

AMČR API

Uživatelská příručka

1. 6. 2020

! Upozornění

Následující stránka k 1. červenci 2024 neprošla revizí a informace, které obsahuje, nemusí být plně platné pro aktuálně spuštěné verze nástrojů AMČR. Zároveň mohou být některé odkazy, které stránka obsahuje nefunkční, screenshoty ze zastaralých verzí nástrojů apod.

API v současné době nabízí službu *AMCR Data Provider*, která poskytuje metadatové záznamy z databáze AMČR, a to pomocí protokolu *OpenArchives Initiative Protocol for Metadata Harvesting* ([OAI-PMH](#)).

Naše implementace protokolu OAI-PMH podporuje reprezentaci metadat ve standardech:

- [Dublin Core](#) – poskytuje část datasetu týkající se dokumentů a jejich popisu;
- [CIDOC-CRM](#) – nabízí archivovaná, volně přístupná data z AMČR ve formátu RDF v syntaxi odpovídající této ontologii;
- **AMČR XML** – nativní formát, který umožňuje plné vytěžování databáze při zohlednění přístupových práv.

[Uživatelé](#) s vyšším oprávněním mohou využít své přihlašovací údaje do AMČR pro přístup k nearchivovaným či jinak chráněným záznamům. Aktualizace dat probíhá na denní bázi.

API najdete na adrese <https://api.aiscr.cz>.

Technické řešení

Architektura API se skládá z 5 komponent: *AMCR Filter*, *RDF Converter*, *Scrub*, *AMCR Pass* a *AMCR Data Provider*. Všechny komponenty jsou napsány v jazyce Java. Řešení používá open-source projekt ~~X3ML Engine pro konverzi XML na RDF~~.



**Archaeological
Information System
of the Czech Republic**

Public API
AMCR Data Provider

Mgr. Petr Pajdla
Institute of Archaeology, Czech Acad Sci, Brno

Introduction

The Archaeological Information System of the Czech Republic (AIS CR) is a research infrastructure integrating digital resources on Czech archaeology. In our core information system (Archaeological Map of the Czech Republic, AMCR), we collect and curate data on new and past archaeological activities conducted in the Czech Republic. We strive to follow FAIR principles and introduce new digital services for the archaeological community and the general public.

Public API

The API provides metadata records from the AMCR database in a machine readable format using the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH). OAI-PMH implementation supports representation in native AMCR XML, CIDOC-CRM and Dublin Core.

Example Request

List identifiers of documents that were published between 1st and 28th February 2021 in CIDOC-CRM.

```
https://api.aiscr.cz/dapro/?oaiVerb=ListIdentifiers&
  set=dokument&
  from=2021-02-01&
  until=2021-02-28&
  metadataPrefix=oai_rdf
```

click me

```
<?xml version="1.0" encoding="UTF-8" standalone="no"?>
<url xmlns="http://www.openurl.org/oai-2.0/" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xsi:schemaLocation="http://www.openurl.org/oai-2.0/ http://www.openurl.org/oai-2.0/oai-2.0.xsd">
  <request verb="ListIdentifiers" metadataPrefix="oai_rdf" from="2021-02-01" until="2021-02-28" set="dokument" http://api.aiscr.cz/dapro/oai/request?>
    <ListIdentifiers>
      <header>
        <identifier>https://api.aiscr.cz/id/cr-00000000/<identifier>
        <setSpec>2021-02-02/<setSpec>
        <setSpec>dc-dokument/<setSpec>
      </header>
      <body>
        <identifier>https://api.aiscr.cz/id/cr-00000000/<identifier>
        <setSpec>2021-02-02/<setSpec>
        <setSpec>dc-dokument/<setSpec>
      </body>
    </ListIdentifiers>
  </request>
</url>
```

To get another page, use `resumptionToken=` as a single argument to a given verb.

Verbs

```
https://api.aiscr.cz
/dapro/oai?verb=
Identify
returns information on the repository
ListMetadataFormats
returns available metadata formats
ListIdentifiers
returns record headers
arguments: metadataPrefix
optional: set, from, until
ListRecords
returns records
arguments: metadataPrefix
optional: set, from, until
GetRecord
returns individual record
arguments: metadataPrefix,
identifier
ListSets
lists possible sets
```

Arguments

```
& metadataPrefix=
oai_dc oai_rdf oai_amar
Dublin Core CIDOC-CRM Native XML
```

```
& identifier=
filters based on identifier,
see ListIdentifiers
```

```
& set=
filters based on sets, see ListSets
```

```
& from= & until=
date in ISO-8601, i.e. 2021-04-01
```

For more information on the infrastructure, check the Structuring archaeological evidence book by Martin Kuna et al.

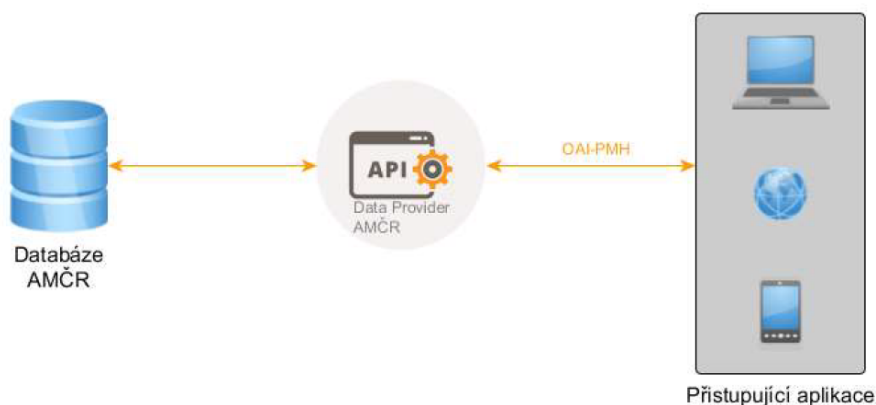




EUROPEAN UNION
European Structural and Investment Funds
Operational Programme Research, Development and Education



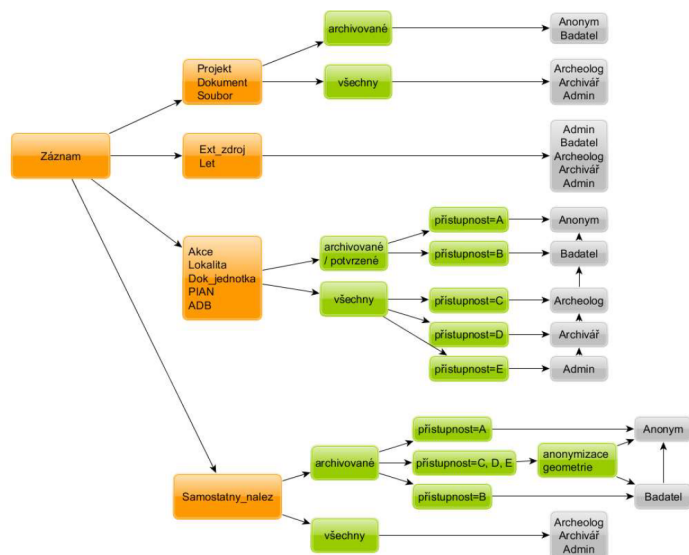
Ministry of Education, Youth and Sports



Řešení filtrace

Stav a přístupnost záznamu

Přístupnost záznam je ovlivněna jak [procesními stavy](#) daných metadatových záznamů, tak [uživatelskou rolí](#).



Obr. 5: Schéma přístupnosti záznamů dle uživatelských rolí

Sety

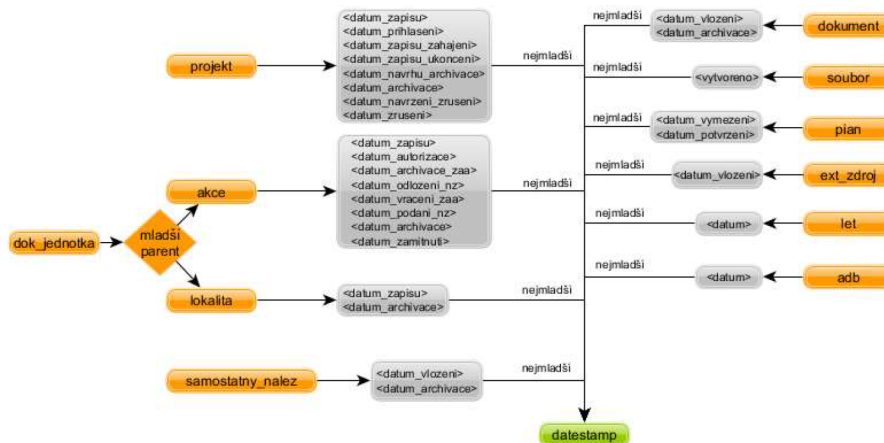
Selektivní výběr je umožněn pomocí nadefinovaných setů, viz také Obr. 4.

Tab. 1: Sety

| setSpec | setName |
|------------------|---|
| projekt | Projekty / Projects |
| akce | Akce / Fieldwork Events |
| lokalita | Lokality / Sites |
| dok_jednotka | Dokumentační jednotky / Descriptive Units |
| adb | Archeologický dokumentační bod / Archaeological Documentation Point |
| dokument | Dokumenty / Documents |
| soubor | Soubory / Files |
| pian | PIAN / Spatial Units |
| ext_zdroj | Externí zdroje / External Sources |
| let | Lety / Flights |
| samostatny_nalez | Samostatné nálezy / Individual Finds |

Datestamps

Pro dotazy `ListRecords` a `ListIdentifiers` je umožněno filtrování na základě volitelného argumentu `datestamp` (from – od, until – do), kdy `datestamp` je vždy nejmladší z dostupných datumů.



Obr. 6: Specifikace nastavení `datestamp` pro jednotlivé záznamy

AMČR nepodporuje evidenci smazaných záznamů. Po smazání záznamu z databáze proto dojde k odstranění záznamu z API bez náhrady ve formě hlavičky smazaného záznamu (`status="deleted"`). Některé operace se v AMČR nemusí projevit změnou `datestamp` v API. Důvodem je způsob logování změn, který se váže pouze na specifické operace (změna [stavu](#) záznamu). Při využití API proto doporučujeme data pravidelně obnovovat, nebo umožnit uživateli vynucenou aktualizaci konkrétního záznamu. Podpora úplného sledování změn bude implementována v budoucích verzích API po provedení nutných úprav v databázi AMČR.

Přihlášení

Pokud chcete pro stahování dat používat svůj účet AMČR, je nutné přihlášení pomocí Basic access authentication. Protokol https zajišťuje bezpečnou komunikaci. Způsob přihlášení se liší v každém nástroji, který je pro stahování dat používán. Níže jsou detailněji popsány způsoby přihlášení pro *cURL* a *Postman*.

cURL

cURL je nástroj skládající se ze softwarové knihovny (libcurl) a z nástroje příkazového řádku (curl) sloužících ke stahování souborů přes počítačovou síť. Nástroj je dostupný pro různé operační systémy Linux, macOS a Windows. Své uživatelské jméno a heslo je možné zadat pomocí přepínače `-u`.

```
curl -u username:password <GET request>
```

Přihlašovací údaje je nutné posílat v každém requestu.

Postman

Postman je multiplatformní aplikace pro vývojáře sloužící k návrhu a interakci s HTTP API. Uživatelské jméno a heslo lze zadávat na kartě *Autorizace*. V rozevírací nabídce *TYPE* zvolte možnost *Basic Auth* a stiskněte tlačítko *Odeslat*. Přihlašovací údaje jsou poté posílány automaticky v každém requestu.

Verbs

Dotazování v protokolu OAI-PMH je možné pomocí tzv. sloves (verbs). Význam jednotlivých sloves s příkladem užití jsou popsány v následujících kapitolách. Podrobná specifikace protokolu OAI-PMH je k dispozici [zde](#).

Identify

Sloveso *Identify* se používá k získání informací o úložišti.

Dotaz: <https://api.aiscr.cz/dapro/oai?verb=Identify>

ListMetadataFormats

Sloveso *ListMetadataFormats* se používá k načtení formátů metadat dostupných z úložiště.

Dotaz: <https://api.aiscr.cz/dapro/oai?verb=ListMetadataFormats>

Odpověď:

```
<OAI-PMH xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/ http://www.openarchives.org/OAI/2.0/">
  <responseDate>2023-07-10T14:10:00Z</responseDate>
  <request verb="ListMetadataFormats">https://api.aiscr.cz/dapro/oai</request>
  <ListMetadataFormats>
    <metadataFormat>
      <metadataPrefix>oai_dc</metadataPrefix>
      <schema>http://www.openarchives.org/OAI/2.0/oai_dc.xsd</schema>
      <metadataNamespace>http://www.openarchives.org/OAI/2.0/oai_dc/</metadataNamespace>
    </metadataFormat>
    <metadataFormat>
      <metadataPrefix>oai_rdf</metadataPrefix>
      <schema>https://api.aiscr.cz/dapro/media/oai_rdf.xsd</schema>
      <metadataNamespace>https://api.aiscr.cz/schema/oai_rdf/</metadataNamespace>
    </metadataFormat>
  </ListMetadataFormats>
</OAI-PMH>
```

```

    <metadataFormat>
      <metadataPrefix>oai_amcr</metadataPrefix>
      <schema>https://api.aiscr.cz/dapro/media/oai_amcr.xsd</schema>
      <metadataNamespace>https://api.aiscr.cz/schema/oai_amcr/</metadataNamespace>
    </metadataFormat>
  </ListMetadataFormats>
</OAI-PMH>

```

Implementace OAI-PMH podporuje reprezentaci metadat v:

- Dublin Core – metadataPrefix oai_dc,
- CIDOC CRM – metadataPrefix oai_rdf,
- interním AMČR XML – metadataPrefix oai_amcr.

Jednotlivá schémata jsou k dispozici na: https://api.aiscr.cz/dapro/media/*.xsd.

ListIdentifiers

Sloveso `ListIdentifiers` se používá pro načtení záhlaví záznamů (header) z úložiště. Povinným argumentem je `metadataPrefix`. Nepovinné argumenty umožňují filtraci záhlaví na základě nastavených [setů](#) a/nebo [datestamp](#). Uživatel dostane v odpovědi první stránku se záhlavím záznamů. Pro načtení dalších stránek je nutné použít [resumptionToken](#).

Dotaz: https://api.aiscr.cz/dapro/oai?verb=ListIdentifiers&metadataPrefix=oai_amcr

ListRecords

Sloveso `ListRecords` se používá pro načtení záznamů z úložiště. Povinným argumentem je `metadataPrefix`. Nepovinné argumenty umožňují filtraci záznamů na základě nastavených [setů](#) a/nebo [datestamp](#). Uživatel dostane v odpovědi první stránku se záznamy. Pro načtení dalších stránek je nutné použít [resumptionToken](#).

Dotaz: https://api.aiscr.cz/dapro/oai?verb=ListRecords&metadataPrefix=oai_amcr

Odpověď:

```

<OAI-PMH xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/ http://www.openarchives.org/OAI/2.0/xsd" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance">
  <responseDate>2023-07-11T07:21:17Z</responseDate>
  <request verb="ListRecords" metadataPrefix="oai_amcr">https://api.aiscr.cz/dapro/oai</request>
  <ListRecords>
    <record>
      <header>
        <identifier>https://api.aiscr.cz/id/C-201773056</identifier>
        <datestamp>2022-06-09</datestamp>
        <setSpec>projekt</setSpec>
      </header>
    </record>
  </ListRecords>
</OAI-PMH>

```

```

</header>
<metadata>
  <oai_amcr:amcr xsi:schemaLocation="https://api.aiscr.cz/schema/oai_amcr https://ap
    <projekt>
      <ident_cely>C-201773056</ident_cely>
      <stav>6</stav>
      <typ_projektu>záchranný</typ_projektu>
      <datetime_born>2016-01-01 00:00:00</datetime_born>
      <okres>PLZEŇ-SEVER</okres>
      (...)
      <termin_odevzdani_nz>2019-12-30</termin_odevzdani_nz>
      <child_akce>C-201773056A</child_akce>
    </projekt>
  </oai_amcr:amcr>
</metadata>
</record>
<resumptionToken cursor="0" expirationDate="2023-07-12T22:30:00Z">2023-07-10.oai_amcr.
</ListRecords>
</OAI-PMH>

```

GetRecord

Sloveso **GetRecord** se používá k získání individuálního záznamu metadat z úložiště. Požadované argumenty jsou **metadataPrefix** a identifikátor položky (**identifier**), ze které je záznam požadován. Položky jednotlivých setů nesou v hlavičce persistentní identifikátory, odvozované z pole *ident_cely* u jednotlivých záznamů nejvýše v XML struktuře záznamu v daném setu. Výjimkou je set soubor, kde jako identifikátor slouží pole *filepath*. K základnímu formátu identifikátoru je vždy jako prefix připojena URL: <https://api.aiscr.cz/id/>. Odkazy mezi záznamy napříč sety jsou řešeny na úrovni základního identifikátoru bez prefixu. Další informace ke struktuře identifikátorů a vazbám naleznete v samostatné [kapitole](#).

Dotaz: https://api.aiscr.cz/dapro/oai?verb=GetRecord&identifier=https://api.aiscr.cz/id/C-DL-200400001&metadataPrefix=oai_amcr

Odpověď:

```

<OAI-PMH xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/ http://www.openarchives.
  <responseDate>2023-07-11T07:37:18Z</responseDate>
  <request verb="GetRecord" metadataPrefix="oai_amcr" identifier="https://api.aiscr.cz/id/
  <GetRecord>
    <record>
      <header>
        <identifier>https://api.aiscr.cz/id/C-DL-200400001</identifier>
        <datestamp>2018-06-15</datestamp>
        <setSpec>dokument</setSpec>

```



```

</header>
<metadata>
  <oai_amcr:amcr xsi:schemaLocation="https://api.aiscr.cz/schema/oai_amcr https://ap
    <dokument>
      <ident_cely>C-DL-200400001</ident_cely>
      <let>C-LET-00253</let>
      (...)
      <jednotka_dokumentu>
        <ident_cely>C-DL-200400001-D01</ident_cely>
        (...)
        <komponenta_dokumentu>
          <ident_cely>C-DL-200400001-K01</ident_cely>
          <parent>C-DL-200400001-D01</parent>
          <obdobi>pr.zem</obdobi>
          <areal>sídliště nesp.</areal>
          <aktivita_sidlistni>1</aktivita_sidlistni>
          <aktivita_pohrebni>0</aktivita_pohrebni>
          (...)
          <nalez_dokumentu>
            (...)
          </nalez_dokumentu>
        </komponenta_dokumentu>
      </jednotka_dokumentu>
    </dokument>
  </oai_amcr:amcr>
</metadata>
</record>
</GetRecord>
</OAI-PMH>

```

ListSets

Sloveso **ListSets** se používá k načtení nastavené struktury úložiště, tzv. setů.

Dotaz: <https://api.aiscr.cz/dapro/oai?verb=ListSets>

Parametry

Stránkování

Pomocí sloves uživatel získává první stránku požadovaných metadat. Pro získání dalších záznamů je nutné použít parametr **resumptionToken**, který uživatel dostal v odpovědi na svůj dotaz. Parametr **resumptionToken** je vždy na konci odpovědi. Počet záznamů na stránku je dán konfigurací na straně poskytovatele API.

Odpověď na dotaz v sekci [ListRecords](#) vrátila následující XML element obsahující `resumptionToken`, který je použit v dalším dotazu:

```
<resumptionToken cursor="0" expirationDate="2023-07-12T22:30:00Z">
  2023-07-10.oai_amcr.1000.projekt.1000.1000...f
</resumptionToken>
```

Dotaz: https://api.aiscr.cz/dapro/oai?verb=ListIdentifiers&resumptionToken=2023-07-10.oai_amcr.1000.projekt.1000.1000...f

Parametr `resumptionToken` je v tomto případě společně se slovesy `ListRecords` a `ListIdentifiers` výhradním parametrem.

Filtrování

Tzv. selektivní sklizeň umožňuje uživatelům omezit požadavky na sběr dat na části metadata dostupných z úložiště. OAI-PMH podporuje selektivní sklizeň se dvěma typy kritérií pro sklizeň, která mohou být kombinována: `datestamp` a `set`. Data provider AMČR navíc uživatelům filtruje záznamy podle jejich přístupnosti.

Podle přístupnosti záznamu

Při zpracování dotazu jsou uživatelé automaticky záznamy filtrovány podle přístupnosti, která je nadefinovaná v AMČR. Anonymní uživatel má vždy přístupnost pouze k archivovaným záznamům. Archeologové či uživatelé s vyšším oprávněním mají přístup i k nearchivovaným či jinak chráněným záznamům. Více viz kapitoly k [procesním stavům](#) a [uživatelským rolím](#).

Pomocí Setů

Protokol OAI-PMH umožňuje filtrování pomocí nadefinovaných setů, a to přidáním parametru `set`. Pomocí slovesa `ListSets` je možné získat hodnotu `setSpec` pro všechny nadefinované [sety](#), viz také Obr. 4. Pro zvolené filtrování se poté do requestu se slovesy `ListRecords` nebo `ListIdentifiers` přidává parametr `set=setSpec`. Filtrování pomocí setů lze kombinovat s filtrováním pomocí `datestamp`.

Dotaz: https://api.aiscr.cz/dapro/oai?verb=ListRecords&set=akce&metadataPrefix=oai_amcr

Pomocí datestamp

Protokol OAI-PMH umožňuje filtrování pomocí datumu (**datestamp**), přiřazeného ke každému záznamu. Pro požadované filtrování se poté do requestu se slovesy **ListRecords** nebo **ListIdentifiers** přidává parametr **from** (od) a/nebo **until** (do).

Dotaz: https://api.aiscr.cz/dapro/oai?verb=ListIdentifiers&metadataPrefix=oai_amcr&from=2022-12-10&until=2022-12-20

Pajdla, Petr. 2021. „AIS CR Public API: AMCR Data Provider”. Poster prezentováno v LiSeH 2021, GAMS. <https://hdl.handle.net/11471/562.40.3>.