Historical Research meets Semantic Interoperability: The Documentation System SYNTHESIS and its Application in Art History Research

Pavlos Fafalios

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 roles, interface, data model, functionalities, usage statistics
- Data transformation to a (CIDOC-CRM based) semantic network
- Conclusion
  - + Lessons learned and future work

Introduction

Historical Science and Computing

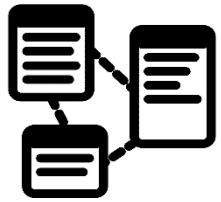
- Historical Science
  - A vast area of research concerns the **collection**, **documentation** and **interpretation** of information about **cultural artefacts** and related **evidence**

- Computing in the field has developed enormously over the last years
  - Nevertheless, **data management problems** still exist and are still vast and varied

https://www.pngkey.com/maxpic/u2e6w7r5r5r5y3e6/

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

Our approach: the ***Synthesis*** system

- Web-based and collaborative system for the **documentation** of data and knowledge in **cultural heritage** and the **humanities**
  - Can be easily **configured** for specific fields!

- Focus on **semantic interoperability**
  - Making use of **standards** for data modelling and storage (**CIDOC-CRM**, **RDF**)
  - Aiming at the production of **sustainable data** of **high value**

- Application in **History of Art**
  - In the context of a large 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.)
  - What are the paths and the mediums of their transfers as well as the moving factors?
  - What are the aesthetic, ideological, political and social factors that shaped the context of their reception in the various social and cultural environments?
  - What is their influence on the visual culture of the host societies?

RICONTRANS – The data

- Information and data about:
  - **Art objects** (icons, triptychs, crosses, …)
  - **Object transfers** (from/to, purpose, …)
  - Other **relevant entities**:
    - **Historical figures** (archbishops, priests, saints, …)
    - **Events** (archbishop ordination, church erection, …)
    - **Locations** (cities, villages, monasteries, churches, museums, …)

- Primary Sources
  - Archival sources
  - Oral history sources
  - Old books / newspapers

- Secondary Sources
  - Bibliography

- Research data
  - Findings, comments, …

- Digital files
  - Images, scans, docs, …

Data Documentation with ***Synthesis***

System overview, user roles, interface, data model, functionalities, usage statistics

Synthesis – System overview

- Web-based system for the **collaborative documentation** of data and knowledge in cultural heritage and (digital) humanities
  - Configurable, 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)

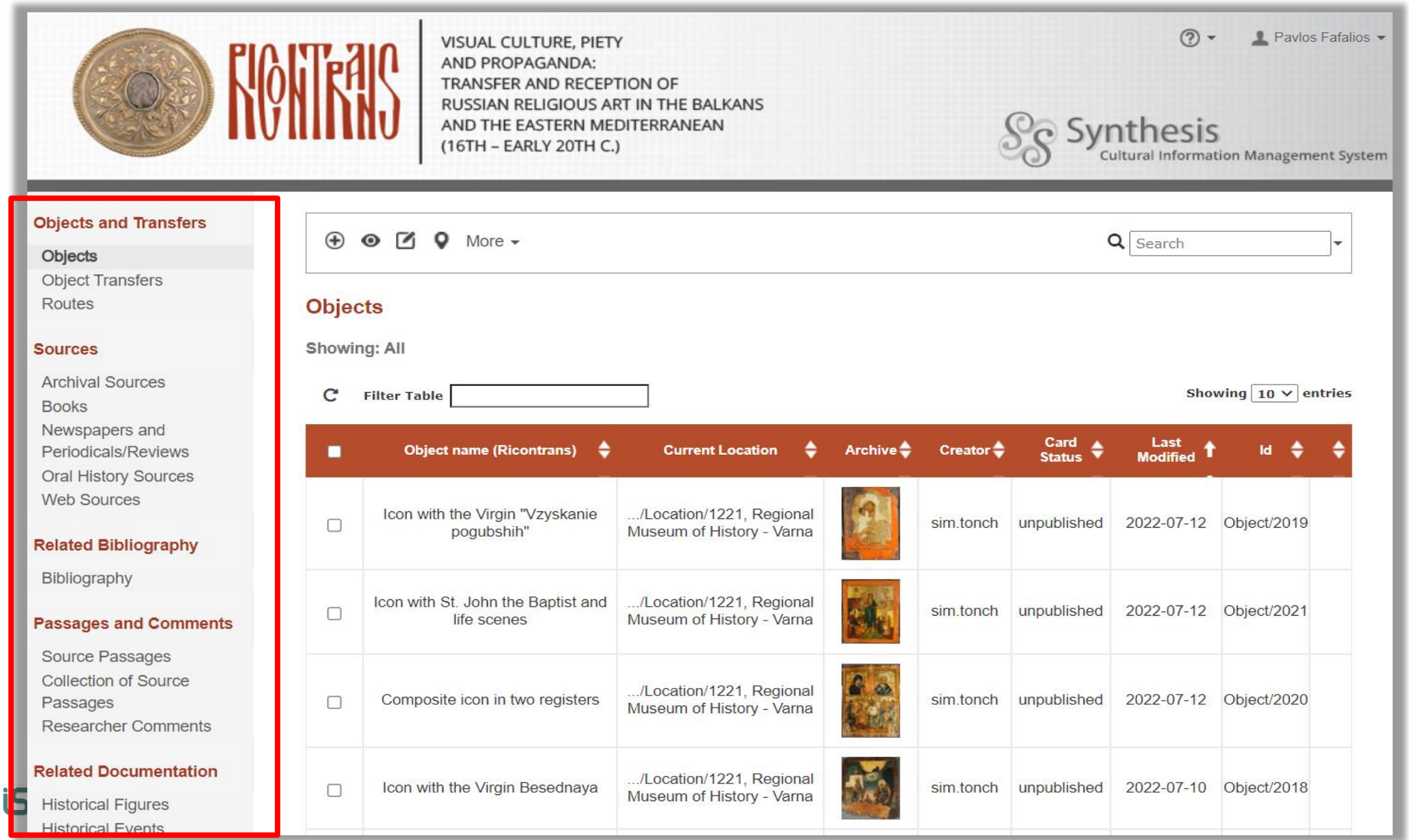
Synthesis – Users Roles

- User roles
  - **System administrator** – can create new 'organizations' (groups of users)
  - **Organization administrator** – can create editors and guests for a particular organization
  - **Editor** – can create and document entities for a specific organization
  - **Guest** – can only view the documented entities of a specific organization
- The management of rights can be easily adjusted for any specific need
  - E.g., user with edit access to all documented entities (for making corrections, etc.)

- **Editors** create and document **entities** organized in **entity types**
  - Example: the **entity** "*Brass icon depicting the Three Hierarchs (Benaki Museum)*" is of **entity type** "*Object*"

Synthesis – Web Interface

Entity types



Synthesis – Entity Types

- The **entity types** in the case of **RICONTRANS**

Objects and Transfers

- Objects
- Object Transfers
- Routes

Related Bibliography

- Bibliography

Related Documentation

- Historical Figures
- Historical Events
- Collections
- Locations
- Persons
- Organizations

Sources

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

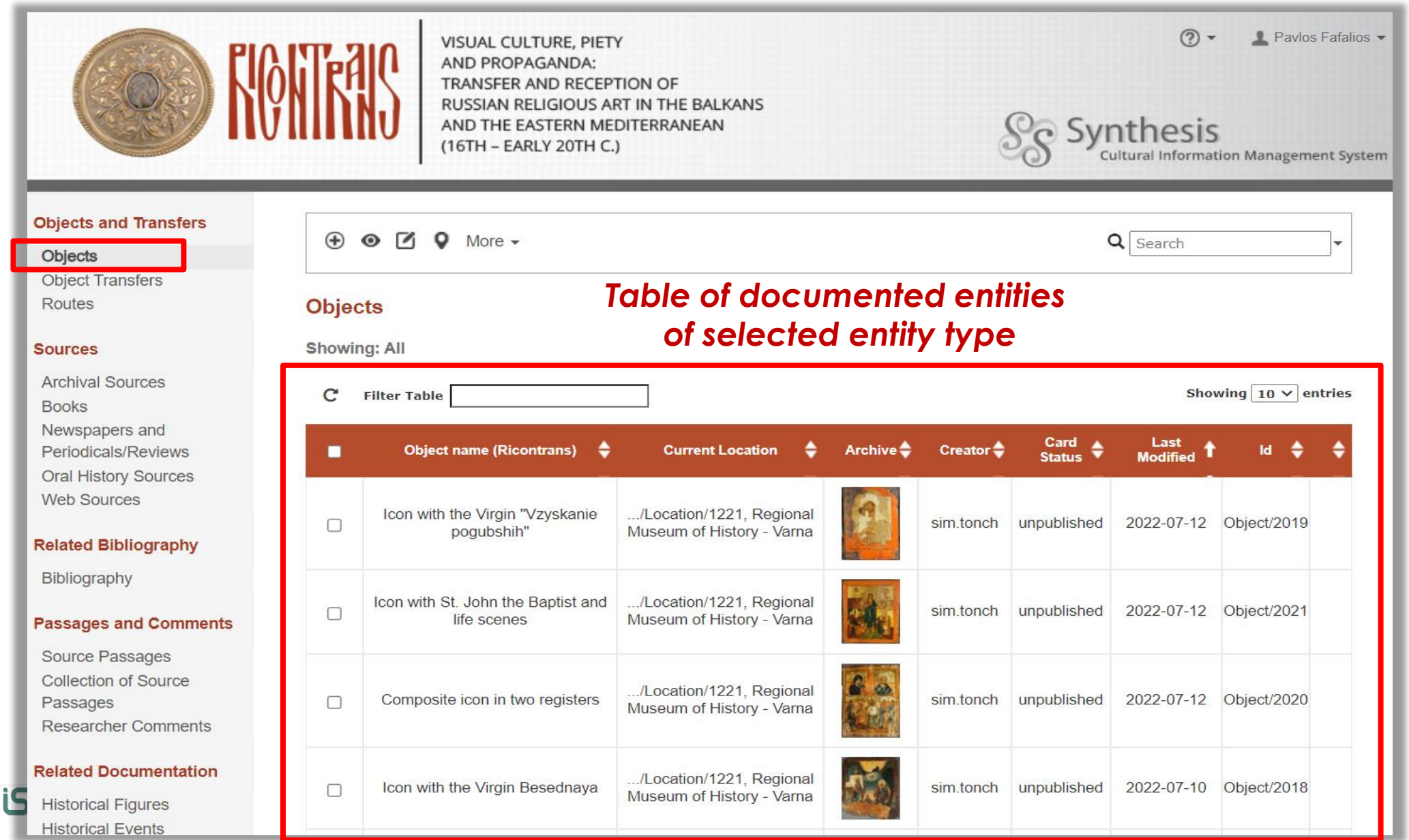
Passages and Comments

- Source Passages
- Collection of Source Passages
- Researcher Comments

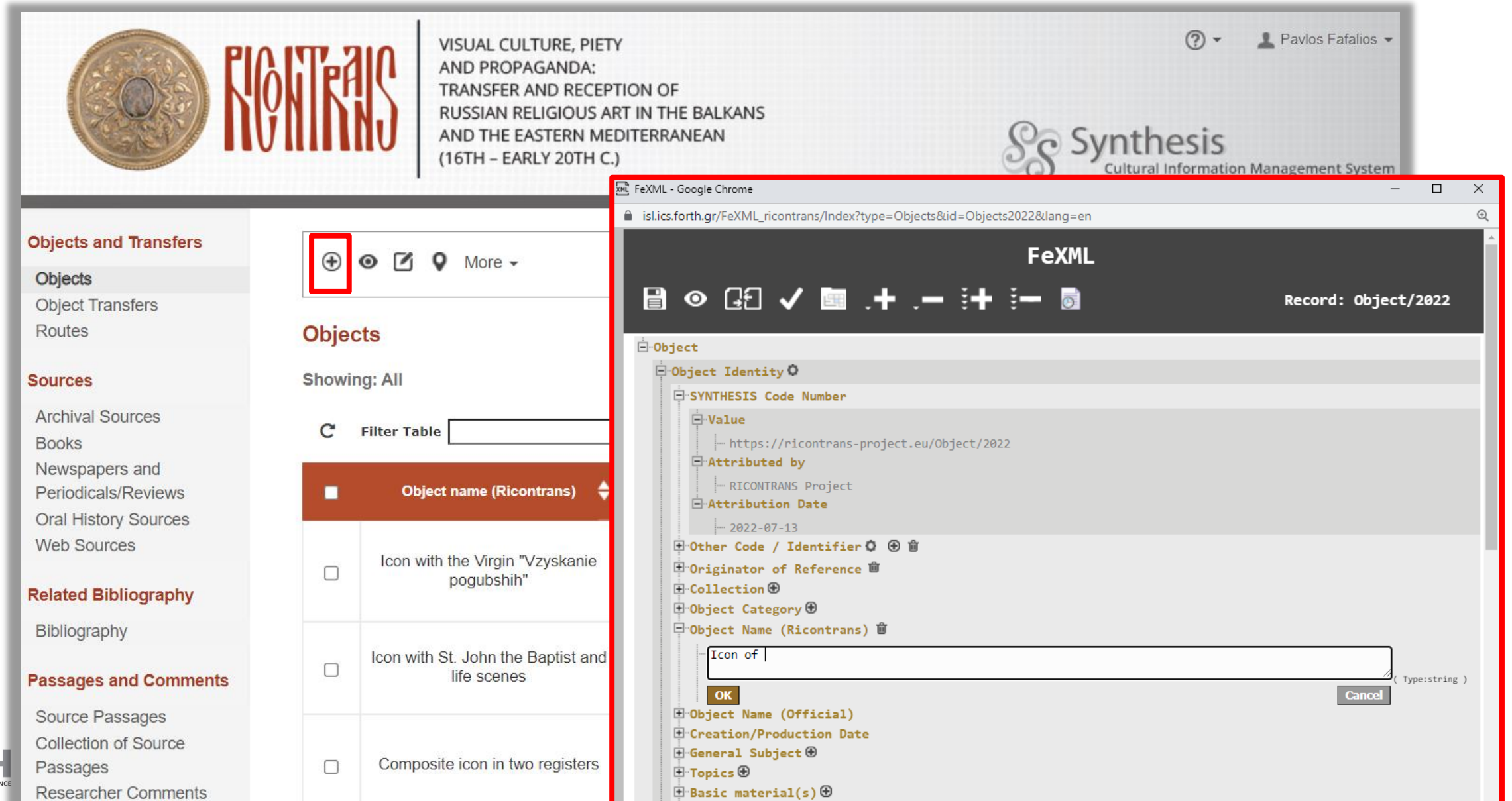
Presentation

- Digital Objects
- Information Texts

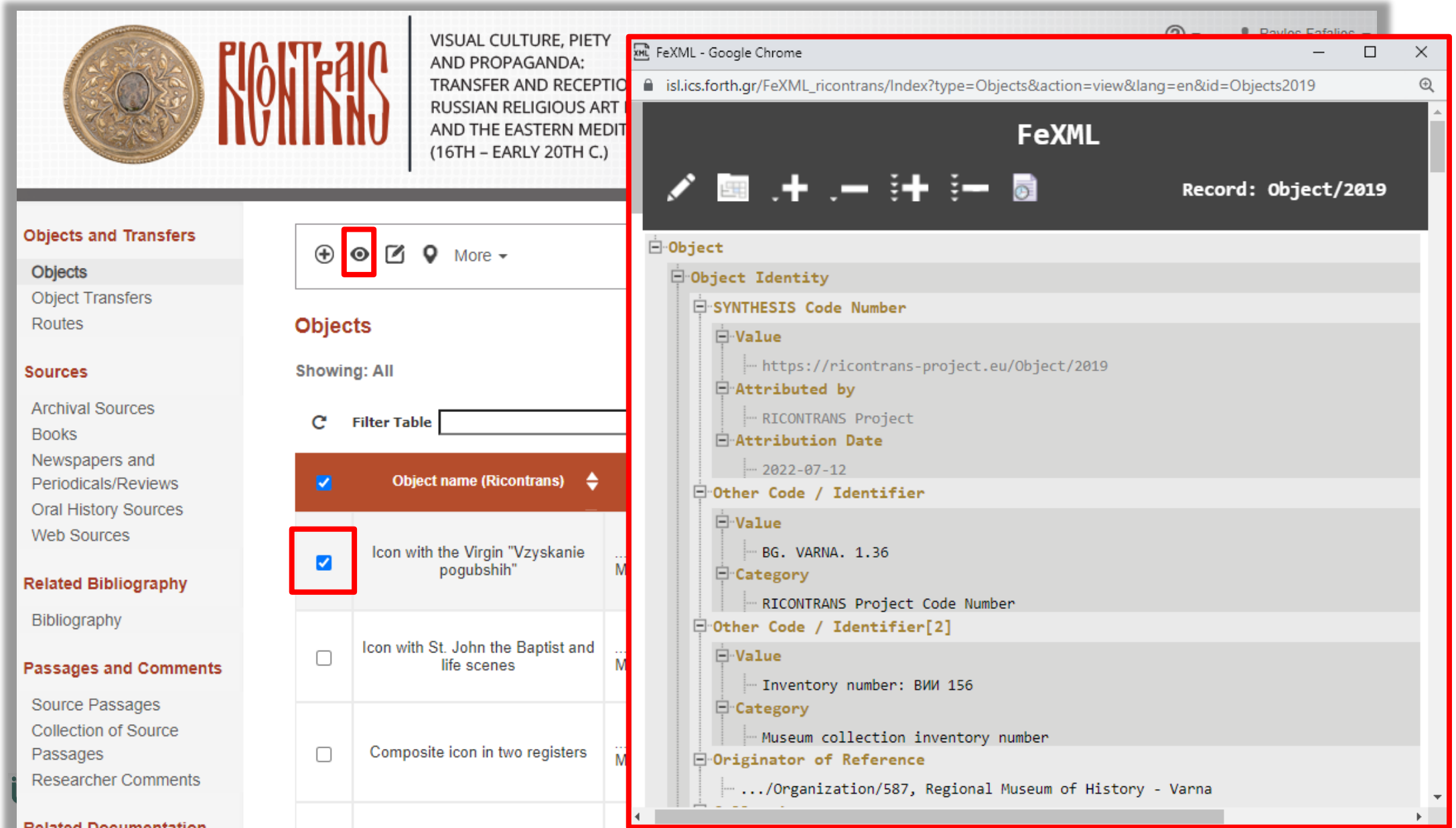
Synthesis – Inspecting the documented entities



Synthesis – Creating a new entity for documentation



Synthesis – Viewing an existing entity



Synthesis – Editing an existing entity

Synthesis – Data Model

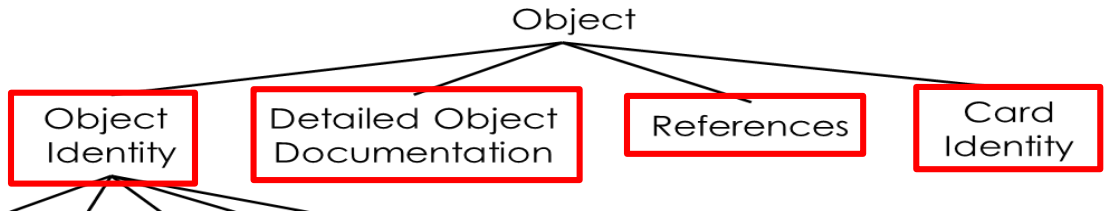
- Specially-designed for the domain of **History of Art** and the particular needs of the **RICONTRANS** project

  - 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

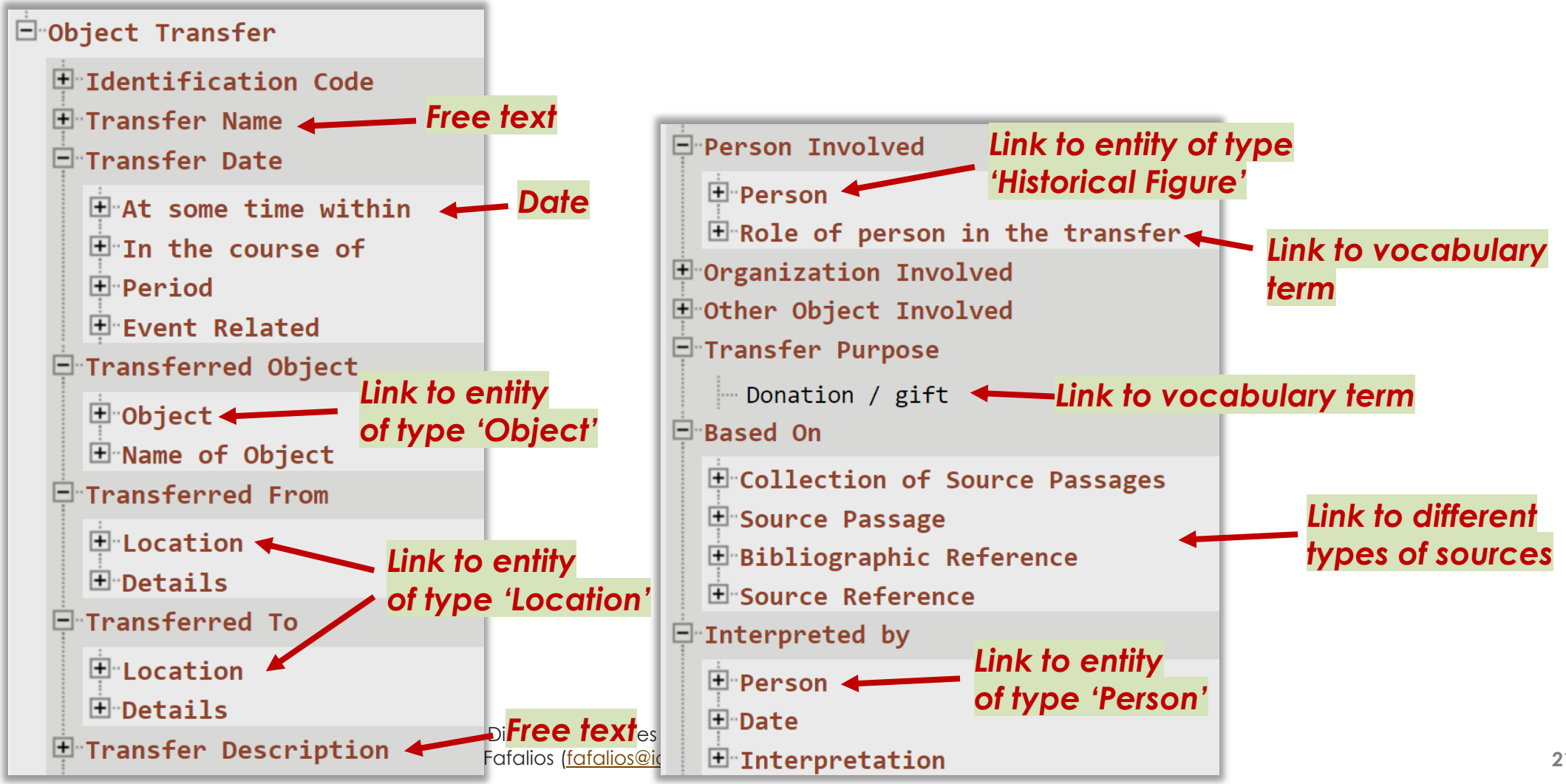
Entity type: **Object**



- Object
  - Object Identity
    - 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/Informa
        - 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

Synthesis – Data Model

Entity type: Object Transfer



Synthesis – Types of documentation fields

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

https://www.ics.forth.gr/isl/themas-thesaurus-management-system

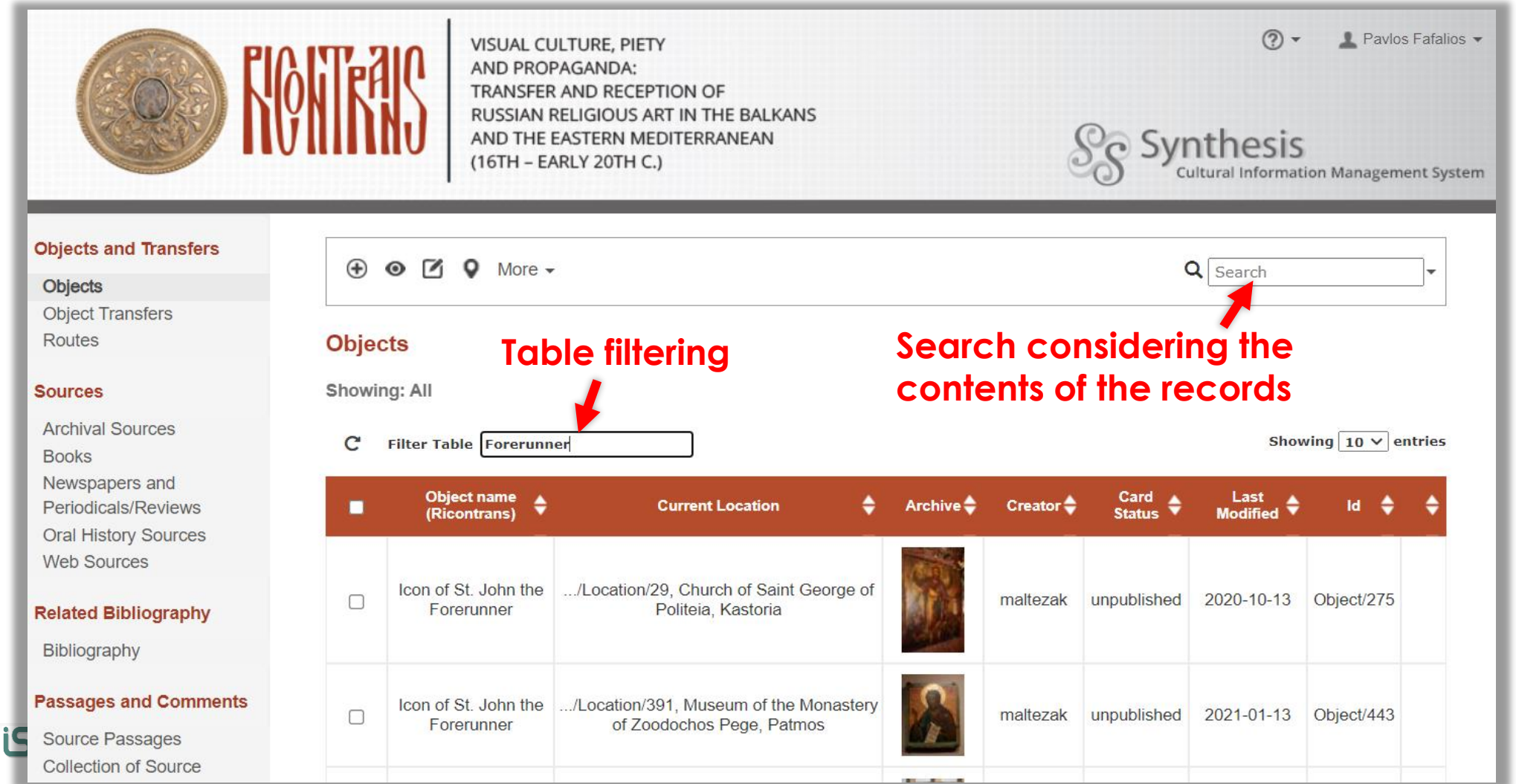
- **Unformatted** free text
- **Formatted** free text
- Number
- Date expression (**date range** in an accepted format) →
- Location coordinates (**point** or **polygon**)
- Location ID (**TGN** or **Geonames**)
- Upload file(s)

Examples of accepted time expressions

- 1821 January 2
- 1996 February
- 1945
- decade of 1970
- seventh decade of 20th century
- 20th century
- 1920 - 1950
- 3rd century - 5th century
- decade of 1920 - decade of 1950
- 18th century - decade of 1850
- early 16th century
- mid 20th century
- late 19th century
- 1st half 4th century
- 3rd quarter 1st century
- ca. 1920
- 1500 BCE
- 23rd century BCE
- early 4th century BCE
- 1st half 3rd century BCE
- 1800 - 1500 BCE
- 300 BCE - 300 CE
- 7th century - 5th century BCE
- 3rd century BCE - 1st century CE
- 3rd century - 5th century

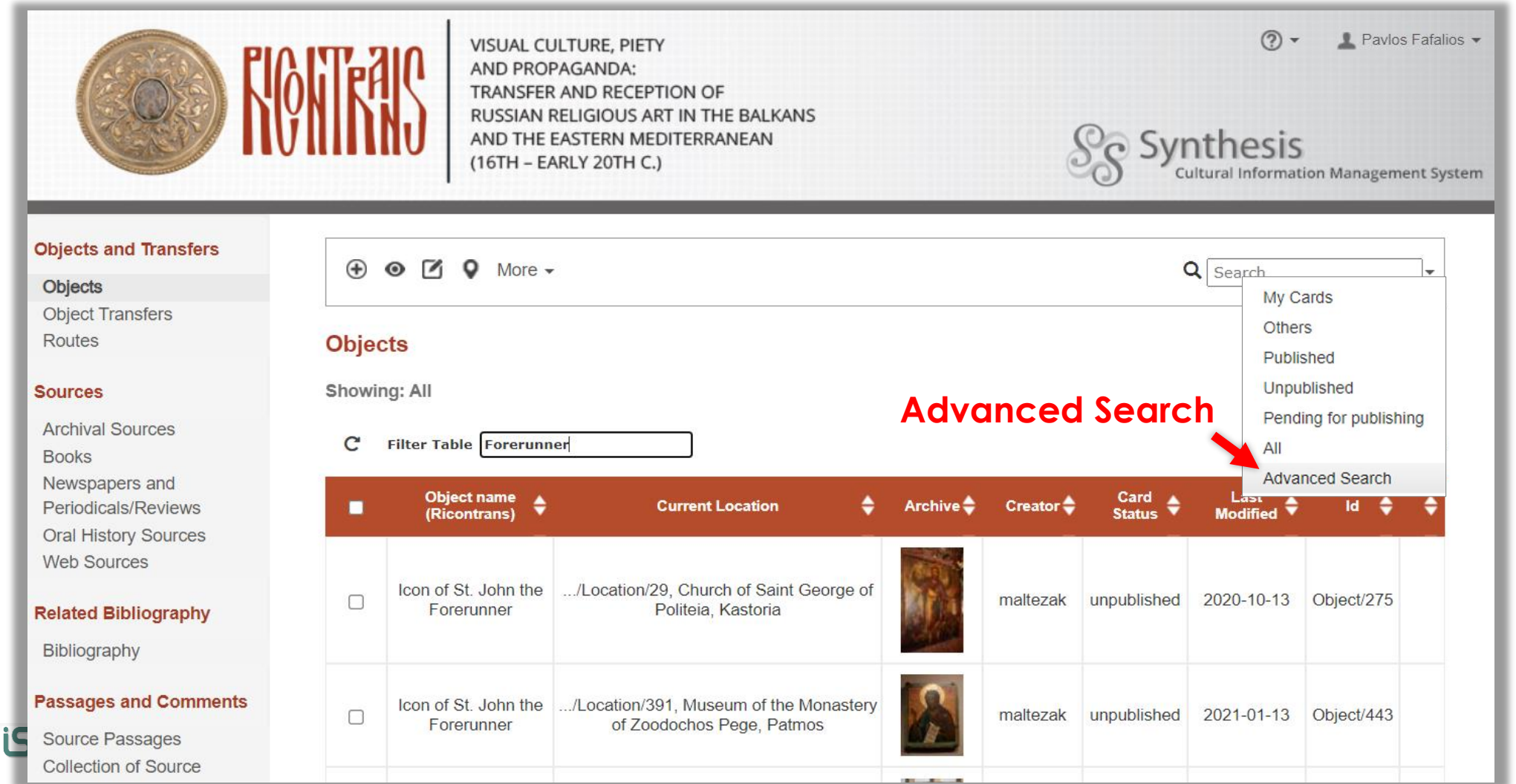
Synthesis – Other functionalities

- Table filtering / Search



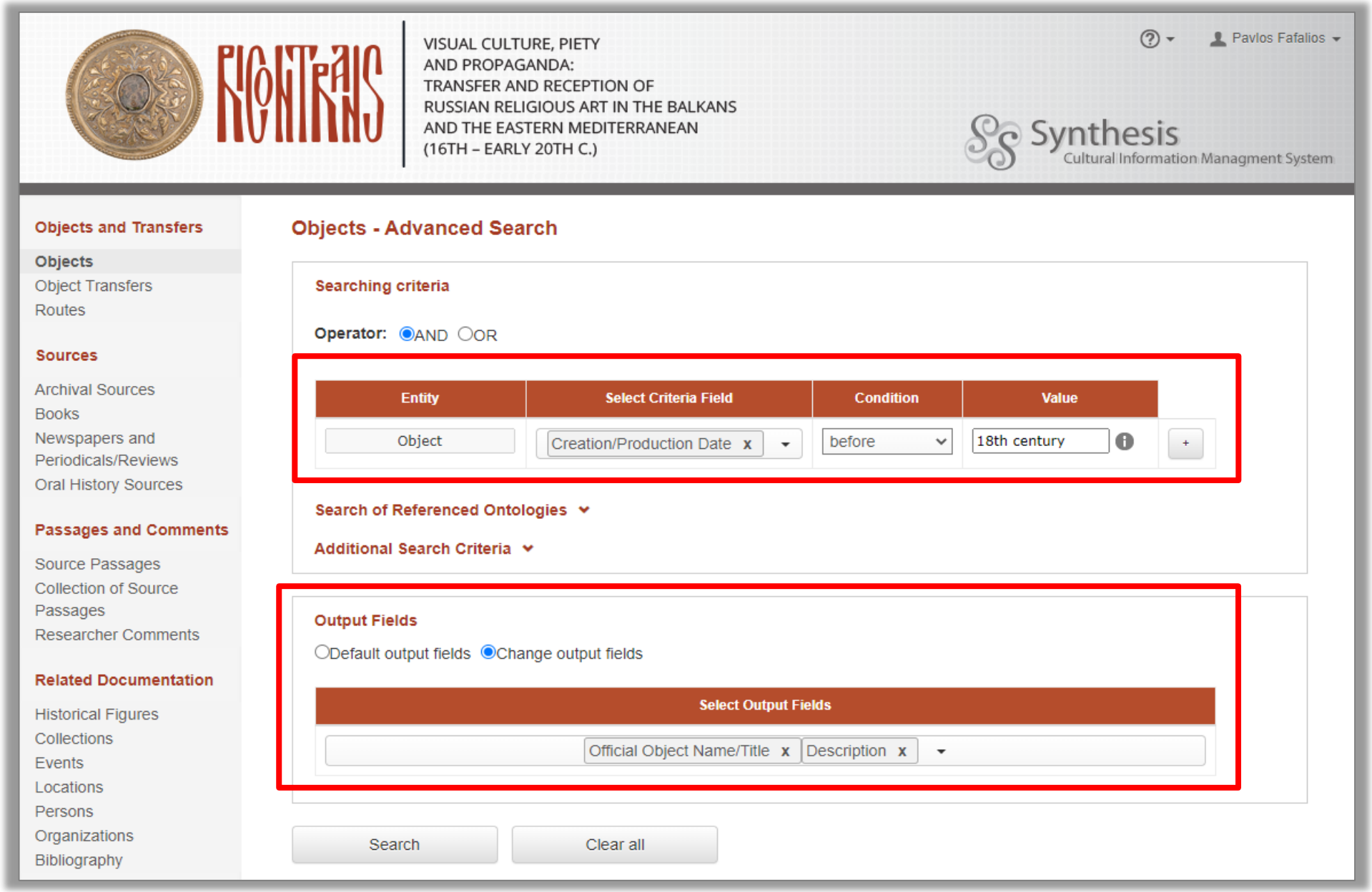
Synthesis – Other functionalities

- Table filtering / Search



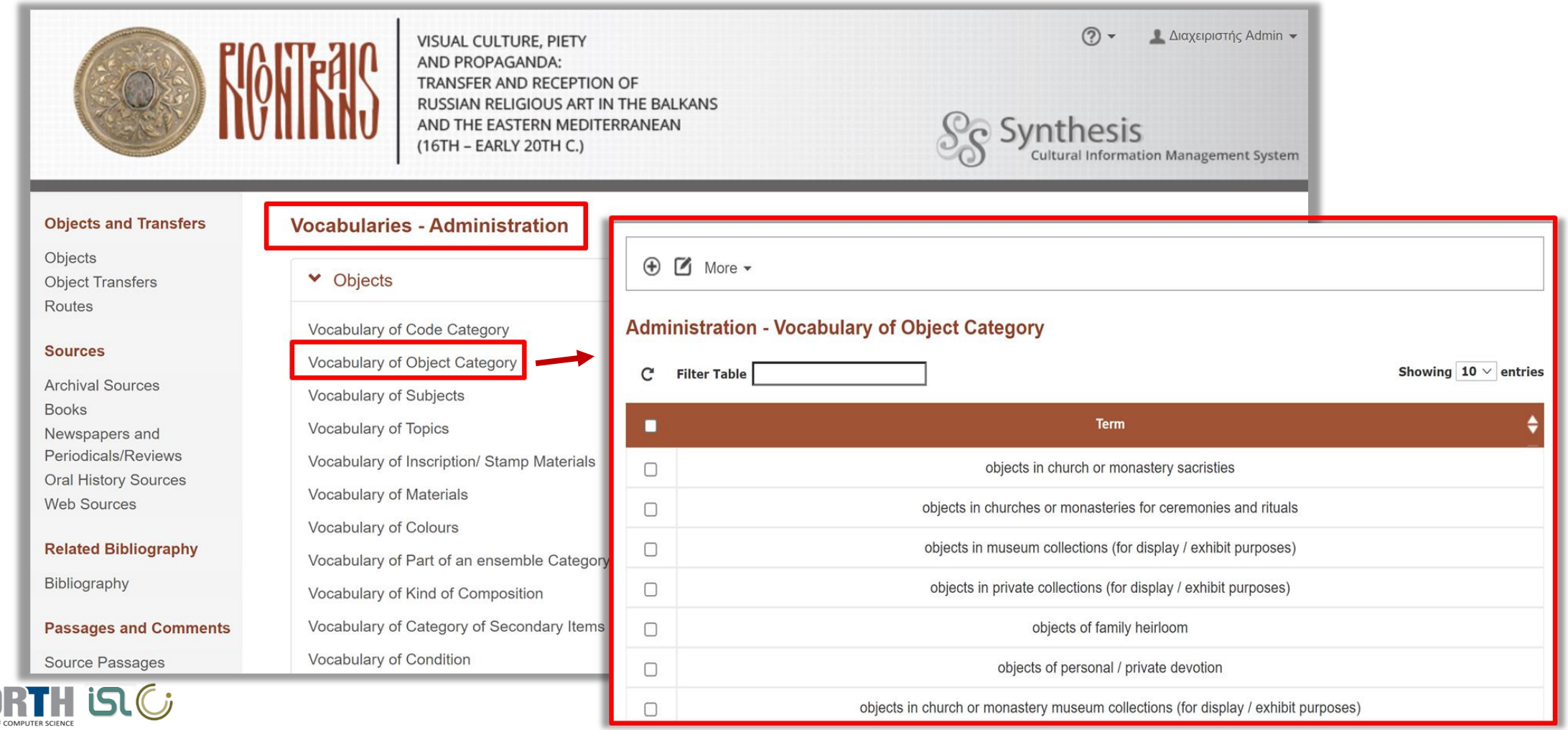
Synthesis – Other functionalities

- Advanced Search



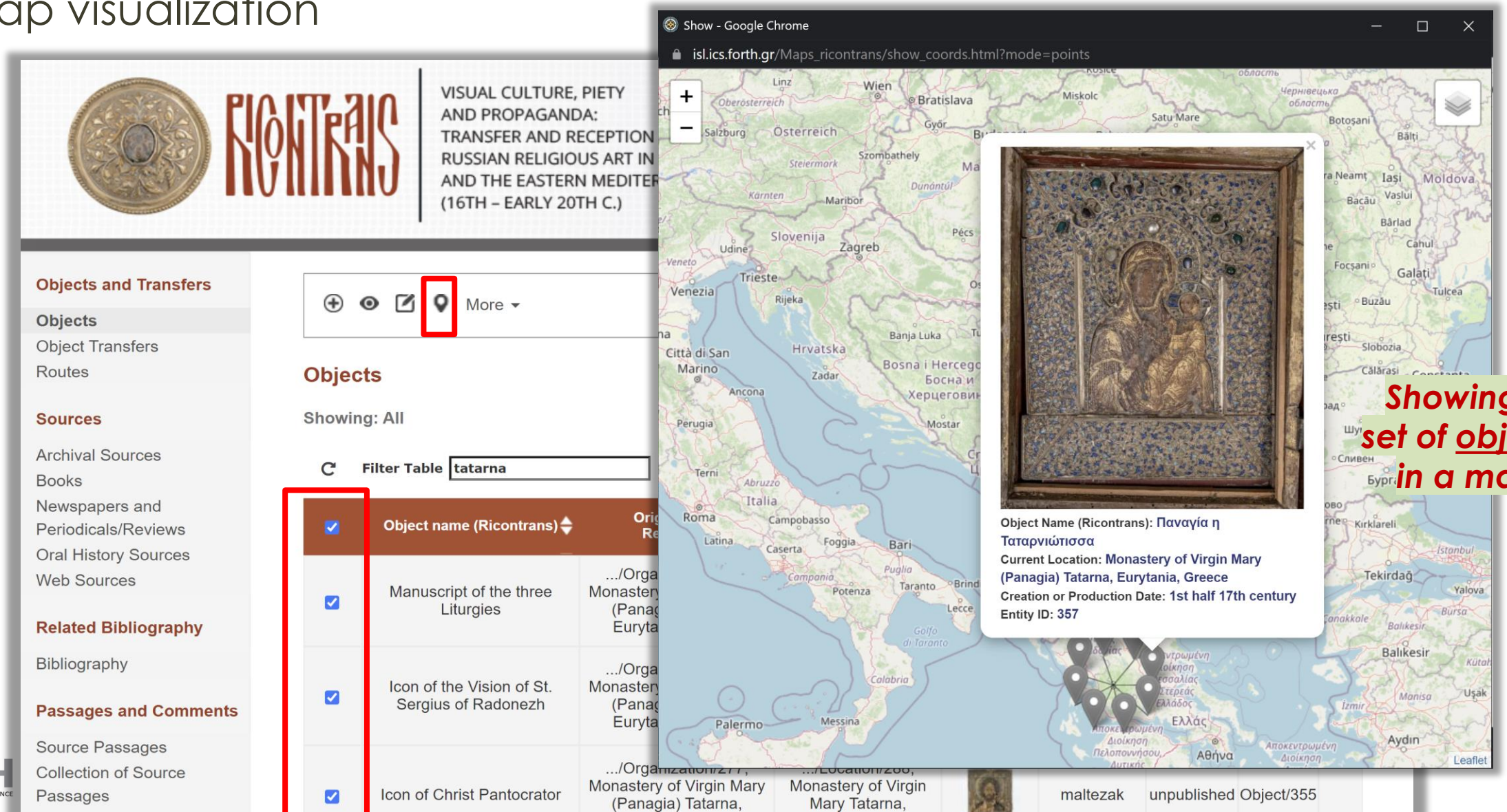
Synthesis – Other functionalities

- Management of vocabularies (*add*, *delete*, *rename* terms)



Synthesis – Other functionalities

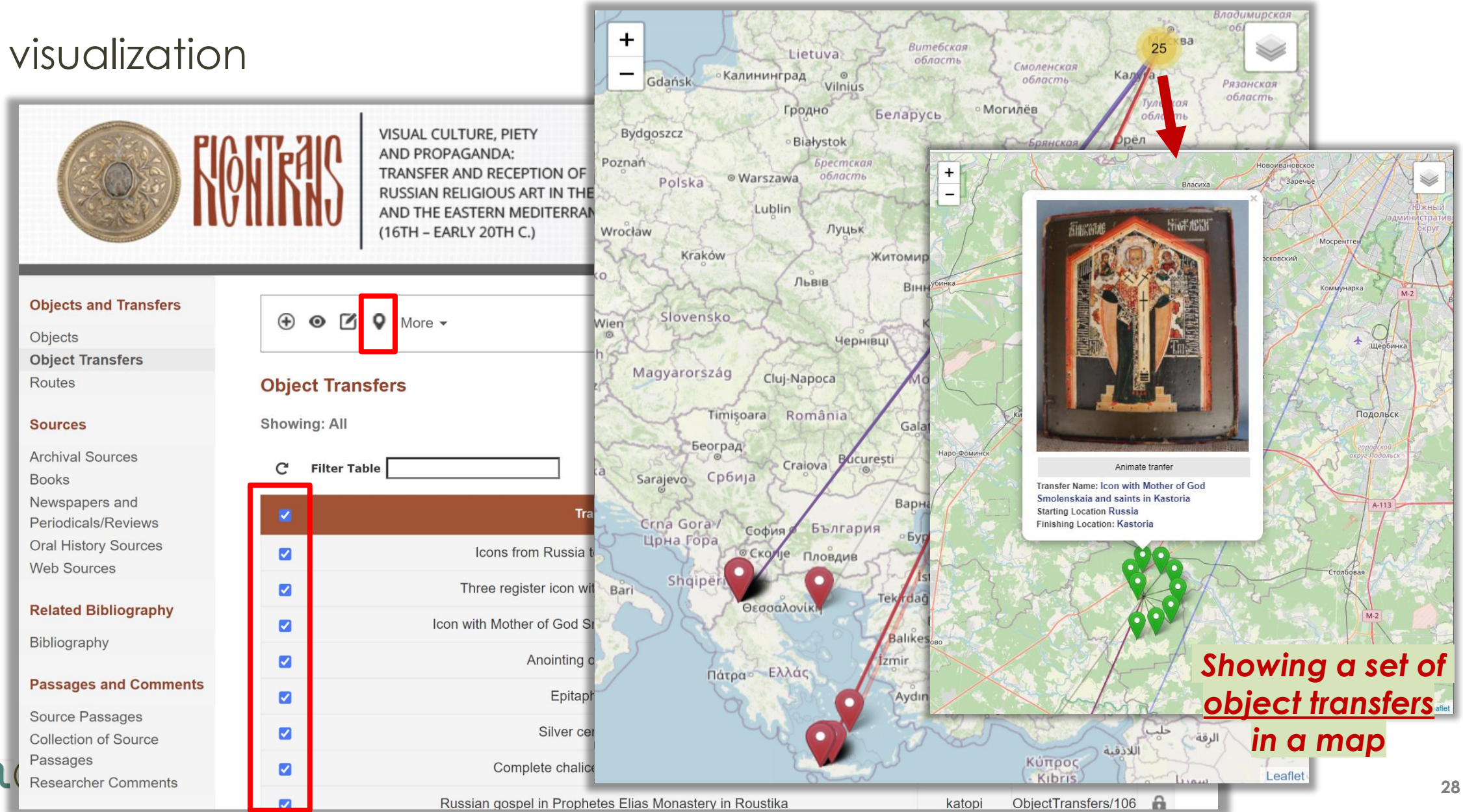
- Map visualization



Showing a set of objects in a map

Synthesis – Other functionalities

- Map visualization



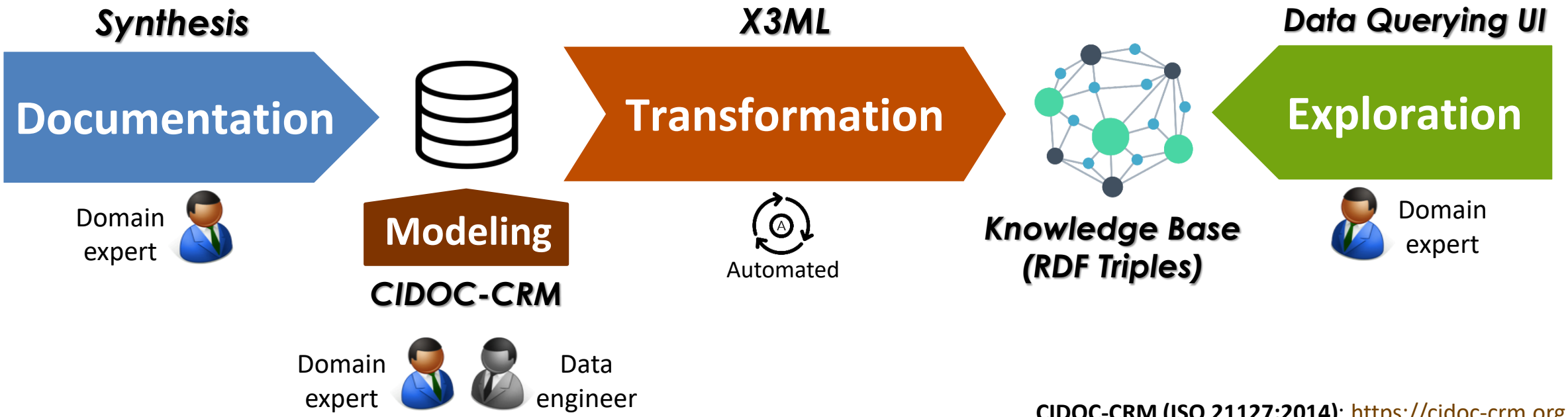
Use in RICONTRANS

- ~40 users (5 countries)
- Current number of documented entities (as of July 2022):
  - Objects: 1,928
  - Object transfers: 714
  - Routes: 97
  - Archival sources: 230
  - Books: 58
  - Newspapers and Periodicals/Reviews: 155
  - Oral History Sources: 3
  - Web Sources: 147
  - Bibliography: 497
  - Source Passages: 1059
  - Collection of Source Passages: 8
  - Researcher Comments: 0
  - Historical Figures: 259
  - Historical Events: 38
  - Collections: 210
  - Locations: 665
  - Persons: 136
  - Organizations: 488
  - Digital Objects: 1,880

Data transformation to a rich **semantic network**

Creating a **Semantic Network** – Process

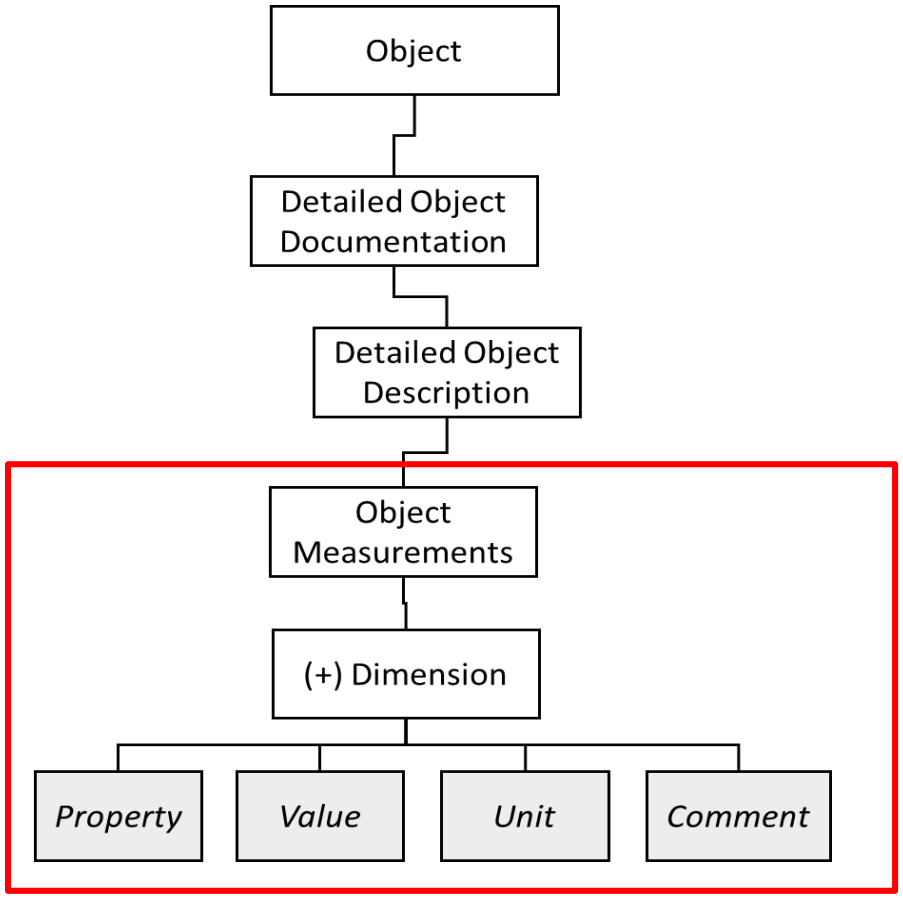
- *Synthesis* has embedded processes for transforming the data stored in the XML documents to an **ontology-based RDF dataset (Knowledge Base)**
  - It **decouples** data entry *(made by the research team)* from the ontology-based integration and creation of the KB *(supported by data engineers)*



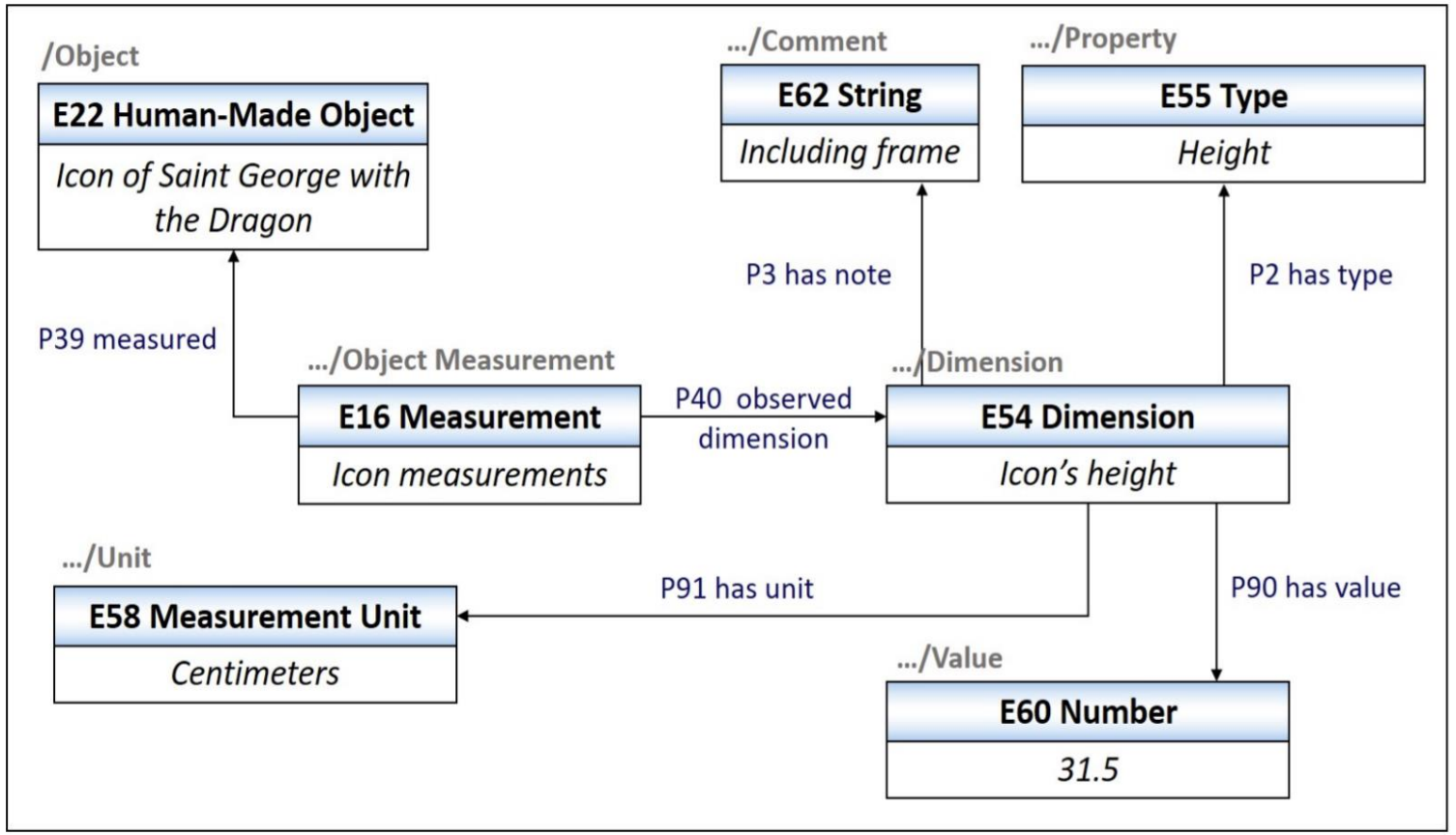
CIDOC-CRM (ISO 21127:2014): https://cidoc-crm.org/
X3ML Toolkit: https://www.ics.forth.gr/isl/x3ml-toolkit

Creating a **Semantic Network** – Example

Part of object's documentation schema



Modelling of object measurements in CIDOC-CRM



Creating a **Semantic Network** – Why?

- Enables **semantic interoperability**
  - The ability of computer systems to exchange data with unambiguous, shared meaning
- Facilitates **data integration**
  - With other, external datasets that also make use of **CIDOC-CRM**
- Supports **advanced data querying**
  - *"Find the routes of **icons** transferred to **Mount Athos** before the **18th century** and the **purpose** of these transfers"*

Creating a **Semantic Network** – 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**

Conclusion

+ Lessons learned and **future work**

Conclusion

- Data documentation and management with ***Synthesis***
  - Web-based and collaborative
  - Focus on semantic interoperability (compatibility with **CIDOC-CRM**)
  - **Aim:** production of **sustainable** data with **high value** and **long term validity**
- 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**

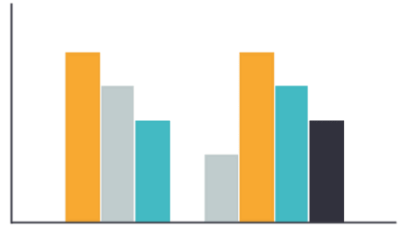
Lessons Learned

- 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 (which can be laborious)
    - *How could we support a better management of the dynamic vocabularies?*

Future Work

- Additional **data visualizations** (dynamic production of **charts**)
  - E.g., group all **objects** by 'object type' and show a **bar chart**



- Creation of '**private entities**' that can be only viewed by the creator
  - For documenting data related to **ongoing (unpublished) research**

More info:

- Pavlos Fafalios et al. "*Towards Semantic Interoperability in Historical Research: Documenting Research Data and Knowledge with Synthesis", International Semantic Web Conference*. Springer, 2021. https://arxiv.org/pdf/2107.13957.pdf
- RICONTRANS project: https://ricontrans-project.eu/
- Synthesis-core: https://www.ics.forth.gr/isl/synthesis-core

The research & development team:

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

Centre for Cultural Informatics (CCI)
Institute of Computer Science (ICS)
FORTH

Thank you!

Pavlos Fafalios
fafalios@ics.forth.gr

Centre for Cultural Informatics
FORTH-ICS