

Building cross-field national data repository

Miroslav Šimek

Lukáš Košťenský



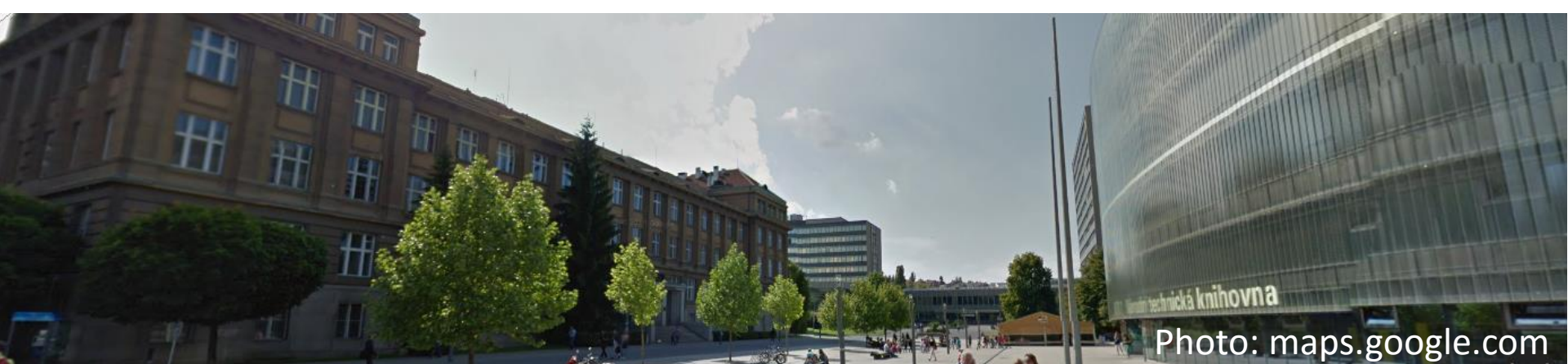
UNIVERSITY OF
CHEMISTRY AND TECHNOLOGY
PRAGUE

NTK

50°6'14.083"N, 14°23'26.365"E
Národní technická knihovna
National Library of Technology

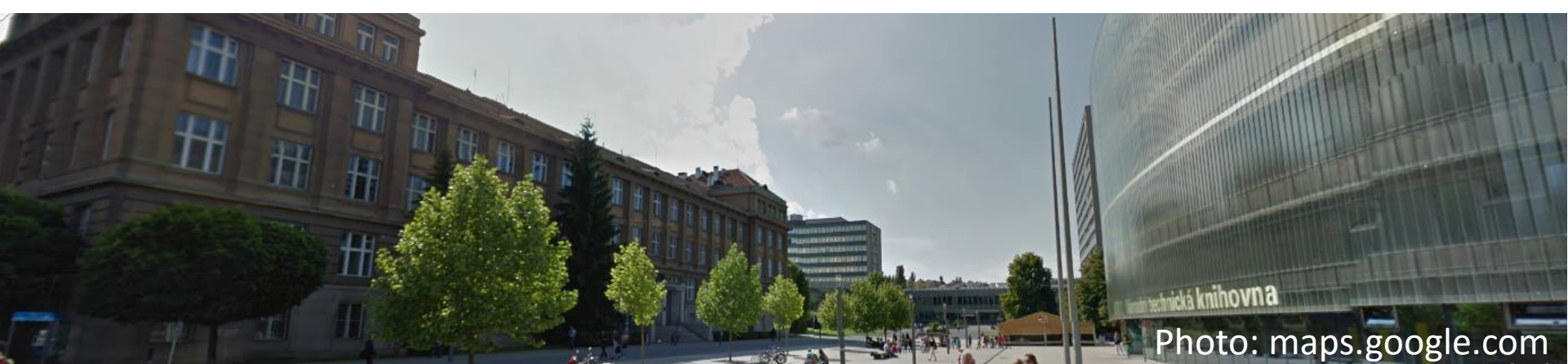
Why NTK and UCT Prague?

- Centre for Information Services, UCT Prague
 - University libraries
 - UCT publishing house
 - IT support



Why NTK and UCT Prague?

- National Library of Technology (NTK)
 - The largest and the oldest library of science and technology literature in the Czech Republic
 - Who else should know everything about metadata for scientific data



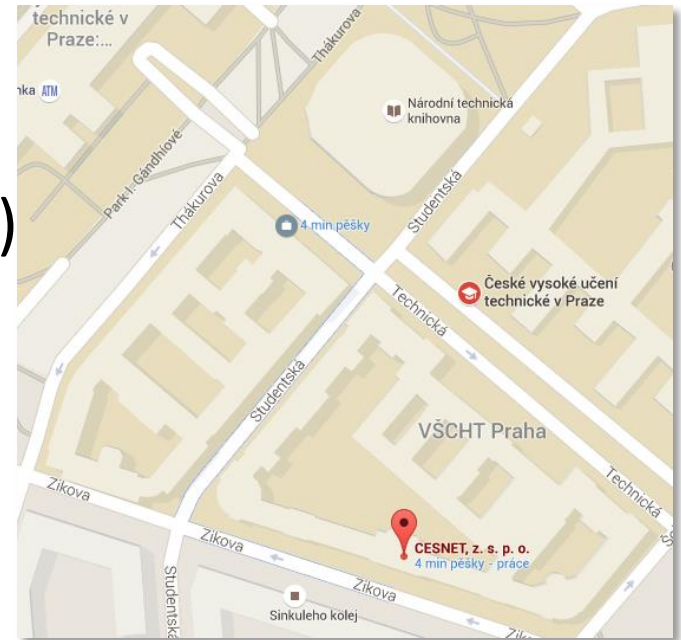
Why NTK and UCT Prague?

- In 2013, CIS and NTK joined forces and founded ChemTK, a shared library of chemistry, sharing not only books but also e-sources and people
- Department of Organic Chemistry of the Czech Academy of Sciences joined ChemTK in 2015



Why CESNET?

- An association of universities of the Czech Republic and the Czech Academy of Sciences
- Operates and develops the national e-infrastructure for science, research and education
- encompasses
 - computer network
 - identity management (eduid.cz)
 - **data storage (21PB of space)**
 - collaborative environment
- Level of trust



Source: maps.google.com

Cross-field repository

- Initial group of users – science
 - Department of Organic Chemistry, UCT Prague
 - Romany Studies, Faculty of Arts, Charles University
- Used as normal document repository
 - UCT Prague
 - institutional repository (qualification works, articles, other documents)
 - media repository

Department of Organic Chemistry

Department of Organic Chemistry



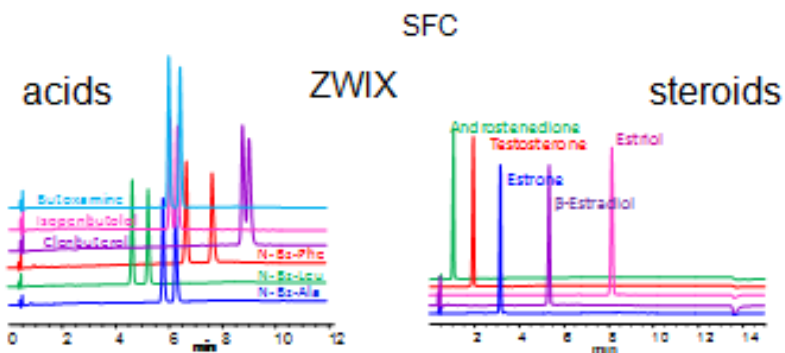
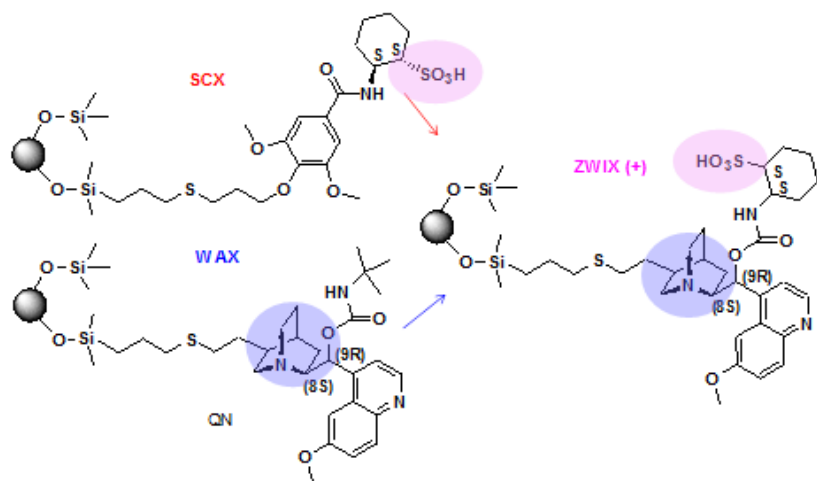
Source: uoch.vscht.cz

Department of Organic Chemistry

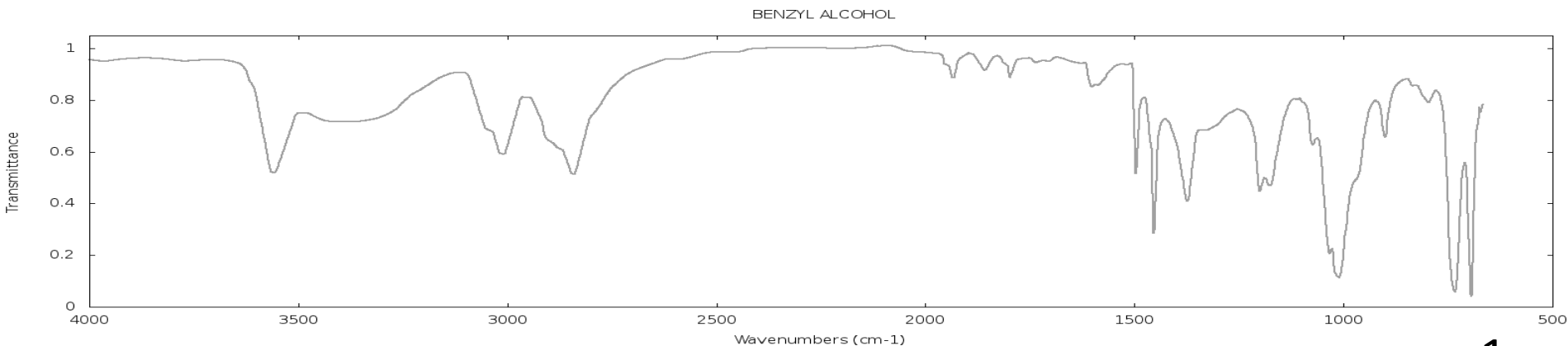


Source: uoch.vscht.cz

Department of Organic Chemistry

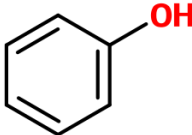


Source: bafa.vscht.cz



Department of Organic Chemistry

- Carefully log the source compounds, reactions, reaction conditions and tools used, synthesized products and the data coming from spectroscopic measurements, NMR, chromatography and other analytical methods
- Want to
 - have the data safe, backed-up
 - make sure that the data are searchable and available for all the team members, regardless of their geographical location

- Make sure that the data are searchable
 - Give me data about chemical compounds
 - with  substructure
 - Give me all data containing a sharp peak in IR spectrum in the region of 2500-3000

Department of Organic Chemistry

- Do not want to publish the data at the time of measurement because:
 - They are writing an article about the reactions/compounds/data (publish or perish)
 - They want to patent the reactions
 - They are members of the Laboratory of New Psychotropic Drugs and their results can't be published in detail at all

Romany Studies, Faculty of Arts, Charles University



Source: Multi-disciplinary approaches to Romany Studies,
CEU 2007-2009

Romany Studies, Faculty of Arts, Charles University

- Need to store recordings of interviews with Roma people from the Czech Republic and Slovakia together with different searchable metadata
 - Technical (where, when, data format, ...)
 - Personal (who, personal relations, ...)
 - Transcriptions, interview topics, tags, ...
- Problems:
 - Both data (recordings) and metadata contain personal information that should not be published beyond the team members
 - Some parts of the recordings may contain descriptions of activities that could not be disclosed at the moment, should not be available even to the team but might be disclosed later

Romany Studies, Faculty of Arts, Charles University

- From the data point of view, one resource (in this case a video recording) should have several manifestations:
 - Full version, archived but not available to anyone but the author
 - Version with some of the information stripped/blackened for the team members
 - Even more stripped version for public use
- DTTO for metadata
 - Highest protection: SSN and other identifiers
 - Medium protection: date of birth
 - No protection: age
- The sets of data and metadata with a given level of protection changes in time

Comparison

	Chemists	Romany Studies	Classical libraries
Data formats	<p>Mainly specialized data formats:</p> <p>Molecules, reactions, spectra, other measurements</p>	<p>Mainly annotated multimedia:</p> <p>Videos, metadata about people, personal relations</p>	<p>Classical DCTerms as you know them:</p> <p>Title, Author, Abstract, ...</p>
Search	Substructure search, searching by features in graphs, in metadata ann.	Searching in metadata annotations	Searching in metadata annotations
Data amounts	Huge (in case of computational chemistry)	Large	If not storing digitized data, moderate
Security	Very well defined and enforced	Very well defined and enforced, to the level of single metadatum	“Everything should be public, open access is the goal”
Data entry	Do not bother me with that	I know I have to add subtitles	Metadata specialists

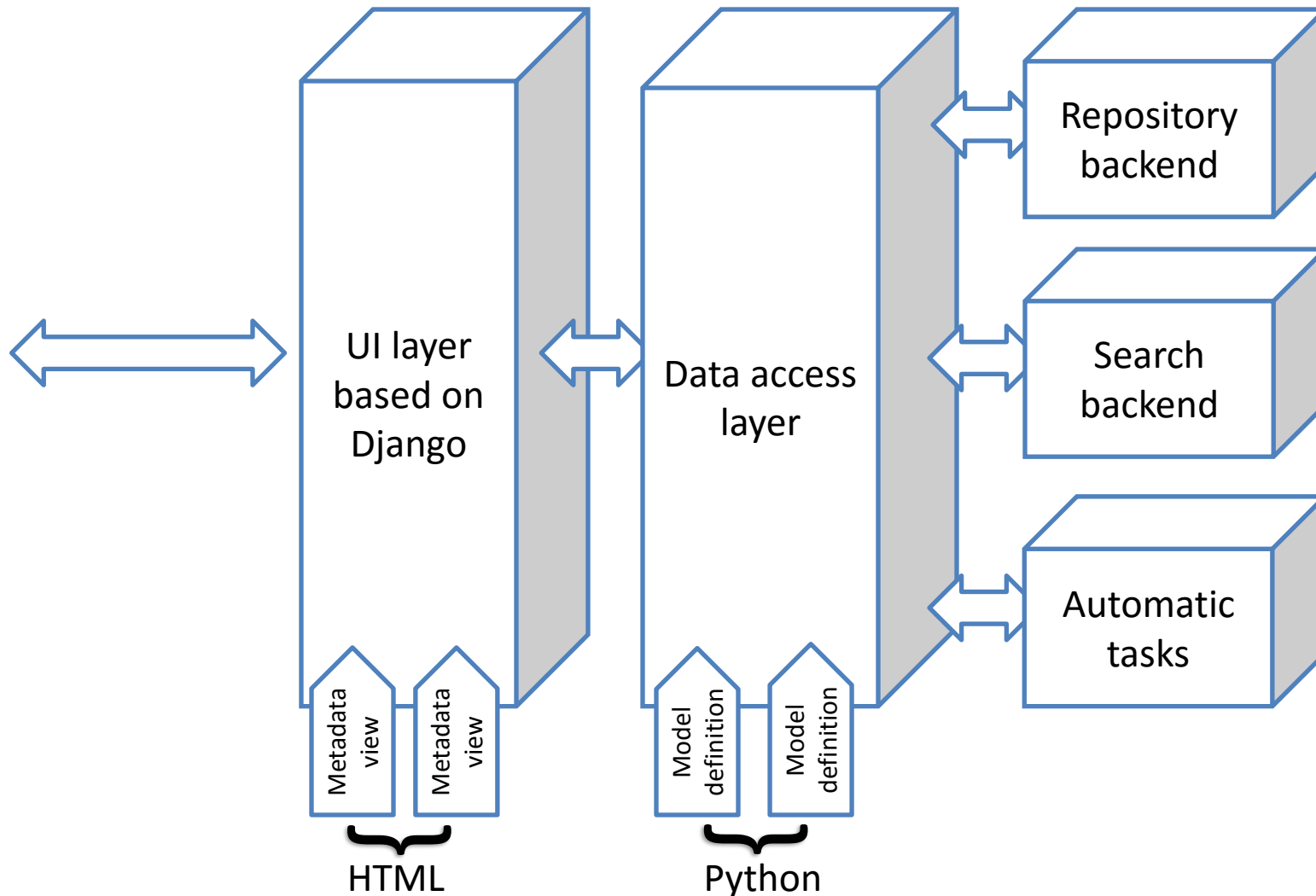
The Quest

- Data repository containing diverse data and associated metadata and providing customizable UI for searching, viewing and editing the data
- Easy to use tools for data filling
 - For chemistry, automatically send laboratory data to the server, run algorithms on measured data to identify and index search features
 - For Romany studies, have a mobile application for audio/video capture together with tagging and transcription tools, sending the data directly to the repository

Cross-field repository

- Reuse repository backend, do not write our own
- Reuse search, do not write our own
- Keep the design modular
 - adding new field should not mean changing the implementation of previous fields
- Keep the design open for interdisciplinary data
- Let it be in Python

Architecture



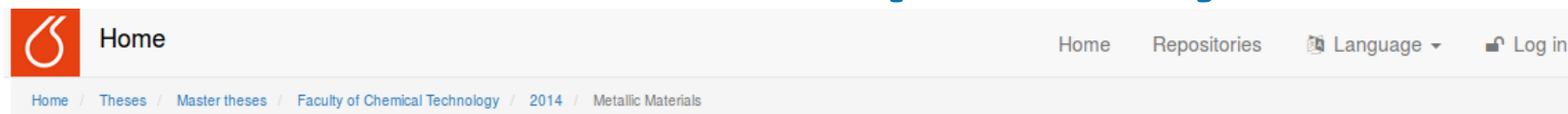
Cross-field repository

- Repository backends
 - DSpace
 - Invenio
 - Fedora
- Search engines
 - SOLR
 - Elastic Search

Cross-field repository

- Current status
 - Implemented proof of concept – institutional repository of UCT Prague which is now in production
 - Implemented proof of concept – repository for Romani studies
 - Initial steps in OCH data
- Waiting for Screenshots?

Look and Feel – classical repository



Master theses defended in academic year 2014/15 Faculty of Chemical Technology

Kompozitní materiály na bázi intermetalických sloučenin

Author: Ing. Milan Valalík
Supervisor: doc. Ing. Pavel Novák, Ph.D.



Metalurgické metody pro zpracování odpadu s obsahem chromu z výroby niklu

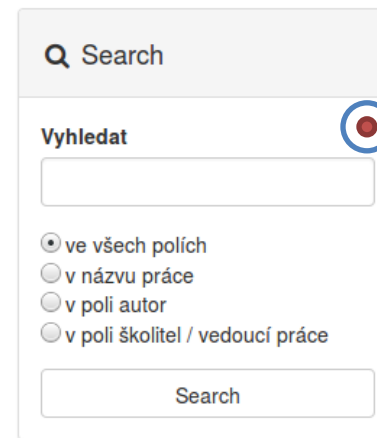
Author: Ing. Tomáš Frýdl
Supervisor: Ing. Hong Vu, Ph.D.

Příprava a vlastnosti nanostrukturovaného povrchu slitin TiNbZr


Author: Bc. Matyáš Peterka
Supervisor: Ing. Jaroslav Fojt, Ph.D.

Struktura a vlastnosti slitiny hořčíku s kovy vzácných zemin

Author: Ing. Jan Luštinec
Supervisor: prof. Dr. Ing. Dalibor Vojtěch

The search sidebar is located on the right side of the page. It has a 'Search' input field at the top. Below it is a section titled 'Vyhledat' (Search) with a red circular icon. Under this section are four radio button options: 've všech polích' (selected), 'v názvu práce', 'v poli autor', and 'v poli školitel / vedoucí práce'. At the bottom of the sidebar is a 'Search' button.

Look and Feel – classical repository

 Home

Home Repositories Language ▾ Log in

Home / Theses / Master theses / Faculty of Chemical Technology / 2014 / Metallic Materials / 20052

Repository of theses, UCT Prague - thesis detail

Title in the language of thesis

Kompozitní materiály na bázi intermetalických sloučenin



Author

Ing. Milan Valalik

Title in Czech

Kompozitní materiály na bázi intermetalických sloučenin

Title in English

Composite materials based on intermetallic compounds

Supervisor

doc. Ing. Pavel Novák, Ph.D.

Opponent(s)

Ing. Pavel Hanus, Ph.D.

Consultant(s)

Ing. Jan Kříž

Abstract in Czech (scanned automatically, the student's version is in the box on the right)

Abstrakt Práce se zabývá přípravou kompozitních materiálů s matricí tvořenou intermetalickou fází. Materiály jsou zamýšleny jako možná náhrada za slinuté karbidy WC-Co v oblasti obráběcích nástrojů. Prášky intermetalických sloučenin NiTi a TiAl, které tvoří matrici kompozitu, jsou připraveny a kompaktizovány práškovou metalurgií. Výztuž v podobě prášků



Download documents

[Abstract in Czech](#)

[Abstract in English](#)




Online fulltext is accessible only to authenticated users. Thesis in printed form is stored in the library of department number 106.

Look and Feel – Romany studies



Recordings

 Search

Text to search

Sort by title ▼

SEARCH

Add a Record

Romové v Evropě



Užaras pro autobus (Čekání na autobus)



Bydlení na Luniku IX



Look and Feel – Romany studies



Romové v Evropě

Download video

- <http://127.0.0.1:8080/fcrepo/rest/romiste/recordings/3b/d8/0a/f4/3bd80af4-0d54-4cde-b37a-8aa843c05ec7/4d/1a/8c/79/4d1a8c79-844d-46f0-9af9-2d0acc85d927>

State

Author is working on data

[Back to List](#)

Title

Romové v Evropě

Creation Place

M50, Ireland

Keywords

Romové v Evropě

Description

Z cyklu evropských televizí odhalujícího kontrasty mezi životem Romů na různých místech starého kontinentu. Rumunsko-irský dokument sleduje osudy Stoicy Rostase, který byl po likvidaci ilegálního romského tábora v irské dálnici M50 deportován zpět do Rumunska (2010)

Respondent

Stoica Rostas

Circumstance

Deportation of Romanians from illegal village near M50 highway in 2007

Recording Type

Video

Technical Notes

25minutes, 2010, Documentary



[Edit](#)

Look and Feel – Romany studies



Romové v Evropě

Download video

- <http://127.0.0.1:8080/fcrepo/rest/romiste/recordings/3b/d8/0a/f4/3bd80af4-0d54-4cde-b37a-8aa843c05ec7/4d/1a/8c/79/4d1a8c79-844d-46f0-9af9-2d0acc85d927>

State

Author is working on data

◀ Back to List

Title

Romové v Evropě

Creation Place

M50, Ireland

Keywords

Romové v Evropě

Description

Z cyklu evropských televizí odhalujícího kontrasty mezi životem Romů na různých místech starého kontinentu. Rumunsko-irský dokument sleduje osudy Stoicy Rostase, který byl po likvidaci ilegálního romského tábora v irské dálnici M50 deportován zpět do Rumunska (2010)

Respondent

Stoica Rostas

Circumstance

Deportation of Romanians from illegal village near M50 highway in 2007

Recording Type

Video

Technical Notes

25minutes, 2010, Documentary

Edit

Look and Feel – Romany studies

Optional data

Middle name(s)

Nickname

Primary language

Birthplace



Date of Death

Place of Death



Phone number



[+ Add a new value](#)

Web page

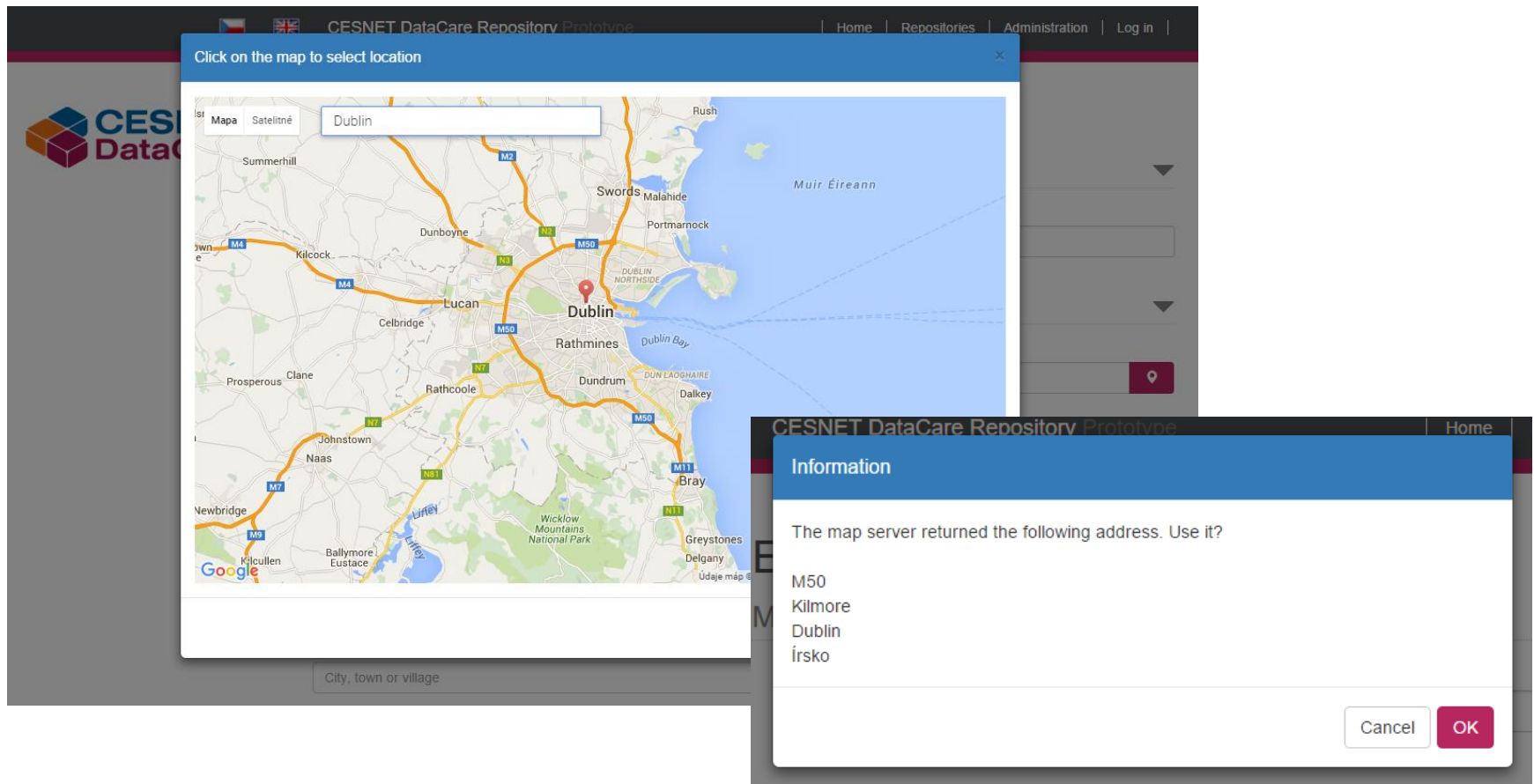


[+ Add a new value](#)

Photo



Look and Feel – Romany studies



Look and Feel – Spectroscopic data


Infrared Spectra

Search

Alcohols/Phenols


Sort by title

Structure search



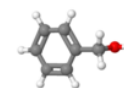
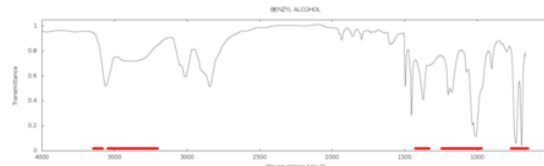
Marvin JS

ChemDraw

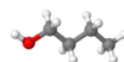
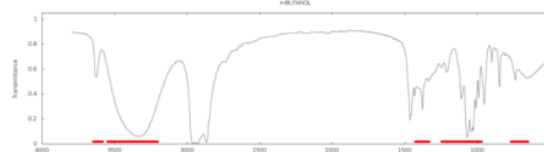


SEARCH

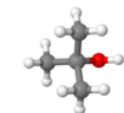
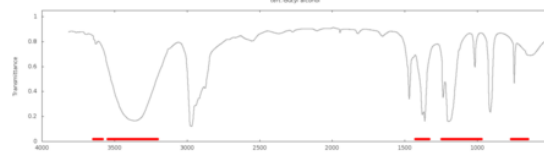
Add an Infrared Spectrum



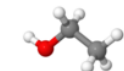
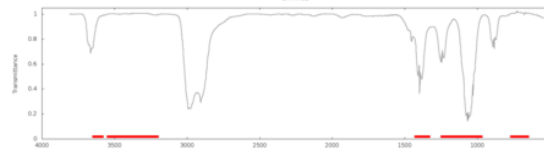
BENZYL ALCOHOL
 PERKIN-ELMER 21
 (GRATING)
 1964/03/21
 COBLENTZ SOCIETY



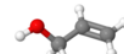
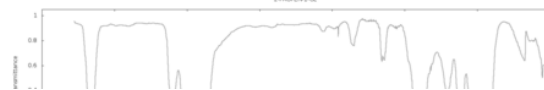
n-BUTANOL
 PERKIN-ELMER 521
 (GRATING)
 1966/02/03
 COBLENTZ SOCIETY



tert.-Butyl alcohol
 PERKIN-ELMER 521
 (GRATING)
 1965/02/22
 COBLENTZ SOCIETY

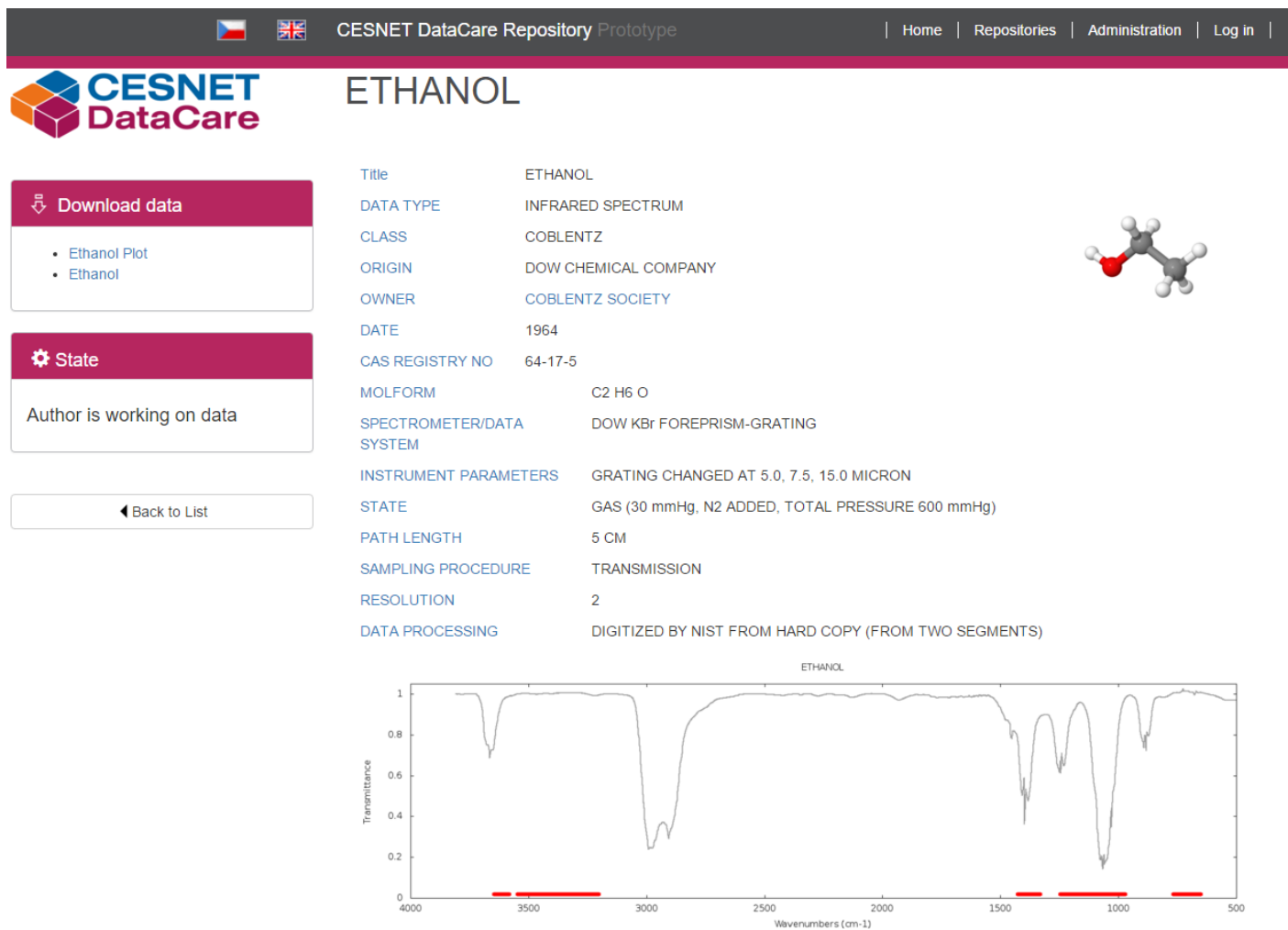


ETHANOL
 DOW KBr FOREPRISM-
 GRATING
 1964
 COBLENTZ SOCIETY

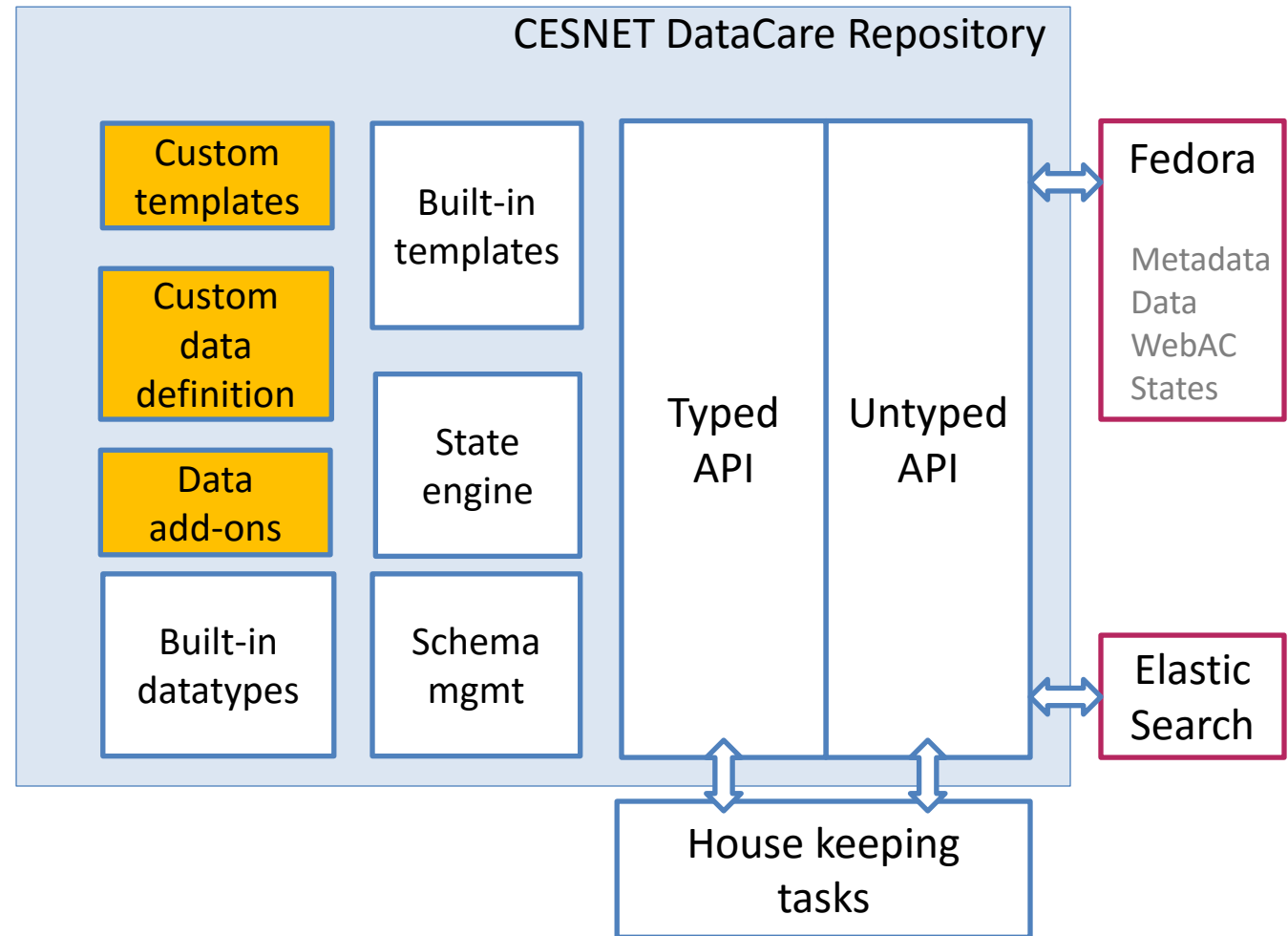


2-PROPEN-1-OL
 BECKMAN IR-9 (GRATING)
 1964/03/11

Look and Feel – Spectroscopic data



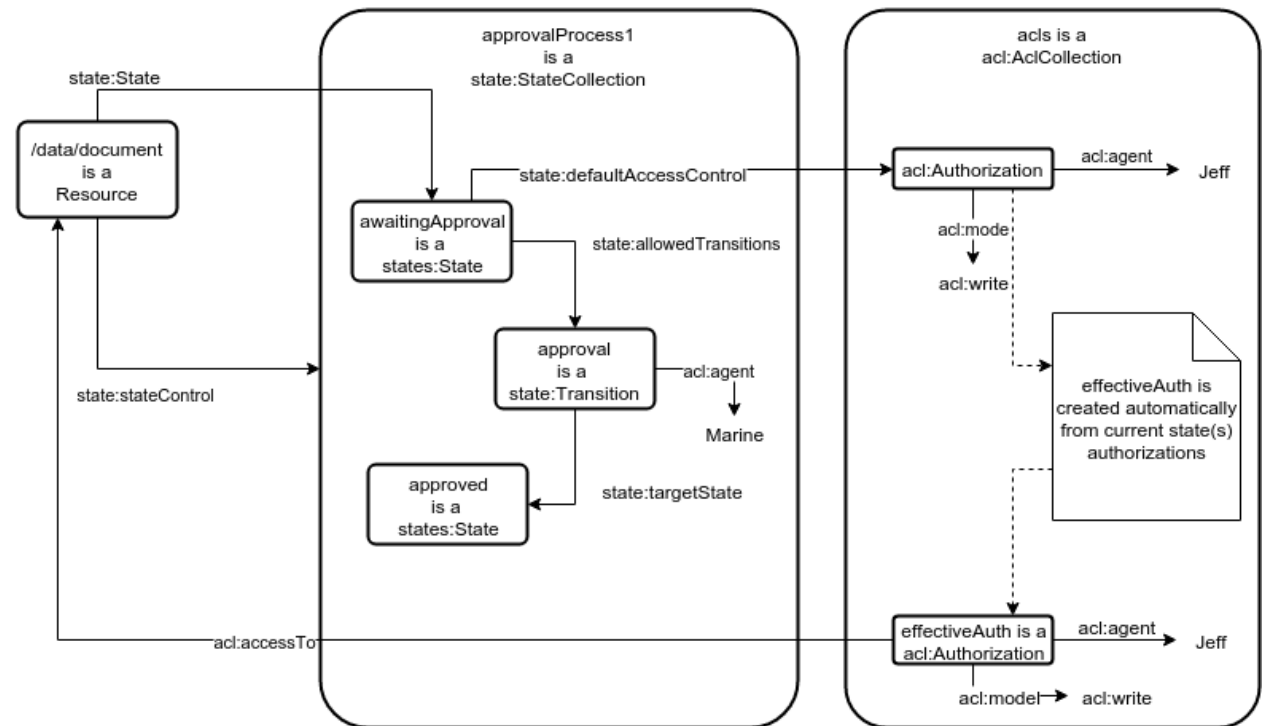
Architecture II



Architecture II - states

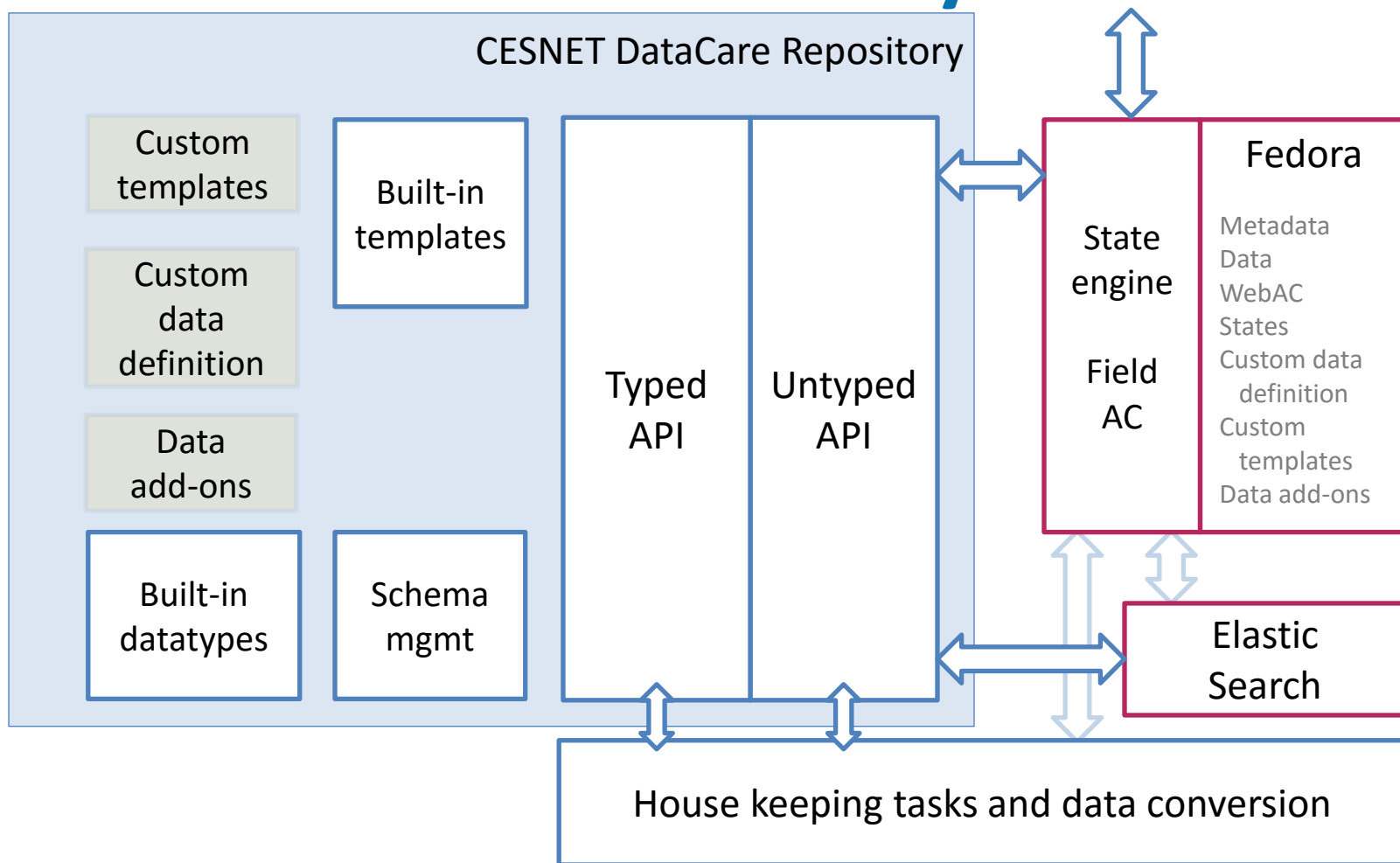
```
@prefix acl: <http://www.w3.org/ns/auth/acl#>.
@prefix state: <...>.
@prefix dc: <http://purl.org/dc/elements/1.1/>.

<> acl:accessControl </fcrepo/rest/acl>;
state:stateControl </fcrepo/rest/states>;
state:state </fcrepo/rest/states/wg1_approving>,
          </fcrepo/rest/states/wg2_approving>;
dc:title "Hello, World!".
```



Architecture II

end of the year



Cross-field repository

- This year
 - Finalize repository for Romani studies
 - Finalize proof of concept for OCH data
 - Start working on media repository at UCT Prague and its connection to e-learning and online teaching materials
- Following steps
 - Other fields



Special thanks to

Fedora
4



django



bleach
django-bootstrap-
pagination
bootstrap

inflection
rdflib
jquery

isodate
requests
bootstrap3-dialogs

pyparsing
python-dateutil
Google APIs