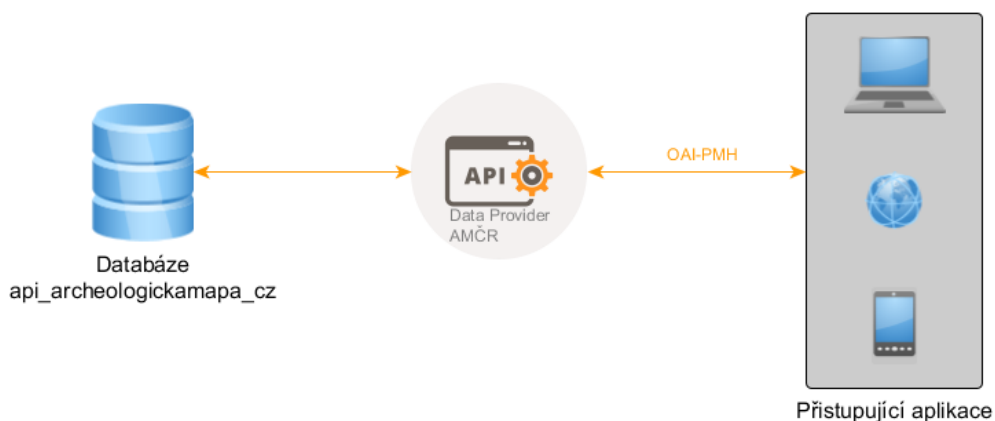


API AIS ČR – Technická specifikace

1. Technické řešení

Aplikace Data Provider AMČR poskytuje metadata o datových objektech uložených v Databázi AMČR, a to pomocí protokolu Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH).



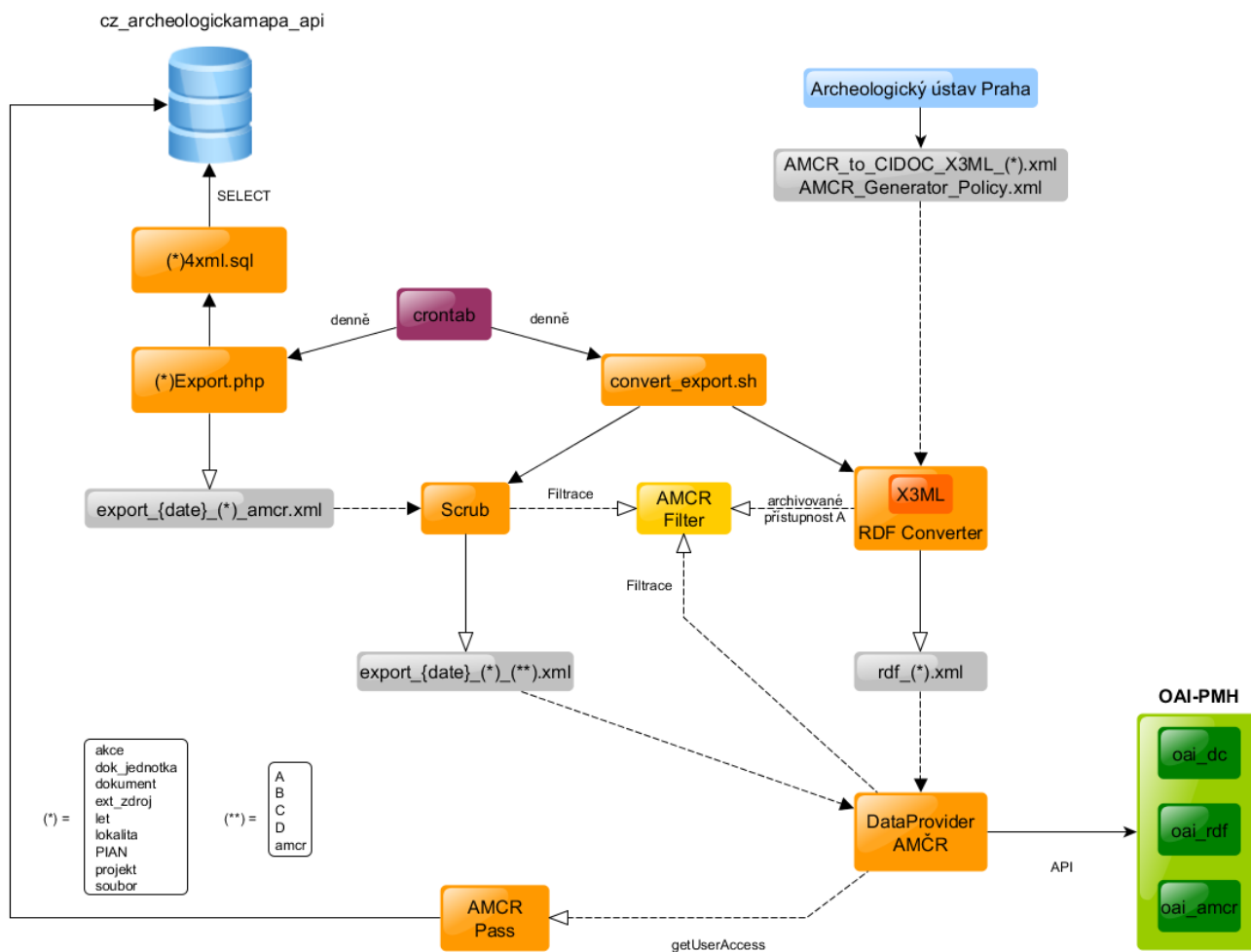
Implementace OAI-PMH podporuje reprezentaci metadat v Dublin Core, CIDOC-CRM a v interním formátu AMČR XML.



Uživatel může využít anonymní přístup nebo svůj registrovaný účet Digitálního archivu AMČR, resp. Archeologické mapy ČR (<http://www.archeologickamapa.cz/>).

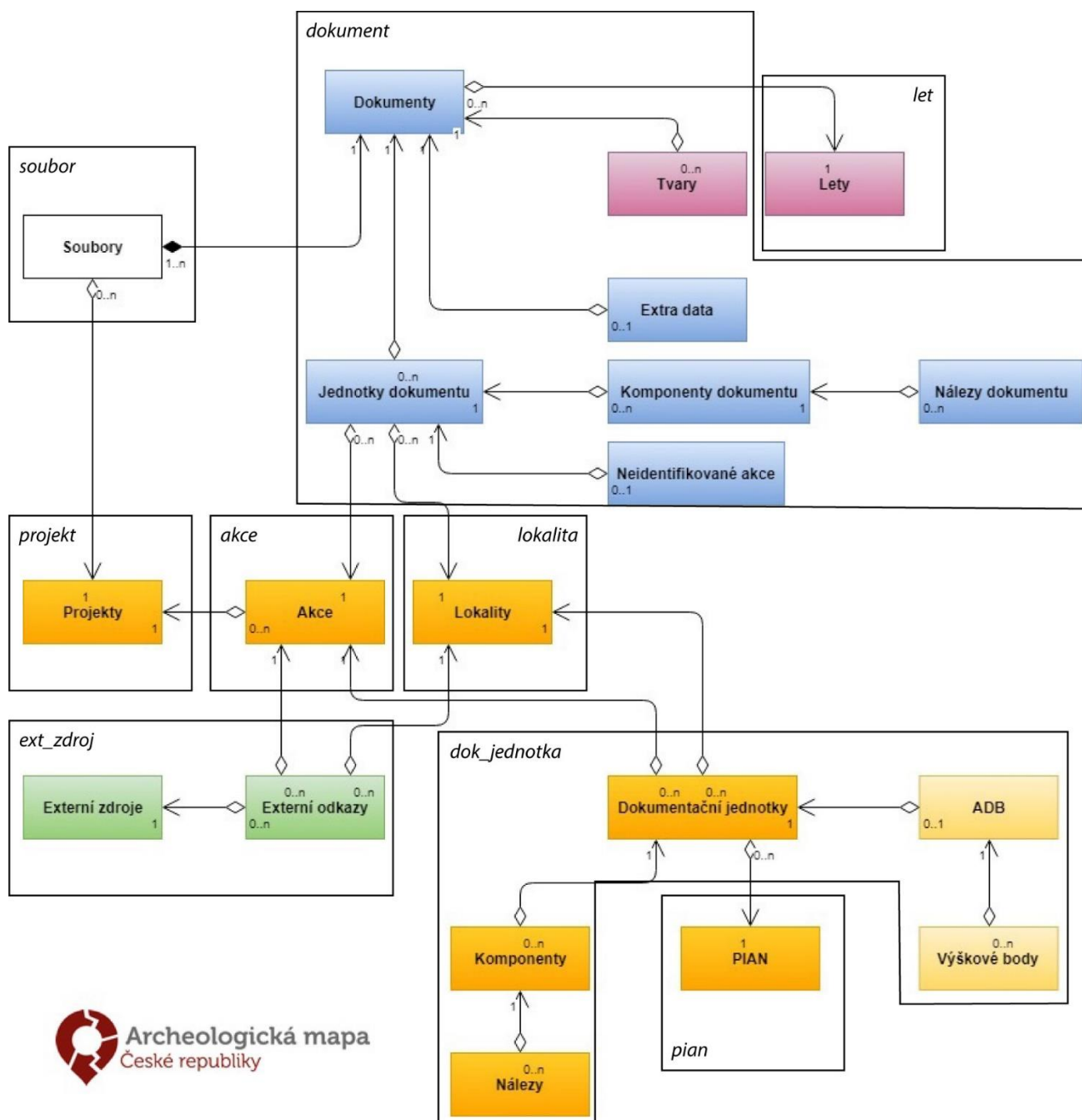
Architektura API se skládá z 5 komponent: AMCR Filter, RDF Converter, AMCR Pass a Data Provider. Všechny komponenty jsou napsány v jazyce Java. Řešení používá open-source projekt X3ML Engine pro konverzi XML na RDF.

1.1 Schéma řešení



Copyright @ 2019 Space Systems Czech, s.r.o.

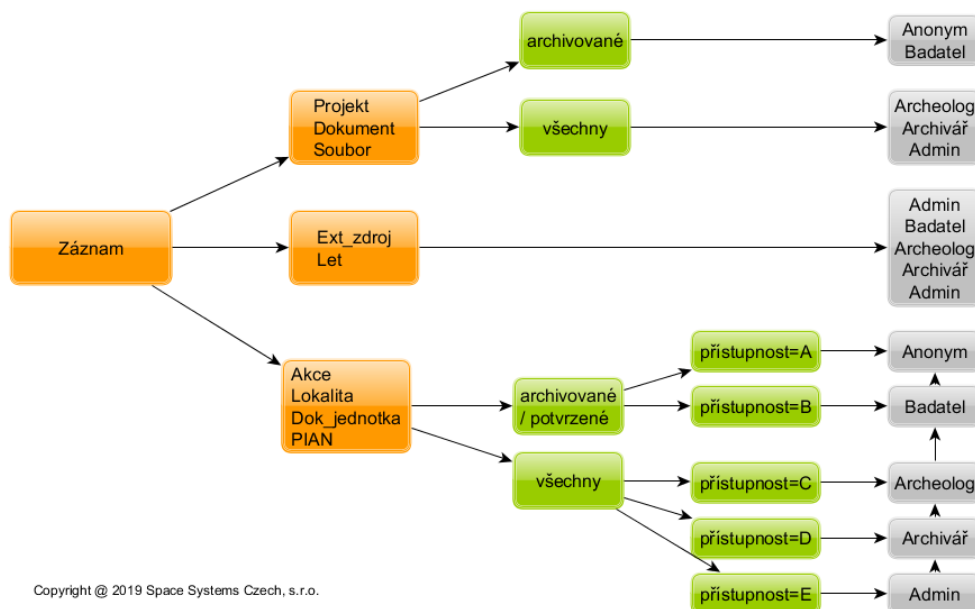
1.2 Schéma databáze AMČR s označením dílčích datových setů



1.3 Řešení filtrace

1.3.1 Stav a přístupnost záznamu

Data jsou filtrována pomocí komponenty AMCR Filter, která uživateli předává data na základě stavu a přístupnosti záznamu.



1.3.2 Sety

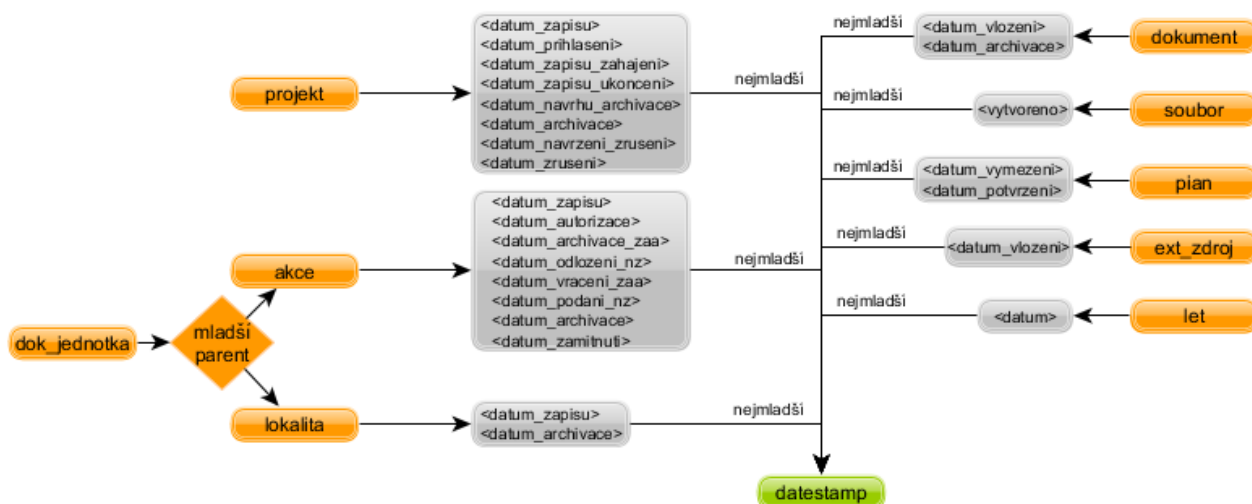
Selektivní výběr je umožněn pomocí nadefinovaných setů.

setSpec	setName
projekt	Projekty / Projects
akce	Akce / Fieldwork Events
lokalita	Lokality / Sites
dok_jednotka	Dokumentační jednotky / Descriptive Units
dokument	Dokumenty / Documents
soubor	Soubory / Files
pian	PIAN / Spatial Units
ext_zdroj	Externí zdroje / External Sources
let	Lety / Flights

1.3.3 Datestamps

Pro dotazy ListRecords a ListIdentifiers je umožněno filtrování na základě volitelného argumentu `datestamp` (`from – od`, `until – do`).

Nastavení `datestamp` pro jednotlivé záznamy je dáno následující specifikací, kdy `datestamp` je vždy nejmladší z dostupných datumů:



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.

API AIS ČR – Uživatelská příručka

1. Obecné informace

API v současné době nabízí službu Data Provider AMČR, která poskytuje metadatové záznamy z databáze Archeologická mapa ČR, a to pomocí protokolu OpenArchives Initiative Protocol for Metadata Harvesting (OAI-PMH). Implementace OAI-PMH podporuje reprezentaci metadat v:

- 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 (<http://www.cidoc-crm.org/>)
- AMČR XML – nativní formát, který umožňuje plné vytěžování databáze při zohlednění přístupových práv

Archeologové či uživatelé s vyšším oprávněním mohou využít registrovaný účet Archeologické mapy ČR pro přístup k nearchivovaným či jinak chráněným záznamům. Aktualizace dat probíhá na denní bázi.

URL API: <https://api.aiscr.cz/dapro>

2. Přihlášení

Pokud chcete pro stahování dat používat svůj účet Digitálního archivu 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.

a) 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:heslo GET_request
```

Přihlašovací údaje je nutné posílat v každém requestu.

b) 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.

3. 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 na těchto stránkách: <http://www.openarchives.org/OAI/openarchivesprotocol.html>.

3.1 Identify

Sloveso „Identify“ se používá k získání informací o úložišti.

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

3.2 ListMetadataFormats

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

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

Odpověď:

```
<OAI-PMH
<responseDate>2019-06-07T08:37:39Z</responseDate>
<request verb="Identify">https://api.aiscr.cz/dapro/oai</request>
<Identify>
<repositoryName>AIS CR - Archaeological Map of the Czech Republic</repositoryName>
<baseURL>https://api.aiscr.cz/dapro/oai</baseURL>
<protocolVersion>2.0</protocolVersion>
<adminEmail>info@amapa.cz</adminEmail>
<earliestDatestamp>1990-01-01</earliestDatestamp>
<deletedRecord>no</deletedRecord>
<granularity>YYYY-MM-DD</granularity>
<description>
<rightsManifest xmlns="http://www.openarchives.org/OAI/2.0/rights/" xsi:schemaLocation="http://www.
openarchives.org/OAI/2.0/rights/
http://www.openarchives.org/OAI/2.0/rightsManifest.xsd" appliesTo="http://www.openarchives.org/OAI/
2.0/entity#metadata">
<rights>
<rightsReference ref="http://creativecommons.org/licenses/by-nc/4.0/rdf"/>
</rights>
</rightsManifest>
</description>
<description>
<oai_dc:dc xmlns:oai_dc="http://www.openarchives.org/OAI/2.0/oai_dc/" xmlns:dc="http://purl.org/dc/
elements/1.1/" xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/oai_dc/
http://www.openarchives.org/OAI/2.0/oai_dc.xsd">
<dc:title>Archeologická mapa České republiky (AMČR)</dc:title>
<dc:creator>Archeologický ústav AV ČR, Praha, v.v.i.</dc:creator>
<dc:creator>Archeologický ústav AV ČR, Brno, v.v.i.</dc:creator>
<dc:identifier>https://api.aiscr.cz/</dc:identifier>
<dc:source>http://www.archeologickamapa.cz/</dc:source>
<dc:relation>http://www.aiscr.cz/</dc:relation>
<dc:language>cs</dc:language>
<dc:language>en</dc:language>
<dc:description>
Archeologická mapa České republiky (AMČR) je informační systém pro sběr, správu a prezentaci dat o
archeologických výzkumech na území ČR a pro poznání minulosti Čech, Moravy a Slezska. Obsažená data
popisují desetitisíce archeologických výzkumů a jejich konkrétních poznatků. Zahrnují informace o
vedoucích výzkumu, o tom, kdy výzkum probíhal, o jeho lokalizaci i to, jaké nálezy a z jakého
období výzkum zjistil. Zvláštním typem záznamů jsou údaje o archeologických lokalitách, zjištěných
povrchovým i dálkovým průzkumem. Většina záznamů je napojena na úložiště digitálních dokumentů a
bibliografický katalog. AMČR je určena profesionálním archeologům, odborníkům z blízkých oborů,
studentům i zájemcům z širší veřejnosti, jimž nabízí možnost efektivně pracovat s poznatky
získanými generacemi archeologů v naší zemi. Je provozována Archeologickými ústavu AV ČR v Praze a
Brně. AMČR je od počátku zamýšlena jako páteří informační systém české archeologie, a proto již v
současné době vznikají její účelová rozšíření pro archeologii pražské památkové rezervace, ale i
další moduly. AMČR je tak nejdůležitější součástí širší infrastruktury Archeologický informační
České republiky (AIS CR), která je zapsána na Cestovní mapě velkých infrastruktur pro výzkum,
experimentální vývoj a inováce, schválené vládou ČR.
</dc:description>
<dc:description>
Pro jednoznačné označení záznamů jsou užívány identifikátory obsažené v polích "ident_cely"
společně se základní doménou "https://api.aiscr.cz/id/". Identifikátor je tak v argumentech vždy
předáván v podobě "https://api.aiscr.cz/id/[ident_cely]" (např. "https://api.aiscr.cz/id/C-
201912345").
</dc:description>
<dc:description>http://www.archeologickamapa.cz/help</dc:description>
<dc:description>
```

```
http://www.archeologickamapa.cz/downloads/AMCR_diagram_cs.jpg
</dc:description>
<dc:description>
http://www.archeologickamapa.cz/downloads/AMCR_diagram_en.jpg
</dc:description>
</oai_dc:dc>
</description>
</Identify>
</OAI-PMH>
```

Implementace OAI-PMH podporuje reprezentaci metadat v:

Dublin Core – metadataPrefix **oai_dc**,

CIDOC-CRM – metadataPrefix **oai_rdf**,

interní AMČR XML – metadataPrefix **oai_amcr**.

Jednotlivá schémata jsou k dispozici na: <https://api.aiscr.cz/schema/>

3.3 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 (kapitola 5.2) a/nebo datestamp (kapitola 5.3). Uživatel dostane v odpovědi první stránku se záhlavím záznamů. Pro načtení dalších stránek je nutné použít resumptionToken (kapitola 4).

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

Odpověď:

```
<OAI-PMH>
<responseDate>2019-05-23T08:19:32Z</responseDate>
<request verb="ListIdentifiers" metadataPrefix="oai_amcr">https://api.aiscr.cz/dapro/oai</request>
<ListIdentifiers>
<header xmlns:ex="http://exslt.org/dates-and-times">
<identifier>https://api-test.aiscr.cz/id/C-201800019</identifier>
<datestamp>2018-12-17</datestamp>
</header>
<resumptionToken cursor="0"> 2019-06-03.oai_amcr.1000.projekt.1000.1000...f</resumptionToken>
</ListIdentifiers>
</OAI-PMH>
```

3.4 ListRecords

Sloveso „ListRecords“ se používá pro načtení záznamů z úložiště. Povinným argumentem je metadataPrefix. Nepovinné argumenty umožňují filtraci záznamů na základě nastavených Setů (kapitola 5.2) a/nebo datestamp (kapitola 5.3). Uživatel dostane v odpovědi první stránku se záznamy. Pro načtení dalších stránek je nutné použít resumptionToken (kapitola 4).

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

Odpověď:

```
<OAI-PMH>
<responseDate>2019-05-23T08:59:43Z</responseDate>
<request verb="ListRecords" metadataPrefix="oai_amcr">https://api.aiscr.cz/dapro/oai</request>
<ListRecords>
<record>
<header xmlns:ex="http://exslt.org/dates-and-times">
```



```
<identifier>https://api-test.aiscr.cz/id/C-201800019</identifier>
<timestamp>2018-12-17</timestamp>
</header>
<metadata>...</metadata>
</record>
<resumptionToken cursor="0"> 2019-06-03.oai_amcr.1000.projekt.1000.1000...f</resumptionToken>
</ListRecords>
</OAI-PMH>
```

3.5 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 (viz kap. 3.6) nesou v hlavičce persistentní identifikátory, odvozené 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 zde: <http://www.archeologickamapa.cz/help/>

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

Odpověď:

```
<OAI-PMH>
<responseDate>2019-05-23T09:11:27Z</responseDate>
<request verb="GetRecord" metadataPrefix="oai_amcr" identifier="https://api-
test.aiscr.cz/id/C-DL-200400001">https://api.aiscr.cz/dapro/oai</request>
<GetRecord>
<record>
<header xmlns:ex="http://exslt.org/dates-and-times">
<identifier>https://api-test.aiscr.cz/id/C-DL-200400001</identifier>
<timestamp>2018-06-15</timestamp>
</header>
<metadata>...</metadata>
</record>
</GetRecord>
</OAI-PMH>
```

3.6 ListSets

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

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

Odpověď:

```
<OAI-PMH>
<responseDate>2019-05-23T09:15:49Z</responseDate>
<request verb="ListSets">https://api.aiscr.cz/dapro/oai</request>
<ListSets>
  <set>
    <setSpec>projekt</setSpec>
    <setName>Projekty / Projects</setName>
    <setDescription>...</setDescription>
  </set>
  <set>
    <setSpec>akce</setSpec>
    <setName>Akce / Fieldwork Events</setName>
    <setDescription>...</setDescription>
  </set>
  <set>
    <setSpec>lokalita</setSpec>
    <setName>Lokality / Sites</setName>
    <setDescription>...</setDescription>
  </set>
  <set>
    <setSpec>let</setSpec>
    <setName>Lety / Flights</setName>
    <setDescription>...</setDescription>
  </set>
  <set>
    <setSpec>dok_jednotka</setSpec>
    <setName>Dokumentární jednotky / Descriptive Units</setName>
    <setDescription>...</setDescription>
  </set>
  <set>
    <setSpec>dokument</setSpec>
    <setName>Dokumenty / Documents</setName>
    <setDescription>...</setDescription>
  </set>
  <set>
    <setSpec>ext_zdroj</setSpec>
    <setName>Externí zdroje / External Sources</setName>
    <setDescription>...</setDescription>
  </set>
  <set>
    <setSpec>pian</setSpec>
    <setName>PIAN / Spatial Units</setName>
    <setDescription>...</setDescription>
  </set>
  <set>
    <setSpec>soubor</setSpec>
    <setName>Soubory / Files</setName>
    <setDescription>...</setDescription>
  </set>
</ListSets>
</OAI-PMH>
```

4. 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.

Ukázka dotazu pro získání další stránky k příkladu v kapitole 3.3:

https://api-test.aiscr.cz/dapro/oai?verb=ListIdentifiers&resumptionToken=2019-06-03.oai_amcr.1000.projekt.1000.100...f

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

5. Filtrování

Tzv. selektivní sklizeň umožňuje uživatelům omezit požadavky na sběr dat na části metadat 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.

5.1 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.

5.2 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 (kapitola 3.6). 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-test.aiscr.cz/dapro/oai?verb=ListRecords&set=akce&metadataPrefix=oai_amcr

5.3 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-test.aiscr.cz/dapro/oai?verb=ListIdentifiers&metadataPrefix=oai_amcr&from=2018-12-10&until=2018-12-20

API AIS ČR – Konfigurace

1. Požadavky na systém

- Java 1.7
- Tomcat 1.7
- cURL (nutný pro vytváření exportů)
- Maven 3.5.4
- 20 GB místo na disku, paměť

2. Změna nastavení

a) nastavení Tomcat:

- `etc/tomcat7/catalina.properties`
`CATALINA_TMPDIR=/var/cache/exports/tmp`
`cz.ssc.pass.confir=/etc/tomcat7`
`cz.ssc.confir=/etc/tomcat7`
- `/usr/share/tomcat7/bin/catalina.sh`
`export CATALINA_TMPDIR=/var/cache/exports/tmp`
- `/etc/default/tomcat7:`
`JAVA_OPTS="-Djava.awt.headless=true -Xms5000m -`
`Xmx5000m -XX:MaxPermSize=256m -`
`XX:+UseConcMarkSweepGC"`

b) nastavení amcr_dapro: `amcr_pass.json`

nastavení přístupu do databáze

c) nastavení amcr_pass: `amcr_dapro.json`

„`filebase`“: cesta ke xml a rdf exportům

„`datesource`“: `true` nebo `false`, specifikuje název souborů s exporty,
zda obsahuje datum nebo ne

„`pagerows`“: počet záznamů na stránku

„`identfierhead`“: <https://api.aiscr.cz/id/> validní absolutní URL

3. Vytváření XML a RDF exportů

XML exporty jsou spouštěny pomocí démonu, a to každý den v 00:30 SEČ. Po ukončení exportu je volán skript `convert_export.sh`, který zajistí konverzi xml do rdf a vytvoření souborů pro jednotlivé přístupnosti (A-D). Pokud všechny kroky proběhnou v pořádku, nastaví se touch na nové soubory.

Editace crontab: `crontab -e`

Přidat řádek:

```
# XML export post-processing
```

```
30 00 * * * find /var/cache/exports/data -xtype f -mmin +2160 -exec rm -
rf {} \; && cd /var/cache/exports/data; curl -s -O -J
"http://localhost/xml_export/0/?do={projekt,akce,externi_zdroj,lokalita,l
et,dokument,dokumentacni_jednotka,pian,soubor}&key=write_private_key";
chown tomcat7:www-data export_*; sudo -u tomcat7
/var/cache/exports/lib/convert_export.sh
```

Kontrola crontab: `crontab -l`

4. Vytváření buildu

Repozitář: <https://github.com/ARUP-CAS/aiscr-dapro>
<https://github.com/ARUP-CAS/aiscr-pass>

Struktura:

```
/var/cache/exports/lib:
    convert_export.sh
    rdfcon-1.0.SNAPSHOT.jar
    srub-1.0.SNAPSHOT.jar
    AMCR_Generator_Policy.xml
    AMCR_to_CDOC_X3ML_*.xml
```

```
/var/cache/exports/data:
/var/cache/exports/var:
```

```
/var/lib/tomcat7/webapps:
    dapro.war
    pass.war
    /dapro
    /pass
```

Homepage: