

Exploring CIDOC CRM datasets with A-QuB

Towards a **generic, user-friendly** and **ready-to-use**
data exploration interface for **end-users**

Pavlos Fafalios

Technical University of Crete and FORTH-ICS

fafalios@ics.forth.gr



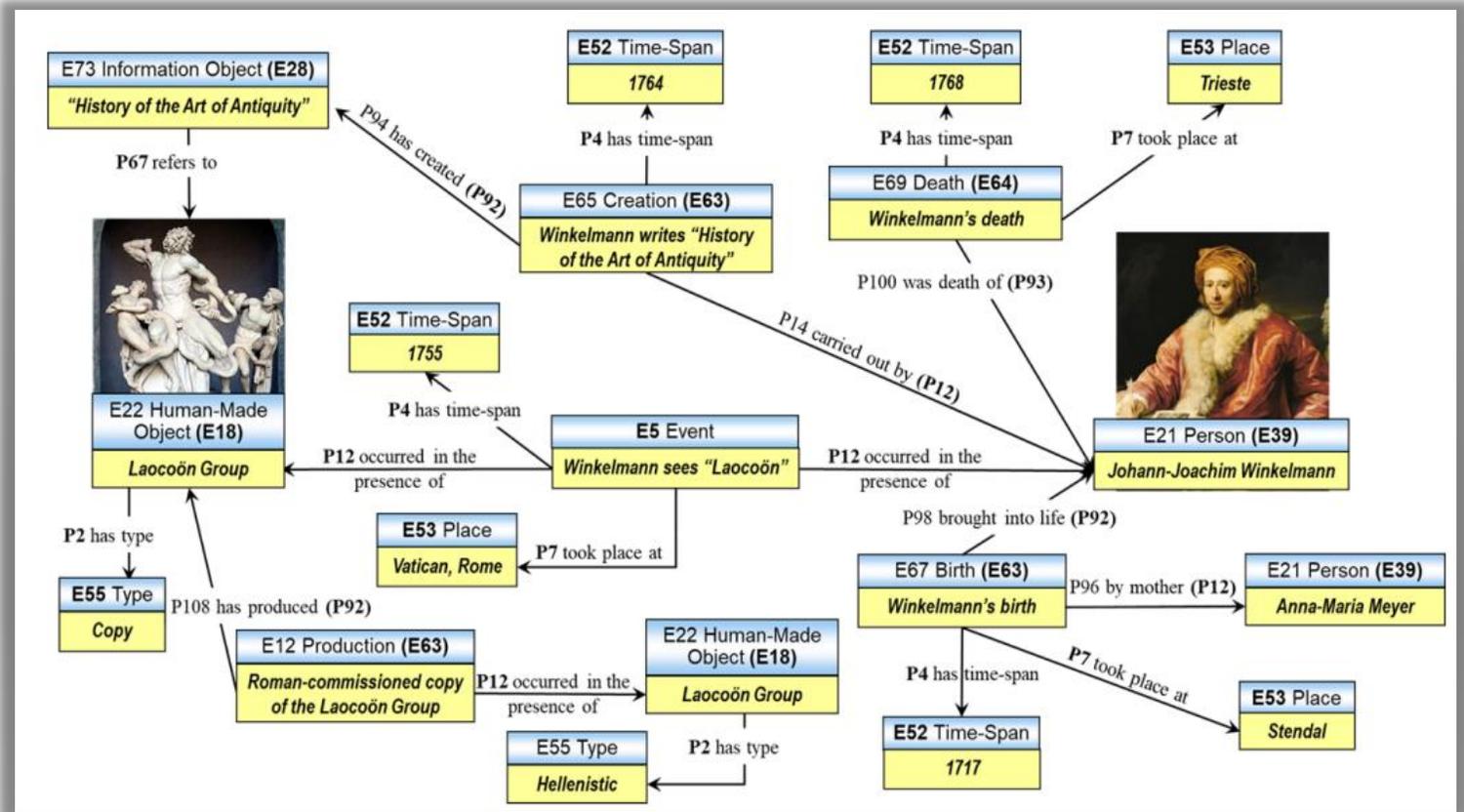
FORTH

FOUNDATION FOR RESEARCH AND TECHNOLOGY - HELLAS
INSTITUTE OF COMPUTER SCIENCE

CIDOC CRM (Conceptual Reference Model)

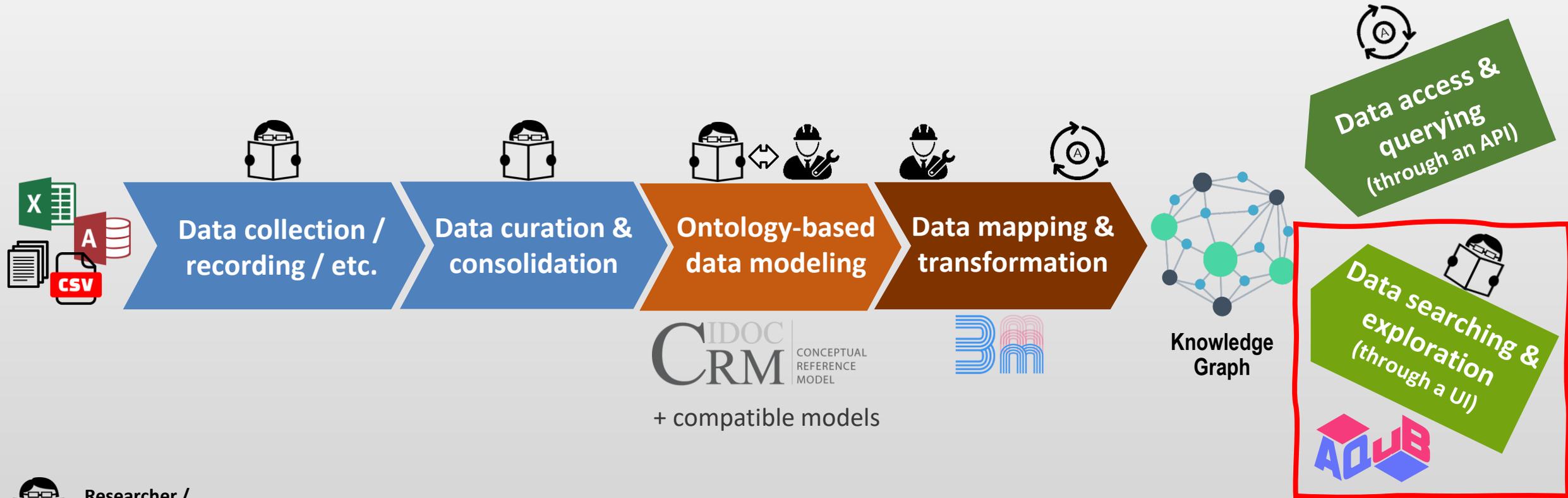
- A theoretical and practical tool for **data integration** in the field of **cultural heritage**
- A high-level **ontology** of human activity, things and events happening in spacetime
- It provides a formal structure (**hierarchy of classes and properties**) for describing concepts and relationships used in **cultural heritage documentation** (and beyond)
- An ISO standard (**ISO 21127:2023**)
- Current version: **7.1.3** (Feb 2024)

The outcome of over 25 years of collaborative development and maintenance work!



<https://www.cidoc-crm.org/>

The usual (**data integration**) workflow



-  Researcher / Domain Expert
-  Data Engineer
-  Software / Automated process

The web application 'A-QuB'

- ❑ **Web application operating over a knowledge graph (RDF triplestore)**
 - It facilitates the **exploration** of semantic (RDF) data **by plain users** (like historians)
 - It offers a **user-friendly** way to build and execute complex queries **step-by-step**
 - It is highly **configurable** (for use over any ontology / data model)
- ❑ **Other functionalities:**
 - data filtering, results browsing, query save, data export
- ❑ **First version developed in 2018, in the context of the VRE4EIC project (H2020)**
 - **Related paper**: Kritsotakis, Vangelis, Yannis Roussakis, Theodore Patkos, and Maria Theodoridou. "Assistive Query Building for Semantic Data." In *SEMANTiCS (Posters & Demos)*. 2018. https://ceur-ws.org/Vol-2198/paper_107.pdf
 - **Current FORTH-ICS development team**: Vangelis Kritsotakis, Konsolaki Konstantina, Pavlos Fafalios



❑ Available at: <https://github.com/isl/A-QuB-2>

A-QuB – Architecture and Technologies

- Architecture:



- Technologies:

- Java / Spring framework
- Angular



A-QuB – The search interface

The screenshot displays the A-QuB search interface. At the top left is the logo for the Archaeological Museum of Messara (MM) and the text "Data Exploration". A search bar contains the text "Explore Data". On the right side of the search bar are icons for a menu, settings, and help. Below the search bar, there are two main panels. The left panel, titled "Searching for...", lists several categories: "Object (291)", "Monument (58)", "Place (98)", "Site (47)", "Digital Object (780)", and "Display Place (22)". A red rectangular box highlights this list. A plus sign icon is positioned between the two panels. The right panel, titled "... related ... to entity", contains two dropdown menus labeled "Select Relation" and "Select Related Entity". A red text annotation points to the category list with the text "Categories configured in a pre-processing step". At the bottom left is a "SEARCH" button, and at the bottom right is a "RESET" button.

ARCHAEOLOGICAL MUSEUM of MESSARA

Data Exploration

Explore Data

Searching for...

- Object (291)
- Monument (58)
- Place (98)
- Site (47)
- Digital Object (780)
- Display Place (22)

... related ... to entity

Select Relation

Select Related Entity

SEARCH

RESET

Categories configured in a pre-processing step

A-QuB – The search interface

Searching for entities of a specific category (e.g. Object)

The screenshot displays the 'Explore Data' search interface. The top navigation bar includes a search icon, the text 'Explore Data', and utility icons for sharing, menu, settings, and help. The main search area is divided into two panels. The left panel, titled 'Searching for...', has a dropdown menu for 'Select Target Entity *' with 'Object (291)' selected, and a text input for 'Containing keyword'. The right panel, titled '... related ... to entity', has dropdowns for 'Select Relation' and 'Select Related Entity'. A plus sign icon is positioned between the two panels. Below the search panels is a blue 'SEARCH' button and a pink 'RESET' button. The bottom section shows 'Data Results (291 results)' with a table header 'OBJECT' and columns for 'NAME', 'TYPE', 'COLLECTION', and 'IMAGE'. Two rows of results are visible, each with a small image of a ceramic object.

OBJECT			
NAME	TYPE	COLLECTION	IMAGE
Τσαγιέρα HTR2971	Τσαγιέρα	Κεραμικά	
Αρύβαλλος AMM442	Αρύβαλλος	Κεραμικά	

A-QuB – The search interface

Browsing the search results

The screenshot displays the A-QuB search interface. At the top, a blue header contains a magnifying glass icon and the text "Explore Data". Below this, there are two main search panels. The left panel, titled "Searching for...", includes a "Select Target Entity *" dropdown menu with "Object (291)" selected, a "Containing keyword" input field, and a blue "SEARCH" button. The right panel, titled "... related", has a "Select Relation" dropdown menu. A blue circle with a white plus sign is positioned between the two panels. Below the search panels is a blue header for "Data Results (291 results)". Underneath is a table with columns for "NAME", "TYPE", and "COLLECTOR". The first row of the table is highlighted in pink and has a red box around the "NAME" cell, which contains "Τσαγιέρα HTR2971". To the right of the search interface is an "Information Card of Τσαγιέρα HTR2971 (Object)". This card is also highlighted with a red border and contains several sections with pink headers and white backgrounds, each listing a property and its value:

- NAME**: Τσαγιέρα HTR2971
- TYPE**: Τσαγιέρα
- CONDITION**: Ελλίπες
- MAIN MATERIAL**: Πηλός
- FOUND IN**: Αγία Τριάδα
- DATE**: Προανακτορική περίοδος (ΜΜΙΑ)
- LOCATION IN MUSEUM**: Μουσείοκατασκευή
- NOTE**: Πήλινη «τσαγιέρα» του τύπου των Πατερικών

A-QuB – The search interface

The screenshot displays the A-QuB search interface. On the left, a panel titled "Searching for..." allows selecting a target entity, with "Object (291)" chosen. Below it is a field for a containing keyword. On the right, a panel titled "... related ... to entity" allows selecting a relation ("has production date") and a related entity ("Date"). The "Date" entity selection includes fields for "From Date" and "Until Date", with a dropdown menu showing "8ος αιώνας π.Χ." and "6ος αιώνας π.Χ.". A green plus sign is between the panels, and a blue arrow icon is on the right. Below the panels are "SEARCH" and "RESET" buttons. At the bottom, a "Data Results (13 results)" table is shown.

OBJECT	DATE
OBJECT NAME	DATE
Επιγραφή Go377	7ος αιώνας π.Χ.
Αρύβαλλος AMM442	8ος αιώνας - 7ος αιώνας π.Χ.
Πλακίδιο AMM498	7ος αιώνας π.Χ.
Αρύβαλλος AMM579	Μέσα 7ου αιώνα π.Χ.
Επιγραφή Go399	600 - 525 π.Χ.
Γεωμετρικός πίθος με αρχαϊκή επιγραφή AMM457	Τέλη 8ου αιώνα π.Χ.

Searching for objects that have production date the period 8th – 6th century BC

A-QuB – The search interface

Searching for monuments and the objects found in them

The screenshot displays the A-QuB search interface. It features two main search panels. The left panel, titled "Searching for...", has a dropdown menu for "Select Target Entity*" with "Monument (58)" selected. Below it is a text input field for "Containing keyword". The right panel, titled "... related ... to entity", has a dropdown for "Select Relation*" with "is finding place of" selected and another dropdown for "Select Related Entity*" with "Object" selected. It also has a "Containing keyword" field and a search icon. A green plus sign is between the panels. Below the panels are "SEARCH" and "RESET" buttons. The results section, titled "Data Results (31 results)", shows a table with two columns: "MONUMENT" and "OBJECT".

MONUMENT	OBJECT
MONUMENT NAME	OBJECT NAME
Ιερό Πυθίου Απόλλωνα στη Γόρτυνα	Επιγραφή Go377
	Επιγραφή Go399
	Άγαλμα Αθηνάς με αιγίδα AMM615
	Άγαλμα Ήρας τύπου Borgese AMM616
	Αφιερωματική επιγραφή AMM593
	Κολοσσιαίο άγαλμα Πύθιου Απόλλωνος AMM629
Ανάκτορο Φαιστού	Πίθος AMM44
	Μόνωτο τροπιδιωτό κύπελλο Π10119
	Τμήματα από τράπεζα προσφορών AMM50
	Σφραγίδα με παράσταση ιχθύων Σ-K60
	Πώμα Π10527
	Σφραγίδα με παράσταση αιγάνου Σ-K345

A-QuB – The search interface

*Searching for objects that
i) are of a specific material,
ii) are of a specific cultural
period, iii) are of a specific
type*

The screenshot displays the A-QuB search interface with three stacked search criteria panels. Each panel has a blue header with a close button (X) and a help icon (downward arrow). The first panel, titled "Searching for...", has a "Select Target Entity *" dropdown menu with "Object (291)" selected. Below it is a "Containing keyword" input field. The second panel, titled "... related ... to entity", has a "Select Relation *" dropdown menu with "has material" selected and a "Select Related Entity *" dropdown menu with "Material" selected. Below the "Material" dropdown is a keyword input field with "Λίθος" (Stone) entered and a search icon. The third panel, also titled "... related ... to entity", has a "Select Relation *" dropdown menu with "was produced during period" selected and a "Select Related Entity *" dropdown menu with "Cultural Period" selected. Below the "Cultural Period" dropdown is a keyword input field with "Νεολιθική περίοδος" (Neolithic period) entered and a search icon. The fourth panel, also titled "... related ... to entity", has a "Select Relation *" dropdown menu with "has type" selected and a "Select Related Entity *" dropdown menu with "Type" selected. Below the "Type" dropdown is a keyword input field with "Συσσωμάτωμα" (Composite) entered and a search icon. "AND" operators are visible between the second and third panels, and between the third and fourth panels.

Other systems offering a similar search interface

❑ ResearchSpace

<https://researchspace.org/>

The screenshot shows the ResearchSpace search interface. At the top, there are tabs for 'Keyword search', 'Semantic search', and 'Records'. Below this, a search query is displayed: 'Find: People WAS CREW AT Ship where ships arrived at Barcelona'. The query is visualized as a sequence of entities and relationships: 'Person' (with a person icon) 'was crew at' 'Ship' (with a ship icon). A 'WHERE' clause is shown below: 'Ship' (with a ship icon) 'arrived at' 'Place' (with a location pin icon) 'Barcelona'. The interface indicates 'Found 604 matches' and provides options for 'Timeline', 'Chart', and 'List'. A 'Filter Results' section is visible on the left, with a search box containing 'subject' and a list of results including 'Agustin Farche Saura', 'Agustin Fernandez', and 'Agustin Gonda Arellano'.

Example in the Maritime History domain (SeaLiT project):

<https://rs.sealitproject.eu/>

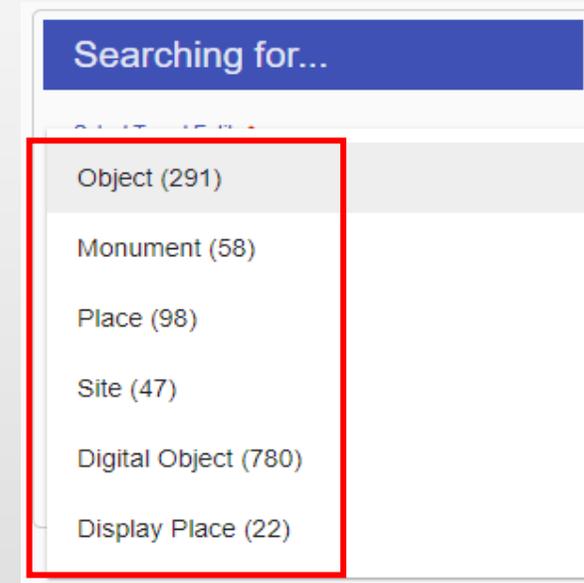
❑ Sparnatural

<https://sparnatural.eu/>

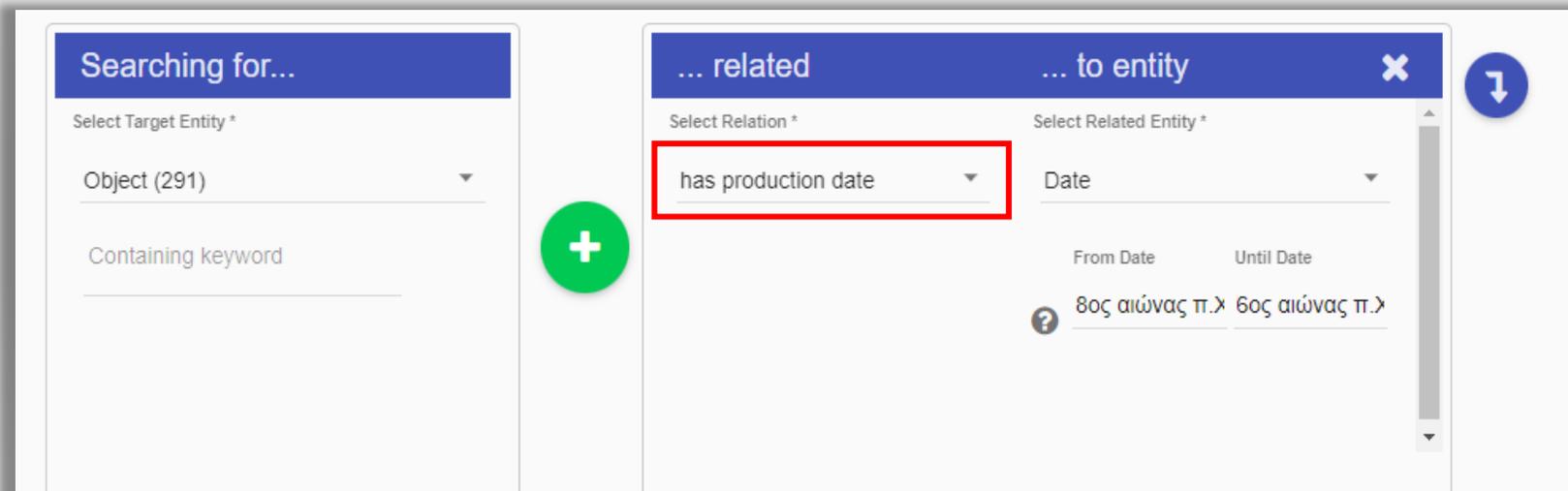
The screenshot shows the Sparnatural search interface, which uses a visual query builder. The query is constructed from several components: 'Artwork' (with a paintbrush icon) 'displayed at' 'Museum' (with a museum icon). A 'Where' clause is applied to the 'Museum' entity, specifying 'country' as 'Country' (with a globe icon) and 'France (222)'. The main query is connected to another 'And' clause: 'Museum' (with a museum icon) 'displays' 'Artwork' (with a paintbrush icon). This second 'Artwork' entity has a 'Where' clause for 'author', listing 'Michelangelo' and 'Caravaggio'. A final 'And' clause is connected to the main query: 'Artwork' (with a paintbrush icon) 'author' 'Person' (with a person icon). This 'Person' entity has a 'Where' clause for 'birth date', specified as 'Date' (with a calendar icon) 'From 01-01-1500 to 31-12...'. Each component in the query builder has a close button (X) for removal.

Configuring A-QuB for a specific use case

- Configuration of the **entity categories**



- Configuration of the **entity relationships**



A-QuB – The configuration model

File: config.properties

Defining an entity category

```
entities = 1,2,3,4,5,6,7,8,9,10,11,12,13,14,15
```

```
entities.1.name = Object
```

```
entities.1.instancesQuery = SELECT COUNT(DISTINCT ?uri) [[FROM_GRAPHS]] WHERE { ?uri a <http:
```

```
entities.1.searchQuery = SELECT DISTINCT ?uri ?name ?type ?collection ?image [[FROM_GRAPHS]]
```

```
entities.1.sourceUriVariable = ?uri
```

```
entities.1.exploreQuery = SELECT ?name ?type ?condition ?main_material ?found_in ?found_in__f
```

```
entities.1.exploreVariableTypes = found_in:4
```

```
entities.1.entityType = LITERAL
```

→ How many entities belong to this category

→ How to search for entities of this category (and what information to show in the results table)

→ What information to show when browsing an entity of this category

A-QuB – The configuration model

- File: config.properties

Defining an entity relationship

```
relations.2.name = has material
relations.2.sourceEntity = 1
relations.2.destinationEntity = 7
relations.2.relationGraphPattern = ?object[[i]] <http://www.cidoc-crm.org/cidoc-crm/P45 consists
relations.2.sourceUriVariable = ?object[[i]]
relations.2.sourceShownVariables = ?object_name[[i]]
relations.2.destinationUriVariable = ?mat[[j]]
relations.2.destinationShownVariables = ?material[[j]]
relations.2.sourceFilterPattern = [[REGEX(?object_name[[i]])]] [[VALUES(?object[[i]])]]
relations.2.destinationFilterPattern = [[REGEX(?material[[j]])]] [[VALUES(?mat[[j]])]]
```

Graph pattern that associates the two categories through this relationship

The problem...

- Configuring the system for use in a particular use case (dataset) can be **very time consuming**
 - depending on the complexity of the data (number of categories, number of category relationships, etc.)
 - need to define the graph patterns for all categories and relationships

AQuB for Messara Archaeological Museum

- 15 categories, 35 relationships
- **Configuring AQuB took several days!**

SeaLiT Project (Maritime History):

- Around 15 categories and **>200 relationships!**
- **Configuring ResearchSpace took months!**

Towards a generic configuration

- Creation of a generic configuration for CIDOC CRM
- Rely on the idea of “**Fundamental Categories and Relationships**”

Tzompanaki, K., & Doerr, M. (2012). Fundamental Categories and Relationships for intuitive querying CIDOC-CRM based repositories. *ICS-FORTH Technical Report, 429*.

[https://publications.ics.forth.gr/tech-reports/2012/2012.TR429 Intuitive querying CIDOC-CRM.pdf](https://publications.ics.forth.gr/tech-reports/2012/2012.TR429%20Intuitive%20querying%20CIDOC-CRM.pdf)

- **Fundamental Categories and Relationships:**
 - A set of **general categories and relationships** that cover all different categories of entities and their relationships
 - For example, an instance of the fundamental category **Thing** can be a *museum object*, a *document*, a *song*, a *ship*, a *photograph collection*, a *cave*, etc.

Creating a generic configuration

7 fundamental categories:

- Actor
- Place
- Thing
- Period/Event
- Concept/Type
- Time
- Dimension

```
entities.1.name = Actor Actor
entities.1.instancesQuery = SELECT COUNT(DISTINCT ?uri) [[FROM_GRAPHS]] WHERE
entities.1.searchQuery = SELECT DISTINCT ?uri ?label ?class ?type [[FROM_GRA
    ?uri a <http://www.cidoc-crm.org/cidoc-crm/E39 Actor> . OPTIONAL {
    OPTIONAL {?uri <http://www.cidoc-crm.org/cidoc-crm/P2 has type> ?t
entities.1.exploreQuery = SELECT ?label ?name ?type ?type_more ?birth_period
entities.1.exploreVariableTypes =
entities.1.sourceUriVariable = ?uri
entities.1.entityType = LITERAL
```

```
entities.2.name = Place Place
entities.2.instancesQuery = SELECT COUNT(DISTINCT ?uri) [[FROM_GRAPHS]] WHERE
entities.2.searchQuery = SELECT DISTINCT ?uri ?label ?class ?type [[FROM_GRA
    ?uri a <http://www.cidoc-crm.org/cidoc-crm/E53 Place> . OPTIONAL { ?uri
    OPTIONAL {?uri <http://www.cidoc-crm.org/cidoc-crm/P2 has type> ?t . ?t
entities.2.exploreQuery = SELECT ?label [[FROM_GRAPHS]] WHERE { [[URI]] <http://www.cidoc-crm.org/cidoc-crm/E53
entities.2.exploreVariableTypes =
entities.2.sourceUriVariable = ?uri
entities.2.entityType = LITERAL
```

```
entities.3.name = Thing Thing
entities.3.instancesQuery = SELECT COUNT(DISTINCT ?uri) [[FROM_GRAPHS]] WHERE
entities.3.searchQuery = SELECT DISTINCT ?uri ?label ?class ?type [[FROM_GRA
    ?uri a <http://www.cidoc-crm.org/cidoc-crm/E70 Thing> FILTER NOT EXISTS
    OPTIONAL { ?uri <http://www.cidoc-crm.org/cidoc-crm/P2 has type> ?t
```

Creating a generic configuration

116 fundamental relationships

- connecting the fundamental categories pairwise

```
relations.1.name = has type
relations.1.sourceEntity = 1
relations.1.destinationEntity = 5
relations.1.relationGraphPattern = ?actor[[i]] a <http://www.cidoc-crm.org/cidoc-crm/E39 Actor> ; <http://www.cidoc-crm.org/cidoc-crm/E39 Actor>
relations.1.sourceUriVariable = ?actor[[i]]
relations.1.sourceShownVariables = ?actor_label[[i]]
relations.1.destinationUriVariable = ?type[[j]]
relations.1.destinationShownVariables = ?type_label[[j]]
relations.1.sourceFilterPattern = [[REGEX(?actor_label[[i]])]] [[VALUES(?actor[[i]])]]
relations.1.destinationFilterPattern = [[REGEX(?type_label[[j]])]] [[VALUES(?type[[j]])]]
```

Actor – has type – Concept

```
relations.17.name = was present at
relations.17.sourceEntity = 1
relations.17.destinationEntity = 4
relations.17.relationGraphPattern = ?event[[j]] <http://www.cidoc-crm.org/cidoc-crm/P12 occurred in the presence of>
relations.17.sourceUriVariable = ?actor[[i]]
```

Actor – was present at – Period/Event

Searching using the generic configuration

The screenshot shows a search interface with two main configuration panels. The left panel, titled "Searching for...", has a dropdown menu for "Select Target Entity" with "Thing (54)" selected. Below it is a text input field for "Containing keyword". The right panel, titled "... related ... to entity", has dropdowns for "Select Relation" and "Select Related Entity". A blue plus sign is between the panels. Below the configuration is a "SEARCH" button and a "RESET" button. The "Results" section shows a table with columns "LABEL", "CLASS", and "TYPE". Two rows are visible, both with "Ship" in the CLASS column. A red box highlights the "Ship" class in both rows. A red text box points to the "Ship" class, stating: "Ship is a (direct or indirect) subclass of 'Thing'".

Important:
Inference for the 'subclassOf' relationship must be enabled (or materialized)

LABEL	CLASS ↑	TYPE
ship Andrea	Ship	brigantino
ship Angelina	Ship	brigantino

Searching using the generic configuration

Searching for...
Select Target Entity *
Actor (195)
Containing keyword

... related
Select Relation *
was present at

... to entity
Select Related Entity *
Period/Event
Containing keyword

Important:
Inference for the 'subpropertyOf' relationships must be enabled (or materialized)

Results

ACTOR LABEL

Francesco Saverio Proto

EVENT LABEL

Voyage of ship Andrea
Birth of Francesco Saverio Proto
Employment of Francesco Saverio Proto
Payment to Francesco Saverio Proto

All these properties are (direct or indirect) subproperties of 'was present at'

Other important aspects

- Other important inferences that should be enabled/applied
 - **'owl:sameAs'** relationships
 - **'broader than'** relationships for types
 - **'falls within'** relationships for places
(and the same for other **transitive properties**)

Ongoing and future work

- Finalize the configuration part that allows exploring the entities in the results
 - Connect with other tools (e.g. 3M/X3ML, RDF visualizer)
 - Improve the UI
- Our (short term) aim: to have (publicly available) an A-QuB deployment which one can directly use for exploring their CIDOC CRM based datasets
- 1. Upload your dataset → 2. Start exploring it**

Thank you!

Questions?

Pavlos Fafalios

Technical University of Crete and FORTH-ICS

fafalios@ics.forth.gr



FORTH

FOUNDATION FOR RESEARCH AND TECHNOLOGY - HELLAS
INSTITUTE OF COMPUTER SCIENCE