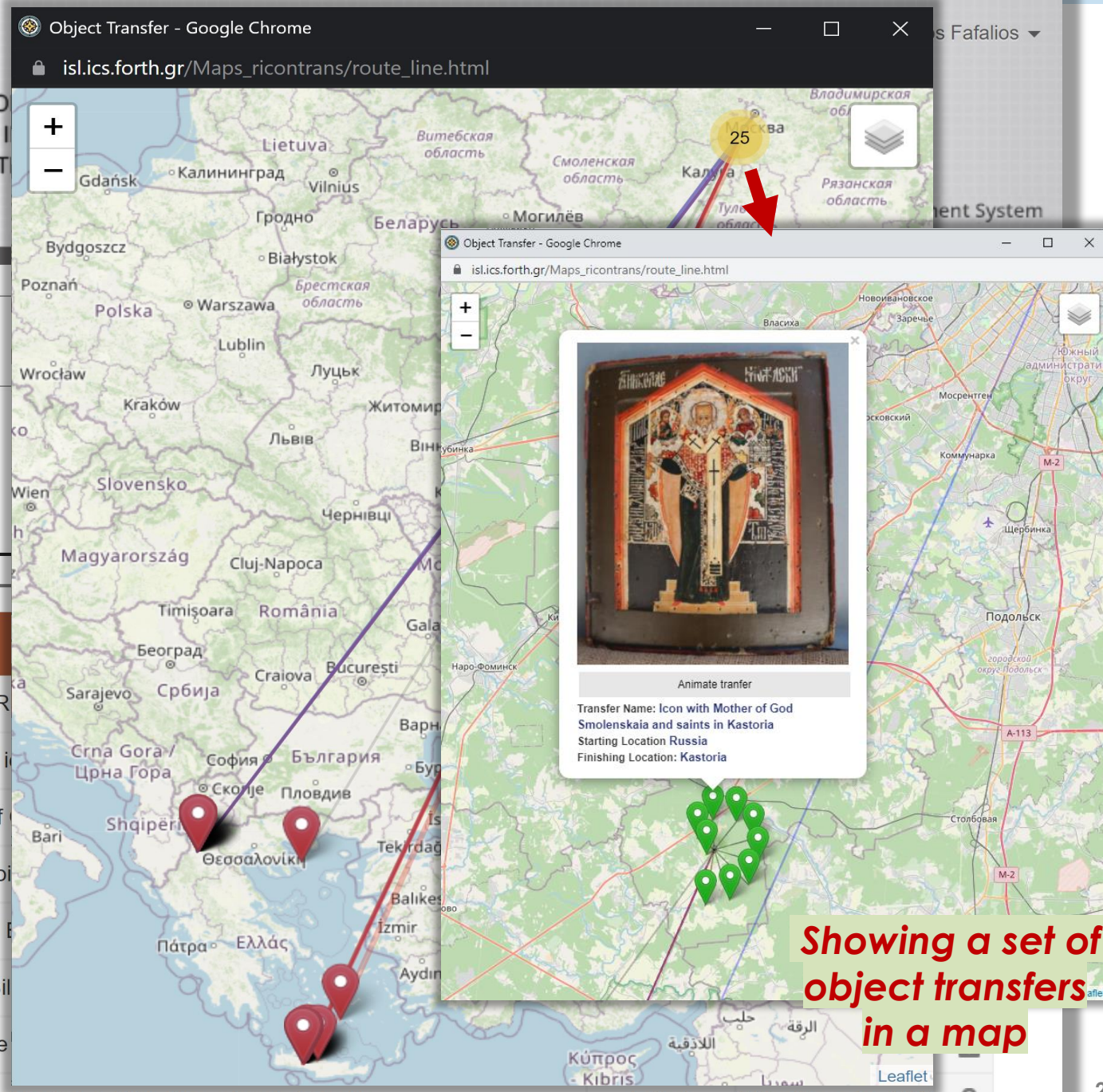Collection of Source
Passages
Researcher Comments

⊕ 👁 ✏ 📍 More ▾

Object Transfers

Showing: All

↻ **Filter Table**

FeXML - Google Chrome

🔒 isl.ics.forth.gr/FeXML_ricontrans/Index?type=ObjectTransfers&action=view&la...

FeXML

✏ 🗔 .+ .— ⋮+ ⋮— 📄   **Record: ObjectTransfers/10**

⊟ Object Transfer
  ⊟ Identification Code
     ├ https://ricontrans-project.eu/ObjectTransfers/10
  ⊟ Transfer Name
     ├ Icons from Russia to Santorini (Megalo Horio)
  ⊟ Transfer Date
     ⊞ At some time within
     ⊟ In the course of
        ⊟ Date
           ├ 1900
     ⊞ Period
     ⊟ Event Related
        ├ Donation
  ⊟ Transferred Object
     ⊟ Object
        ├ .../Object/19, Icons from Russia to Santorini (Megalo Horio)
     ⊞ Name of Object
  ⊟ Transferred From
     ⊟ Location
        ├ .../Location/22, Russia
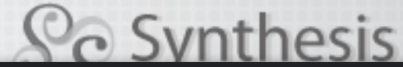   ⊞ Details
  ⊟ Transferred To

*Viewing the
documentation card of
an object transfer*

RICONTRANS

VISUAL CULTURE, PIETY
AND PROPAGANDA:
TRANSFER AND RECEPTION
RUSSIAN RELIGIOUS ART IN
AND THE EASTERN MEDITE
(16TH – EARLY 20TH C.)

System

Objects and Transfers

Objects
Object Transfers
Routes

Sources

Archival Sources
Books
Newspapers and
Periodicals/Reviews
Oral History Sources
Web Sources

Related Bibliography

Bibliography

Passages and Comments

Source Passages
Collection of Source
Passages

⊕ ◉ ☑ 📍 More ▾

Objects

Showing: All

Filter Table tatarna

☑	**Object name (Ricontrans)** ⬍	Ori... R...
☑	Manuscript of the three Liturgies	.../Orga Monaster (Pana Euryta
☑	Icon of the Vision of St. Sergius of Radonezh	.../Orga Monaster (Pana Euryta
☑	Icon of Christ Pantocrator	.../Organization/277, Monastery of Virgin Mary (Panagia) Tatarna,

| .../Location/288, Monastery of Virgin Mary Tatarna, | maltezak | unpublished Object/355 |

Show - Google Chrome

🔒 isl.ics.forth.gr/Maps_ricontrans/show_coords.html?mode=points

Object Name (Ricontrans): Παναγία η Ταταρνιώτισσα
Current Location: Monastery of Virgin Mary (Panagia) Tatarna, Eurytania, Greece
Creation or Production Date: 1st half 17th century
Entity ID: 357

Showing a set of objects in a map

23

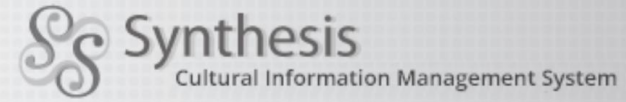*Showing a set of object transfers in a map*

Management of vocabularies

VISUAL CULTURE, PIETY AND PROPAGANDA: TRANSFER AND RECEPTION OF RUSSIAN RELIGIOUS ART IN THE BALKANS AND THE EASTERN MEDITERRANEAN (16TH – EARLY 20TH C.)

Synthesis
Cultural Information Management System

❓▾   👤 Διαχειριστής Admin ▾

Objects and Transfers

Objects
Object Transfers
Routes

Sources

Archival Sources
Books
Newspapers and Periodicals/Reviews
Oral History Sources
Web Sources

Related Bibliography

Bibliography

Passages and Comments

Source Passages

Vocabularies - Administration

❯ Objects

Vocabulary of Code Category
Vocabulary of Object Category
Vocabulary of Subjects
Vocabulary of Topics
Vocabulary of Inscription/ Stamp M
Vocabulary of Materials
Vocabulary of Colours
Vocabulary of Part of an ensemble
Vocabulary of Kind of Composition
Vocabulary of Category of Seconda
Vocabulary of Condition

⊕ 📝 More ▾

Administration - Vocabulary of Object Category

↻ **Filter Table** [   ]   Showing 10 ▾ entries

☐	Term ⬍
☐	objects in church or monastery sacristies
☐	objects in churches or monasteries for ceremonies and rituals
☐	objects in museum collections (for display / exhibit purposes)
☐	objects in private collections (for display / exhibit purposes)
☐	objects of family heirloom
☐	objects of personal / private devotion
☐	objects in church or monastery museum collections (for display / exhibit purposes)

FORTH INSTITUTE OF COMPUTER SCIENCE · isl

Usage Statistics

❑ ~40 users (5 countries)

❑ Current number of documented entities (as of Oct-2021):

- ➢ Objects: 1,270
- ➢ Object transfers: 374
- ➢ Routes: 93

- ➢ Archival sources: 177
- ➢ Books: 45
- ➢ Newspapers and Periodicals/Reviews: 112
- ➢ Oral History Sources: 3
- ➢ Web Sources: 60

- ➢ Bibliography: 328

- ➢ Source Passages: 572
- ➢ Collection of Source Passages: 7
- ➢ Researcher Comments: 0

- ➢ Historical Figures: 208
- ➢ Collections: 169
- ➢ Events: 33
- ➢ Locations: 517
- ➢ Persons: 101
- ➢ Organizations: 420

- ➢ Digital Objects: 1,220

Conclusion and Lessons Learned

Conclusion

❑ Data documentation and management with **_Synthesis_**

➢ Web-based and collaborative

➢ Focus on semantic interoperability (compatibility with **CIDOC-CRM**)

➢ **Aim:** production of data with **high value**, **long term validity** and **longevity**

❑ Application in a large-scale research project in History of Art (RICONTRANS)

➢ Providing full-fledged support for the complete knowledge production life-cycle in historical research

❑ Configurable for use in other fields!

➢ We just need to specify the entity types and their documentation fields

Lessons Learned

❑ Finding the best trade-off between **documentation richness** and **usability** is challenging

➢ **Example:** It is much simpler to record the dimensions of an object in a single text field (e.g., "15cm x 20cm") than breaking it to 3 fields (property, value, unit)

➢ However, the former makes very difficult, if not impossible, to make comparisons

❑ Controlling the **dynamic vocabularies** is difficult

➢ Creation of new terms that already exist with different names

➢ Curation is needed (laborious and time consuming)

❖ *How could we support a better management of the dynamic vocabularies?*

Thank you!

Pavlos Fafalios

fafalios@ics.forth.gr

Centre for Cultural Informatics
FORTH-ICS

ISWC 2021