JAGIELLONIAN UNIVERSITY
IN KRAKOW

The CHExRISH Project

03 April 2025
Luiz do Valle Miranda

Under supervision of: Krzysztoff Kutt and Grzegorz Nalepa
Supported from the CHExRISH project

Outline

1. CHExRISH and data integration

2. CIDOC-CRM in the CHExRISH project

3. JUHMP, CIDOC-CRM and Omeka-S

1. CHExRISH and data integration

The goal:

- Integrating Cultural Heritage (CH) data from different Jagiellonian University (JU) units;

- Creating the Jagiellonian University Heritage Metadata Portal (JUHMP);

- Incorporating Digital Humanities research tools, such as Network Analysis, VR visualization, knowledge discovery;

- Openness for further integration of other collcetions.

CHE×RISH
Cultural Heritage Exploration and Retrieval
with Intelligent Systems at the Jagiellonian University

1. CHExRISH and data integration

The **Archive of the Jagiellonian University** (AUJ):

- The basis for data integration among JU units are the records of AUJ;

- AUJ stores data of students and professors of JU across different periods of the university;

- The Corpus Academicum Cracoviense (CAC) is a SQL database with around 67,000 records on students and graduates of the University of Kraków during the period 1364–1780;

- The Corpus Studiosorum (CS) 1850-1918 is an Excel datasheet containing information about JU students during this period.

1. CHExRISH and data integration

Stanisław Reszka z Buku, syn Stanisława

	+
First name:	Stanisław
Surname:	Reszka
Father's name:	Stanisław
Place of origin:	Buk
Place and date of birth:	Sep 14, 1544
Place and date of death:	Apr 3, 1600

Education and academic degrees

⊞	↕	Event type	Education stage/academic degree	Scientific discipline	Institution	Date
⊞	1	uzyskanie stopnia	doktor		Uniwersytet w Perugii	the beginning of spring 1579 — May 23, 1579
⊞	1	początek studiów/nauki	student		Uniwersytet Krakowski (Akademia Krakowska)	the beginning of summer semester 1558 — the end of summer semester 1558

Functions/offices/roles

⊞	↕	Event type	Function/office/role	Place/Institution	Date
⊞	1	rezygnacja	dziekan (prałat)	kolegiata św. Jana w Warszawie	Mar 21, 1586
⊞	1	początek funkcji	dziekan (prałat)	kolegiata św. Jana w Warszawie	the beginning of 1583 year — the end of 1583 year

Works and books

	Event type	Tilte	Shelf mark	Date	Source of information
🔎	autor	Przestrogę pastyrską do miasta warszewskiego... (Poznań 1585)		the beginning of 1585 year — the end of 1585 year	PSB, t. 31
🔎	autor	Ministromachia (Kr. 1591)		the beginning of 1591 year — the end of 1591 year	PSB, t. 31

Fig. 1: Excerpt of Stanisław Reszka's biogram from CAC.

1. CHExRISH and data integration

The **Jagiellonian Library** (BJ):

- The Jagiellonian Digital Library (JDL) containing digitized collections of the Jagiellonian Library. Data from JDL follows the Dublin Core (DC) data model and can be harvested via the OAI-PMH protocol, with around 700,000 bibliographic records available;

- ALMA is an integrated management system for approximately 9 million bibliographic records. ALMA data can also be harvested via OAI-PMH, formatted in XML following a Marc21-inspired structure.

1. CHExRISH and data integration

```
@prefix dc <http://purl.org/dc/elements/1.1/> .
@prefix jbc <http://www.jbc.bj.uj.edu.pl/Content/> .

jbc:917900  dc:coverage "16 w."@pl
            dc:creator "Hozjusz, Stanislaw (1504-1579)"@pl ,
                       "Reszka, Stanislaw (1544-1600)"@pl ;
            dc:date "1585"@pl ;
            dc:language "lat"@pl ,
                        "pol"@pl ;
            dc:publisher "Drukarnia Jana Wolraba"@pl ;
            dc:rights "Domena publiczna (public domain)"@pl ;
            dc:source "Biblioteka Jagiellonska, BJ St. Dr. Cim. 4583"@pl ;
            dc:title "Przestroga Pastyrska, do Miasta Warszewskiego [...] "@pl ;
            dc:type "starodruk"@pl .
```

Fig. 2: DC description of Stanisław Reszka's work "Przestroga Pastyrska" taken from JDL

CHExRISH
Cultural Heritage Exploration and Retrieval
with Intelligent Systems at the Jagiellonian University

1. CHExRISH and data integration

The **Museum of the Jagiellonian University** (MUJ):

- MUJ holds a rich collection of historical artifacts related to the university's heritage, including scientific instruments, artworks, memorabilia of famous scholars, and valuable books;

- MUJ's data can be harvested via a custom API in a structured format, currently providing access to 75,000 items. Each record contains information about authors, date and place of creation, technique, materials, dimensions, etc..

1. CHExRISH and data integration

BJ and **MUJ** objects have three ways of relating persons to objects:

- A person can be an author of the object;

- A person may have contributed to the existence of the object;

- A person may be mentioned by the object.

2. CIDOC-CRM in the CHExRISH project

The motivation:

- CIDOC-CRM is an international standard (ISO 21127:2014) designed to serve as a shared language among CH institutions;

- With CIDOC-CRM cultural heritage data from different institutions, including libraries, museums, and archives can be connected by means of the events;

- Tools for semantic portal's enrichment are being developed based on this data model;

CHEXRISH
Cultural Heritage Exploration and Retrieval
with Intelligent Systems at the Jagiellonian University

2. CIDOC-CRM in the CHExRISH project

AUJ modelling with **CIDOC-CRM:**

- Academic and occupational events were modeled as CRM's "E7 Activity" in which both a person and an institution "P11 participated in";

- Academic events have several properties, including event type, degre. A solution for aggregating details in an activity is by creating an instance of class "E55 Type". Reszka "P11 participated in" his "academic event 1" which "P2 has type" " beginning of studies ";

- Instances of "E55 Type" of activities were grouped together by the creation of another "E55 Type". For example, both types "beginning of studies" and "end of studies" "P2 has type" "Educational Activity". Similarly "Mathematics " and "Philosophy" "P2 has type" "Academic Degree"

- "First Name" and "Surname" were modelled as "E41Appellation". However, how to deal with second names and maiden names? Maybe using " P48 has preferred identifier"?

2. CIDOC-CRM in the CHExRISH project

BJ and **MUJ** modelling with **CIDOC-CRM:**

- Bibliographic records modeled both as "E33 Linguistic Object" and "E22 Human-Made Object";

- "E65 Creation" and "E12 Production" used for linking a person with an object, though not directly;

- For every related person to an object, a respective "E7 Activity" was created which" P128i is carried by" a person and which " P9i forms part of" "E65 Creation" or "E12 Production". Such "E7 Activity" "P2 has type" the options such as publisher, author, engraving, etc.

- In MUJ other CIDOC-CRM classes are used, such as "E57 Material", "E54 Dimension", and "E58 Measurement Unit". Furthermore, MUJ contains a more explicit reference to the created place of an object, what thus is modeled as "E53 Place" and is connected to a "E12 Production" event via the property "P7i witnessed".

3. JUHMP, CIDOC-CRM and Omeka-S

Omeka S:

- Open-source web publishing platforms for sharing digital collections;

- Linked data, ontology and SPARQL support;

- API or CSV import of data, but does not support integration of linked items from an RDF file;

- CIDOC-CRM is not one of Omeka S pre-built ontologies (though installation is easy);

- does not allow for displaying fields with values nested in another item;

- does not allow customization or filtering of fields according to some nested value (e.g. a connected "E55 type" class)

3. JUHMP, CIDOC-CRM and Omeka-S

First **JUHMP** prototype**:**

- Records for 24 notable individuals in the history of the Jagiellonian University manually selected;

- Data retrieved from from CAC and CS and integrated with their corresponding resources from JDL and MUJ;

- Data modelled according to CIDOC-CRM and imported into the Omeka S via API (custom script);

 Three interfaces for data exploration:

- A browsing and searching interface based on the 24 records.

- A SPARQL endpoint for querying the CIDOC-CRM-based graph.

- A SANTé Semantic Search Engine for enhanced data retrieval.

3. JUHMP, CIDOC-CRM and Omeka-S

First **JUHMP** prototype**:**

- Records for 24 notable individuals in the history of the Jagiellonian University manually selected;

- Data retrieved from from CAC and CS and integrated with their corresponding resources from JDL and MUJ;

- Data modelled according to CIDOC-CRM and imported into the Omeka S via API (custom script);

 Three interfaces for data exploration:

- A browsing and searching interface based on the 24 records.

- A SPARQL endpoint for querying the CIDOC-CRM-based graph.

- A SANTé Semantic Search Engine for enhanced data retrieval.

Live demonstration of first prototype

Thank you for your attention!

Comments?