

# HunCRIS – towards semantic interoperability of CRIS-es

**Adam Tichy-Racs**

*atichy-racs@omikk.bme.hu*

Hungarian National Research Registry Unit  
of the National Technical Information Centre and Library  
at the Budapest University of Economics and Technology  
**(BME OMIKK)**



# Contents

- What is HunCRIS
  - Legal background
  - CERIF in HunCRIS
  - Workflow
- Services of HunCRIS
- Semantic features of HunCRIS
  - Thesauri and list elements
- Project proposal

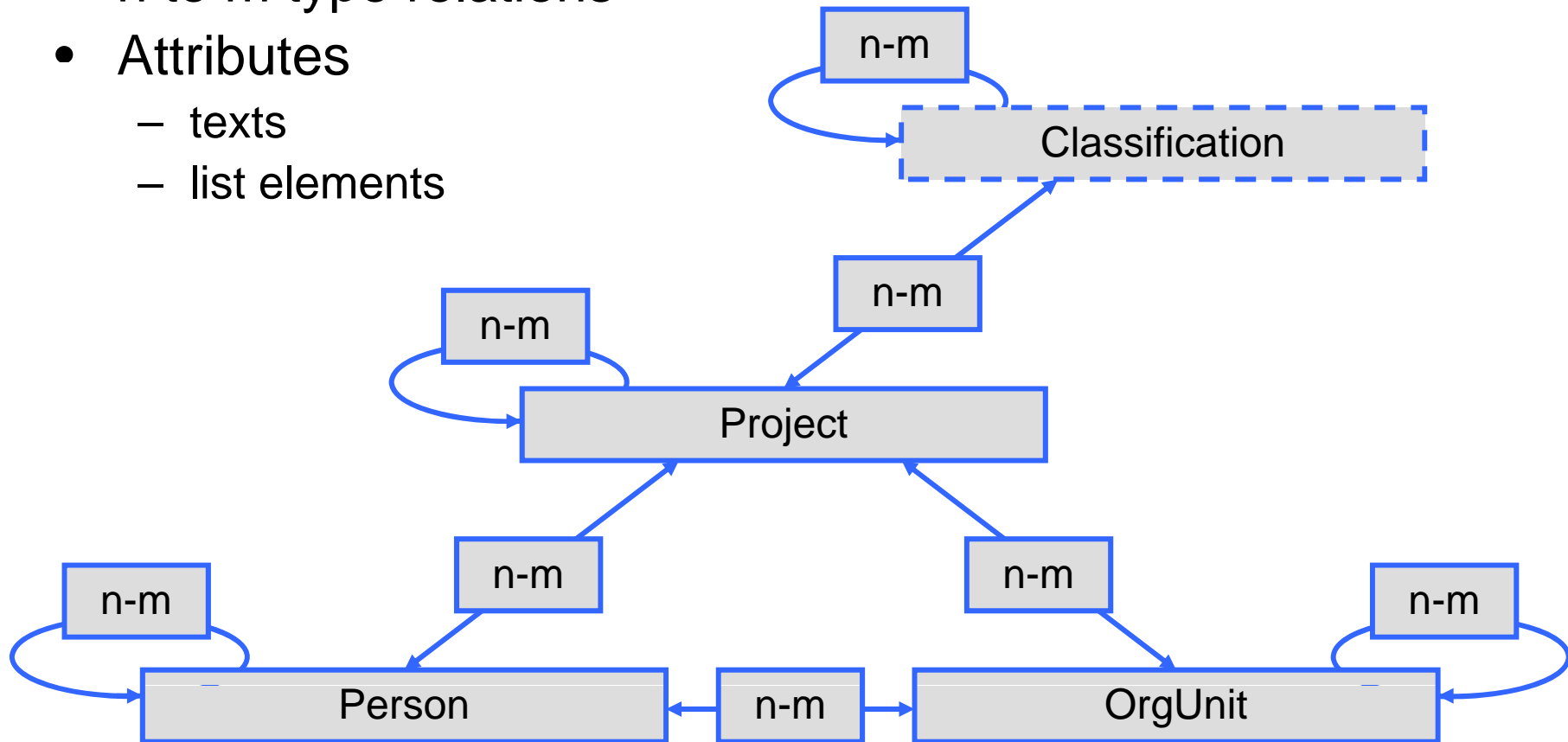
# HunCRIS in the Hungarian National Innovation System

- Information system of publicly financed Hungarian RTD projects, established by Hungarian Government Resolution No. 160/2001
- Adopted Commission Recommendation of 6 May 1991 concerning the harmonization within the Community of research and technological development databases (91/337/CEE)
  - Annex I. CERIF
  - Annex II. a multilingual and multidisciplinary common classification system
- HunCRIS is
  - Owned by the National Office for Research and Technology (NKTH)
  - Operated by BME OMIKK



# CERIF data model

- Entities  
Project, Person, OrgUnit
- n to m type relations
- Attributes
  - texts
  - list elements



# Software and basic workflow in HunCRIS

- CERIF–2000 compatible structure
  - Concentrating on Project entity
  - Upgrades with CERIF–2002, –2004 and –2006 versions
- Workflow under JSP application
  - Data upload
  - Data checking, validation
  - Identification of pre-recorded Persons and OrgUnits
  - Keyword processing
    - Uniform structure, bilingual equivalence
    - Positioning in the structure
- Authority management
  - Roles: data owner, HunCRIS team member, analyst, explorer
- Available at: <https://nkr.info.omikk.bme.hu>
- Software developer: IQSYS Co.



# Basic Services of HunCRIS 1

**Two level Boolean expressions**

**Boolean query**  
Condition groups

Item corresponds to all groups below

Free text or list element

Free Keywords matches all of the following

Equal matches all of the following

Not equal matches any of the following

Settings

Condition categories

Sorted by

Projects

Ascending

Time filter

2002  
2003  
2004  
2005  
2006  
2007  
2008  
2009  
2010  
2011  
2012  
2013

**Persons**

Result: 1-200/864

Agod Attila Péter  
Almássy Zsuzsa  
Andor György  
Anna Péter  
Antal Péter  
Antal Péter  
Aparvi Barnabás  
Aradi Péter  
Arató Péter  
Babarczy Anna  
Bácsi Katalin  
Bajor Gábor  
Balcz Péter  
Balázs Andrea  
Balázs Barbara  
Balázs László  
Balázs Marton  
Balázs Tibor  
Balint István  
Balint Péter  
Ballai Áron  
Balogh Attila  
Balogh Edina  
Balogh László  
Balogh Tibor  
Bánk András  
Bánky Tamás

**Projects**

Sorted by

The end year of the project

Number of items listed

Project names in alphabetic order

Ascending

2007 or later

20

**OrgUnits**

| Project name                            | Start date | End date   |
|---|------------|------------|
| Civilization and liver disease          | 2006.01.01 | 2008.12.31 |
| Fundamental researches in biotechnology | 2005.02.01 | 2008.12.31 |
| Modern data analysis and...             | 2005.02.01 | 2007.12.31 |

# Basic Services of HunCRIS 2

Navigation between projects

Project summary

Project description form

Organisation data form

Department of Process Engineering

Pannon University/Faculty of Engineering/

OrgUnit summary

Department of Process Engineering

Faculty of Engineering/

(unit) data

organisation (unit): Pannon University

organisation (unit): Faculty of Engineering

organisation (unit): Department of Process Engineering

OrgUnit details

organisation (unit) properties

organisation (unit) name in Hungarian

organisation (unit) name in English

Department of Process Engineering

organisation (unit) short name in Hungarian

FMT

organisation (unit) short name in English

Person details

Abonyi János a kémiai tudomány kandidátusa

Abonyi János a kémiai tudomány kandidátusa

Pannon University/Faculty of Engineering/ Department of Process Engineering

Participation of the person in the project

Start Year 2005

Finish Year 2007

Role

vezető kutató

Status (in the organisation (unit))

egyetemi docens

organisation (unit)

Department of Process Engineering

Pannon University/Faculty of Engineering/

Personal data

Family name

Project details

Title of project in Hungarian (maximum 500 characters)

Korszerű adatkezelési technikák és modell alapú algoritmusok

Title of project in English (maximum 500 characters)

Modern data analysis methods and model based algorithms in experiment design and evaluation

Short title of project in Hungarian (maximum 50 characters)

Korszerű adatkezelési és modell alapú technikák

Short title of project in English (maximum 50 characters)

Modern data analysis and model based techniques

Funding programme id

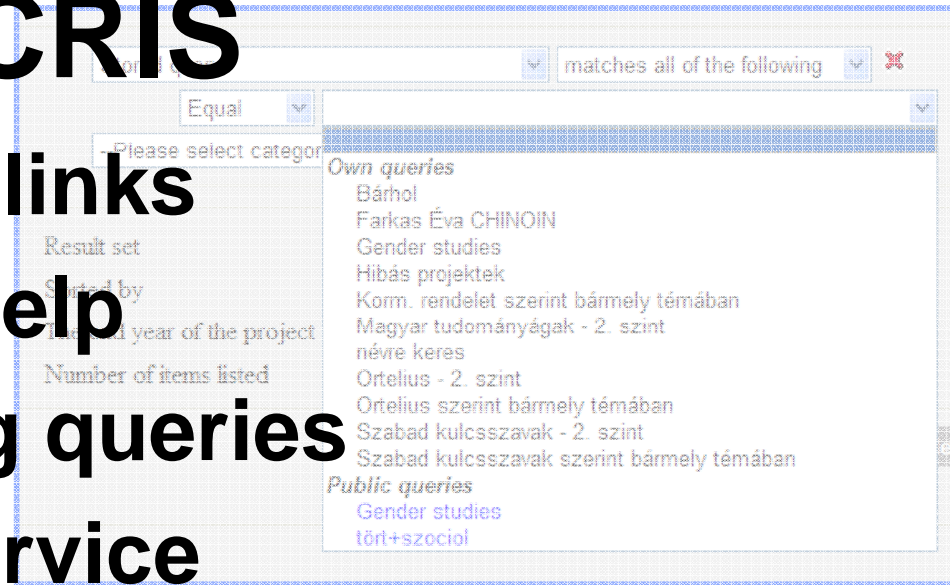
Országos Tudományos Kutatási Alapprogramok, OTKA T17 kutatási program

Contract id in the funding programme

T 049534

# Advanced Services of HunCRIS

- **Providing dynamic links**
- **Content sensitive help**
- **Storing and reusing queries**
- **Instant message service**
  - from registered users to researchers
- **Assistance to query building**
- **Editing documents with projects data that meet pre-set query conditions**



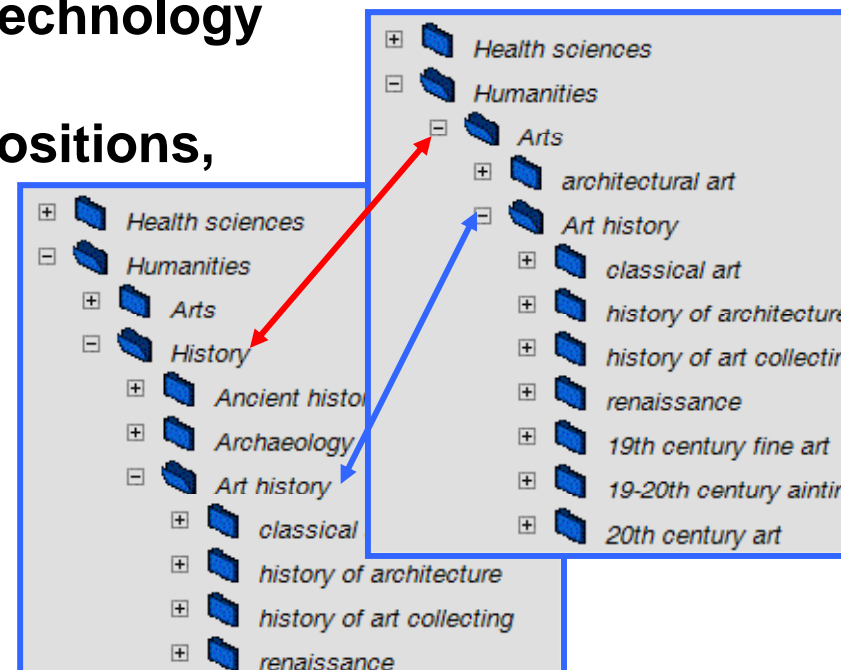


# Thesauri and controlled terms

- **Parallel structured lists of elements**
  - Hungarian national classification system (63 items)
  - Ortelius Thesaurus – classification system for Electronic Proposal Submission System (EPSS) (2067 items)
  - Structured set of key expressions provided with project descriptions (13663 items)
  - OECD's Fields of Science and Technology (not installed yet)
  - Roles, project positions, work positions, programmes, territorial units

- **Structure**

- Bilingual list elements
- Separated linking table
  - Multiple parents  
but no circular references



# Semantic features of HunCRIS

- Parallel use of vocabularies
- „Rebuild query” option (from project)
  - To search similar projects
    - Fine tuning of rebuilt Boolean query
- **Extended indexing**
  - Adding all parents of index terms
  - Advantages of extended indexing
    - „Position” of any project in knowledge space is determined by the index vector in the multidimensional space of concepts (vocabulary)
    - „Distance” and „angle” between projects can be calculated from index vectors
    - Metrics of similarity between any pair of projects or metrics of conformity between a query and a project can be calculated
    - Visual representations of knowledge space in one, two or three dimensions can be generated

Cognitive science, corpus, corpus of child language, Grammar, Humanities, language  
Language sciences, learning, Lexicology, Linguistics, mental grammar, Psycholinguistic  
Psychological sciences, Social sciences, statistical learning mechanism, Syntax

# A thought experiment – parallel use of different vocabularies

- There are three concurrent structured sets of expressions to describe each project in HunCRIS
- The descriptions of one of them looks like

Index set 1

Keywords provided by the project leader

Cognitive science, corpus, **corpus of child language**, Grammar, Humanities, **language acquisition**, Language sciences, learning, Lexicology, Linguistics, **mental grammar**, Psycholinguistics, Psychological sciences, Social sciences, **statistical learning mechanism**, **Syntax**

Index set 2

Scientific classification (according to the Annex of 169/2000. (IX. 29) Regulation of the Hungarian Government)

Humanities, **Informatics**, **Language sciences**, **Psychology**, Technological and engineering sciences

Index set 3

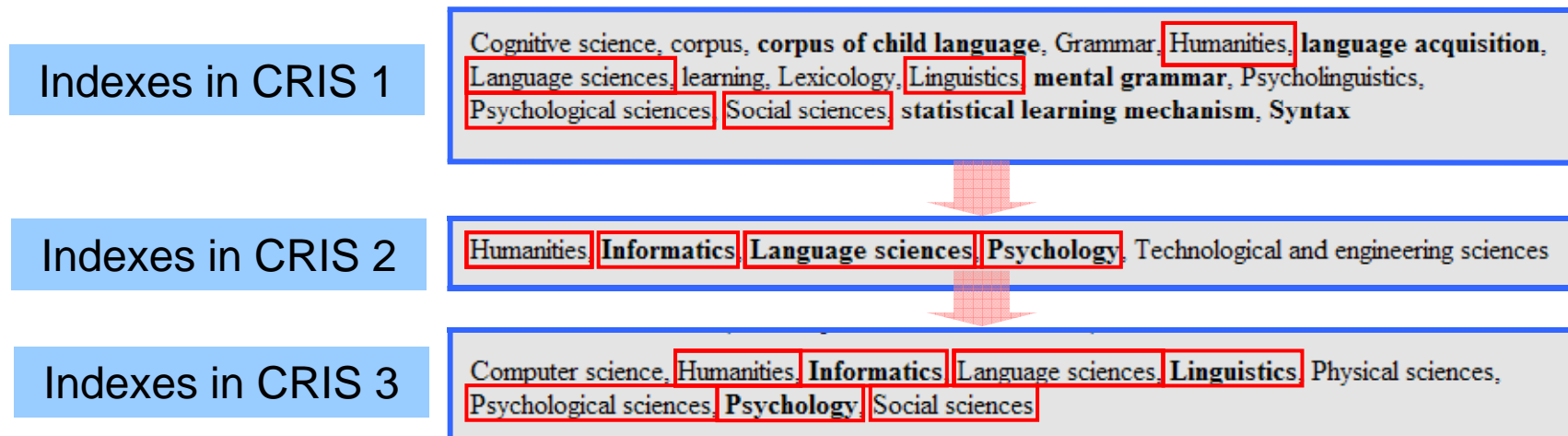
Scientific classification (according to the Ortelius thesaurus)

Computer science, Humanities, **Informatics**, Language sciences, **Linguistics**, Physical sciences, Psychological sciences, **Psychology**, Social sciences

- *Let us suppose that they describe the same project in three different CRIS-es, and start testing the level of semantic similarity!*

# Testing similarities of CRIS-es

- The accuracy of interoperated service



- depends on the difference between vocabularies
  - between index sets 1 and 2 is smaller than either between index sets 1 and 3 or index sets 2 and 3
- and on the number of steps from CRIS to CRIS
  - if CRIS 3 looks up data from CRIS 1 via CRIS 2, the quality is lower than in case of direct linking

# Recommendations for project using semantic interoperability

- **XML wrapper with extended indexes locally**
- **Portal software with**
  - **Metadata library of all involved CRIS-es to ensure syntactic interoperability**
  - **Optimal use of local resources (vocabularies)**
  - **Value added semantic, analytic and visualization tools**
- **Installing portal to all participating CRIS-es in order to**
  - **Gain full power of tools on data of local CRIS-es**
  - **Adopt and emphasise local priorities**
  - **Support for customizing services**
  - **Make the business model sustainable**