

Platforma Invenio

Repozitáře v NRP založené na platformě Invenio

miroslav.simek@cesnet.cz



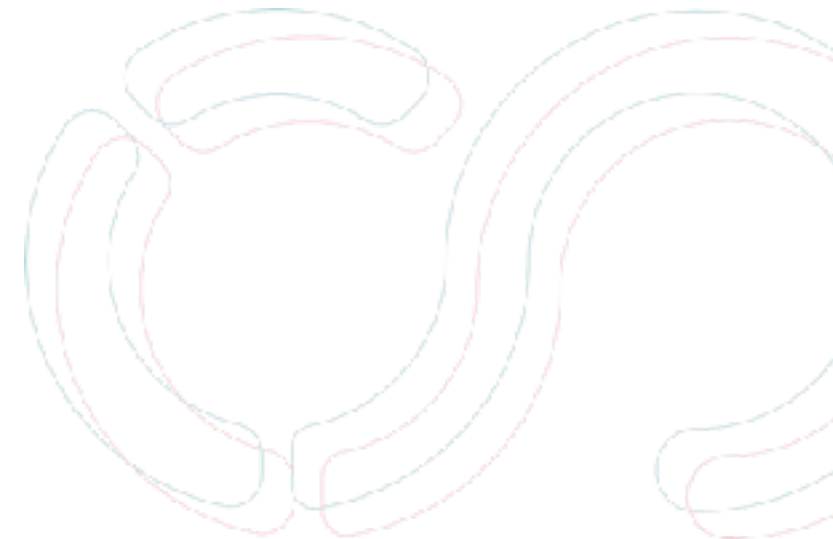
Spolufinancováno
Evropskou unií



MINISTERSTVO ŠKOLSTVÍ,
MLÁDEŽE A TĚLOVÝCHOVY

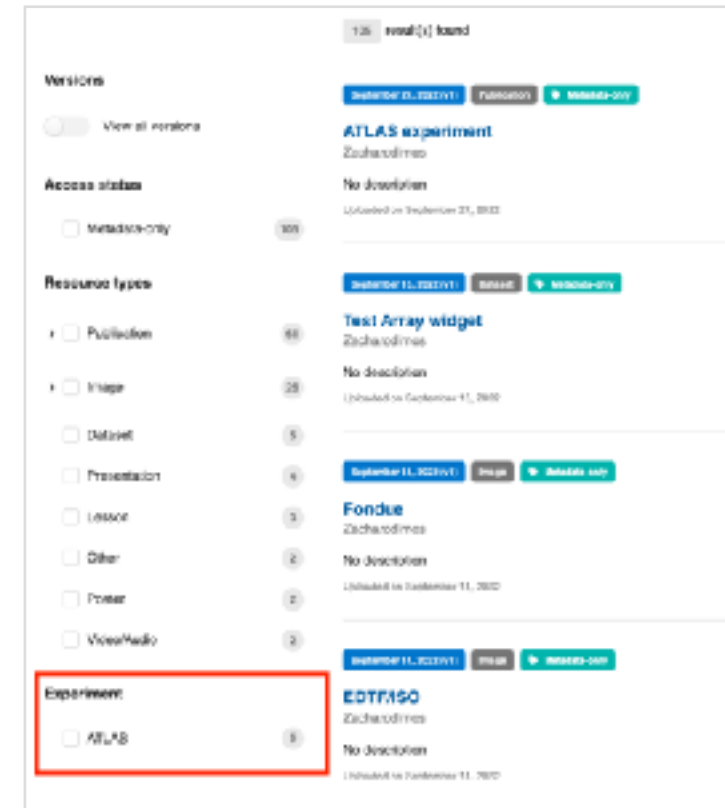
Platforma Invenio

- Obecná repozitářová platforma vyvíjená v Evropské organizaci pro jaderný výzkum (CERN)
- Invenio RDM (research data management)
 - verze platformy určená pro uchovávání výzkumných dat
- Repozitář Zenodo (zenodo.org) je založen na Invenio RDM 12



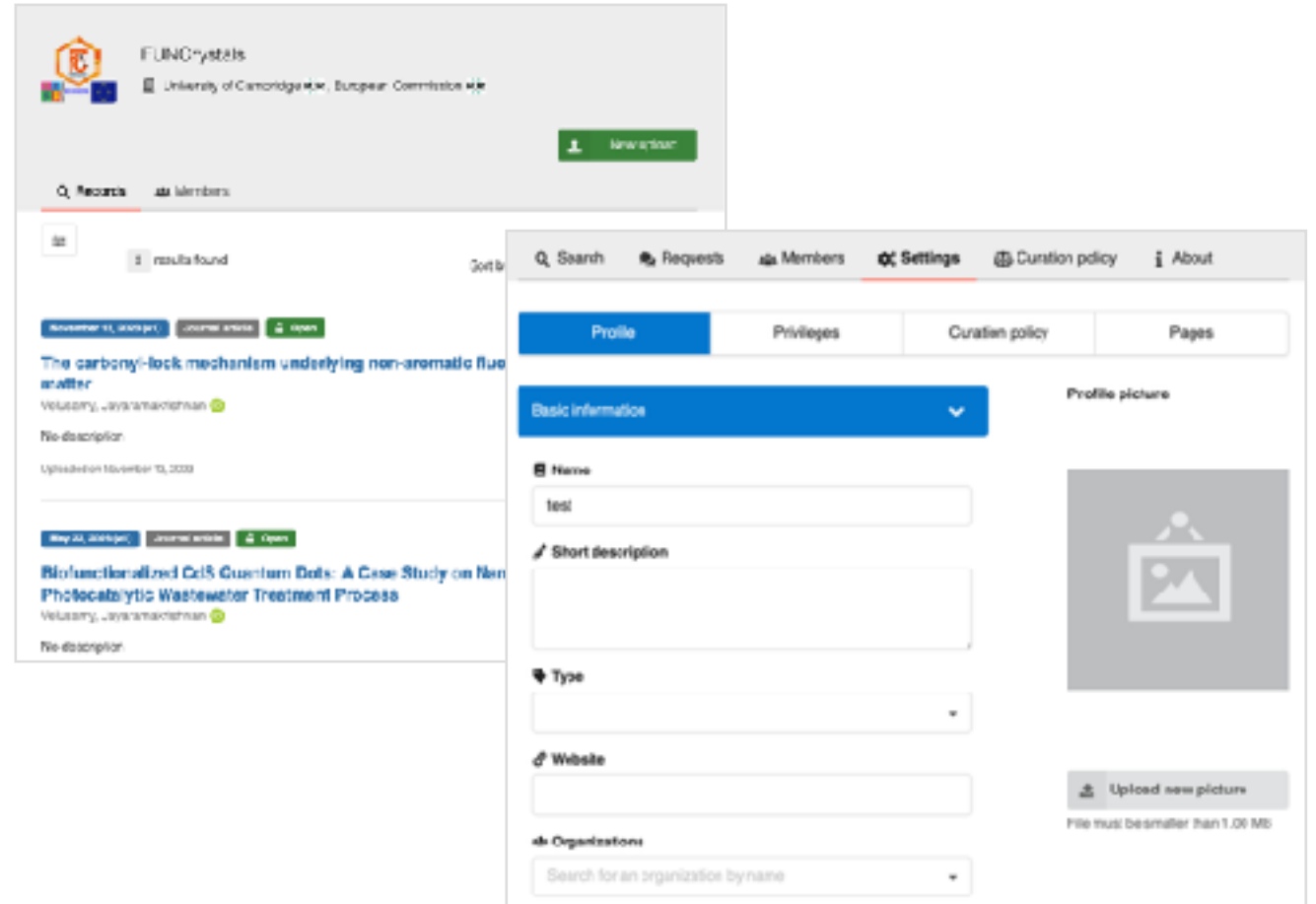
Vlastnosti platformy - obecné

- Ochrana digitálních objektů (přepis, smazání) uložených v S3 úložišti
- Metadata a vyhledávání v nich založené na OpenSearch
- Verzování metadat a digitálních objektů (bitstreamů)
- Validace vkládaných metadat a bitstreamů
- Exportní formáty pro metadata, OAI-PMH, podpora pro CSL
- Transformace a extrakce digitálních objektů
- Optimalizace pro vyhledávače / roboty
- Přidělování persistentních identifikátorů (DataCite DOI)



Vlastnosti platformy - specifické

- Navržena pro
 - velké množství záznamů
 - velké digitální objekty
- Pokročilá úroveň zabezpečení metadat i dat
 - dělení uživatelů do komunit
 - role v rámci komunity opravňující přístup k záznamům
 - workflows (schvalovací procesy)
 - příprava na zabezpečení jednotlivých polí
 - podpora embarga
 - secret links
 - napojení na e-INFRA AAI



Vlastnosti platformy - specifické

- Modifikovatelné uživatelské rozhraní
 - barvy, loga, vlastní komponenty, dynamické části (grafy, editační prvky, ...)
- Hierarchický metadatový model s využitím kontrolovaných slovníků a možností vazeb mezi modely
- Prostředky pro napojení na externí authority (např. NCBI taxonomy)
- Extenze metadatového modelu o vlastní části v konfiguraci
- API-first návrh garantující strojový přístup k téměř všem funkcionalitám repozitáře (JSON REST API)

Hierarchický metadatový model

- Záznam v repozitáři obsahuje výsledky screeningu/ů - interakce sondy se zkoumanou látkou
- Dataset záznamu jsou raw data z měřicího přístroje
- Dotazy:
 - vrať záznamy, kde jsme zaznamenali hit u sondy s id 456 a proteinu ABC



<https://www.chemdiv.com/drug-discovery-services/discovery-biology/hts/>

Hierarchický metadatový model

- Pouze 1 protein a sonda - jednoduché

D

Dataset name	Probe1 CAS	Protein1 ID	Result	Depositor name
A				

- 2 proteiny/sondy

Dataset name	Probe1 CAS	Protein1 ID	Result1	Probe2 CAS	Protein2 id	Result2
A						

Hierarchický metadatový model

- Jak lépe prezentovat? Zavedení hierarchie

Dataset

Dataset name	Well 1			Well 2			Depositor	
	Probe CAS	Protein ID	Result	Probe CAS	Protein ID	Result	Name	ORCID
A								
B								
C								

- Co když bude jamek 96?

Hierarchický metadatový model


- Jak lépe prezentovat (pro člověka)?

Dataset name	Well			
	Well #	Probe CAS	Protein ID	Result
A	1			
	2			
	3			
	4			
B				

- Invenio repozitář je na hierarchiích postaven

Invenio JSON

```
{  
  "name": "A",  
  "wells": [  
    {  
      "no": 1,  
      "probeCAS": 123,  
      "proteinId": 456,  
      "result": true  
    },  
    {  
      "no": 2,  
      "probeCAS": 234,  
      "proteinId": 456,  
      "result": false  
    }  
  ]  
}
```

Invenio JSON - alternativní model

```
{  
  "name": "A",  
  "hits": [  
    {  
      "no": 1,  
      "probeCAS": 123,  
      "proteinId": 456  
    }  
  ],  
  "misses": [  
    {  
      "no": 2,  
      "probeCAS": 234,  
      "proteinId": 456  
    }  
  ]  
}
```

hits.probeCAS:123 AND hits.proteinId:456

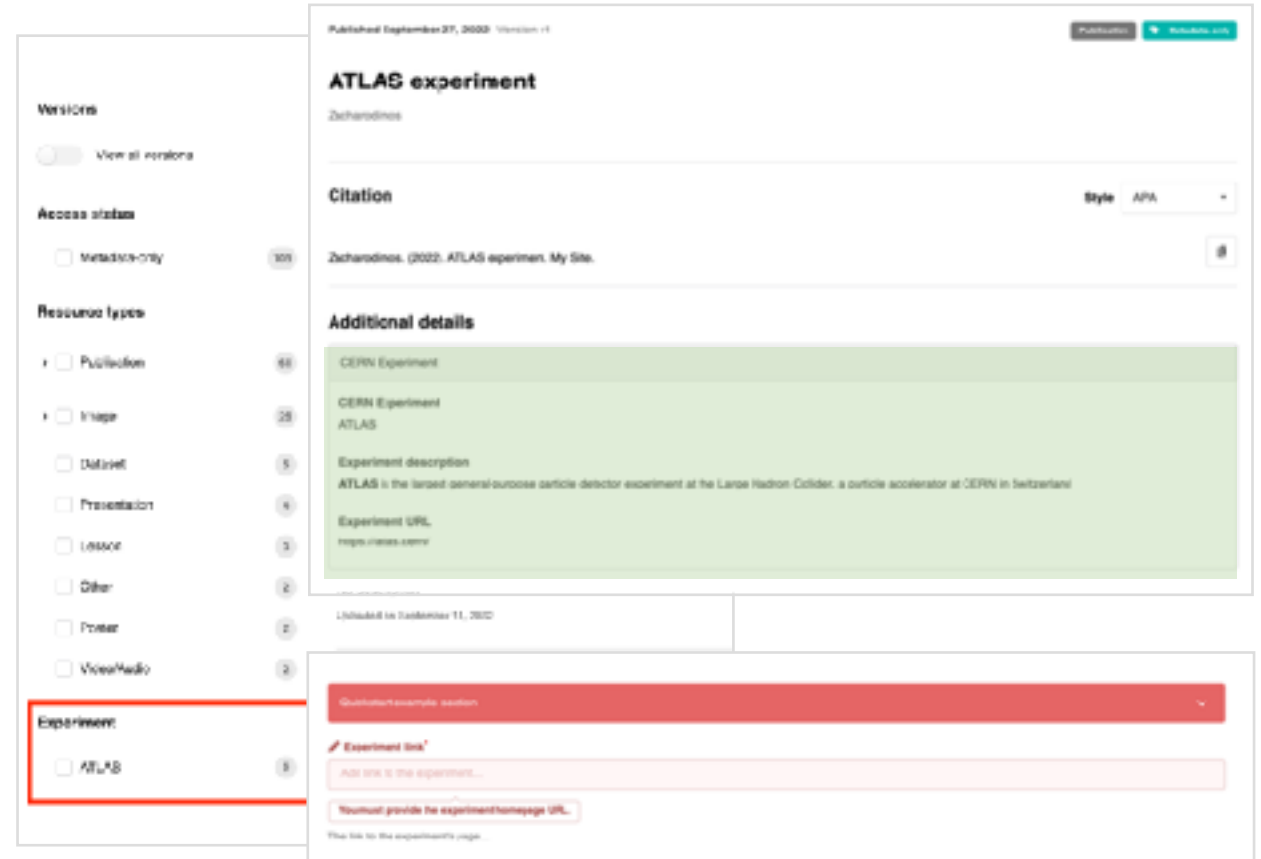


Předkonfigurované repozitáře pro NRP

- Model pro popis výzkumných dat
 - založený na DataCite, Invenio RDM
 - model Národního metadatového adresáře z projektu EOSC-CZ a CESNET datového repozitáře
- Model pro popis dokumentů (článků, ...)
 - dokumentová část národního repozitáře (bývalý NUŠL)

Definice prvků v konfiguraci

- K předkonfigurovanému repozitáři je možno přidat další metadatová pole v konfiguraci
- Pokud jsou jednoduchého typu, není potřeba žádné programování
 - číslo, text, slovníková položka, odkaz (na web či na jiný model), pole hodnot, hierarchie hloubky 1
- Pro složitější metadatová pole je nutno napsat zásuvný modul (jak se prezentují uživateli, jak je uživatel edituje, jak budou validována)



Published September 27, 2022 | Version 1.1

ATLAS experiment

Zachardinos

Citation Style: APA

Zachardinos. (2022). ATLAS experiment. My Site.

Additional details

CERN Experiment

CERN Experiment
ATLAS

Experiment description
ATLAS is the largest general-purpose particle detector experiment at the Large Hadron Collider, a particle accelerator at CERN in Switzerland.

Experiment URL
<https://atlas.cern/>

Unlinked to September 11, 2022

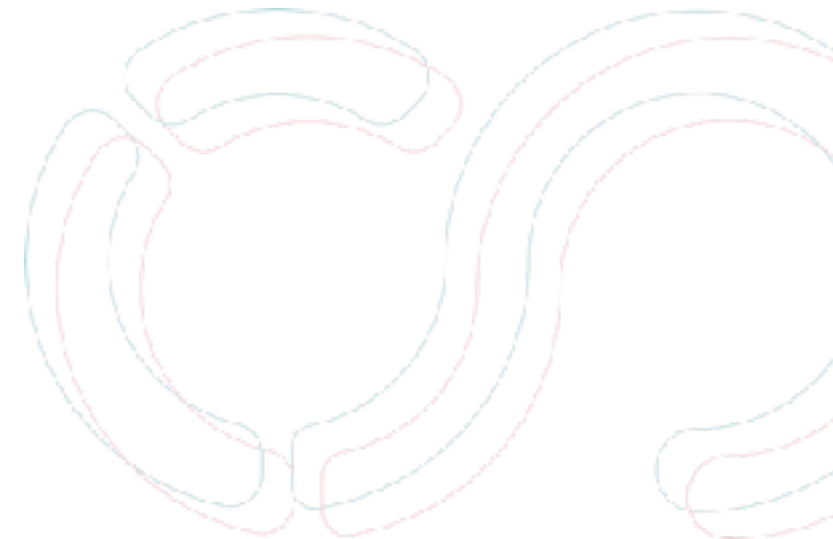
Experiment link*

Add link to the experiment...

* You must provide the experiment homepage URL.
This link is the experiment's page.

Postup nasazení repozitáře

- Konzultace nad metadatovým schématem (modelem) a rozšířeními, která jsou požadována oborovou skupinou (ve spolupráci s NTK)
- Rozhodnutí, zda je možno rozšíření realizovat konfigurací s využitím předem připravených repozitářů v rámci NRP a konfiguračních prvků nebo je nutno změnit metadatové schéma ve vlastní nadstavbě repozitáře
- Příprava testovací instance a odladění modelu
- Uvedení do produkce
- Udržování, aktualizace, kontroly provozu, ...



Díky za pozornost



Spolufinancováno
Evropskou unií