

# Environmentální analytická platforma

---

*Univerzální importní služba - technicko-metodická dokumentace*

---

18.2.2016

*Dokument obsahuje dokumentaci univerzální importní služby Environmentální analytické platformy. Dokumentace obsahuje informace potřebné pro nastavení importního modulu, metodické pokyny pro tvůrce importních konektorů a požadavky na návrháře webových služeb datových zdrojů předávajících data do EAP.*

## Obsah

Historie změn dokumentace .....	4
Koncepce .....	5
Strukturace ukládání dat .....	6
Datový zdroj a index .....	6
Datový objekt a typ .....	6
Datový záznam a dokument .....	6
Konfigurační soubor .....	7
Datum poslední úpravy .....	7
Umístění souborů .....	7
Logovací soubory .....	7
Umístění dat pro import .....	7
Umístění uzlu elasticsearch .....	7
Indexy elasticsearch .....	7
Datové soubory .....	9
Spouštění importní služby .....	11
Doporučení .....	11
Logy .....	12
Konektory .....	13
Doporučení .....	13
Datové zdroje .....	14
Typy datových zdrojů .....	14
Databáze .....	14
Jednoduché soubory .....	15
Webové služby .....	15
Jiné zdroje .....	16
Mapování dat .....	17
Základní datové typy .....	17
Textový datový typ .....	17
Numerické datové typy .....	17
Datový typ datum a čas .....	17
Logický datový typ .....	17

Binární datový typ .....	17
Komplexní datové typy.....	17
Datový typ pole .....	17
Objektový datový typ .....	17
Vnořené datové typy .....	17
Geografické datové typy .....	17
Datový typ geografický bod.....	17
Datový typ geografický tvar.....	17
Speciální datové typy .....	18
Datový typ pro IP adresu v4 .....	18
Kompletační datový typ.....	18
Datový typ počet výskytů .....	18
Hašovací datový typ .....	18
Datový typ příloha .....	18
Vícenásobná pole .....	18
Využití mapování .....	18
Datový slovník .....	19
Požadavky předávání dat na nové a modernizované datové zdroje.....	20
Identifikace environmentálních dat .....	20
Předávání environmentálních dat .....	20
Exportní služba datového zdroje EAP.....	20
Metoda „Export změn“ .....	20
Metoda „Export životních cyklů“ .....	21
Schéma komunikace.....	22
Katalog požadavků .....	23

## Historie změn dokumentace

- 10. 12. 2015 Iniciální dokumentace k verzi 1.0
- 18. 12. 2015 Doplněno doporučení pro nové systémy
- 18. 2. 2016 Restrukturalizace dokumentu. Zpřesněny požadavky na nové a modernizované systémy tak, aby se poslední kapitola dala přímo použít v zadávací dokumentaci

## Koncepce

EAP je koncipována tak, aby poskytovala nad společnými daty společné služby a aby bylo možno doplňovat nové společné služby stavebnicovým způsobem.

Data jsou z datových zdrojů získávána automaticky nebo poloautomaticky pomocí univerzální importní služby, která je pevnou součástí EAP. Ke každému datovému zdroji musí být vytvořen jednoduchý konektor, který převede data dodaná datovým zdrojem do podoby čitelné importní službou.

Univerzální importní služba je aplikace vytvořená v jazyce Java a je spouštěná pomocí virtuálního stroje Java. Importní služba po spuštění načte konfiguraci z konfiguračního souboru a na základě konfiguračních parametrů provede tyto úkony

1. Vyhledá zdrojová data umístěná v souborovém systému na místech specifikovaných v konfiguraci a
2. Připojí cílový uzel platformy elasticsearch
3. Importuje připravená data do cílového uzlu
4. Vytvoří importní log

## Strukturace ukládání dat

Data jsou ukládána do EAP podle velice jednoduchého klíče.

### Datový zdroj a index

Pro každý **datový zdroj**, jehož data mají být analyzována pomocí EAP, je vytvořen jeden nebo více **indexů**. Pro jednoduché datové zdroje obvykle postačí jeden index, zatímco pro komplexní zdroje (jako je například ISPOP) může být výhodné vytvořit více indexů.

**Index** je něco podobného jako **databáze** v relačních databázích.

### Datový objekt a typ

Datové zdroje obsahují nebo generují **datové objekty**. Ke každému **datovému objektu** existuje jeden **typ** dokumentu v EAP. **Typy** EAP tedy reprezentují **datové objekty**.

**Typ** se podobá **tabulce** v relačních databázích.

### Datový záznam a dokument

Datové objekty tvoří instance nebo **záznamy**, které se v EAP nazývají **dokumenty**. Atributy datových **záznamů** se převádějí na **vlastnosti** dokumentů.

Z relačního pohledu odpovídá **dokument řádku** tabulky a **vlastnost** odpovídá **sloupci**.

Na rozdíl od relačního přístupu není při importu dat do EAP potřeba znát jejich strukturu. Součástí EAP je bezschémová databáze, která si dokáže sama vygenerovat struktury z vlastností importovaných dat.

## Konfigurační soubor

Konfigurační soubor config.properties je formátován podle pravidel používaných v jazyce Java a obsahuje několik sekcí

### Datum poslední úpravy

V záhlaví souboru lze umístit nepovinné datum poslední úpravy

```
#Wed Dec 14 18:59:04 CET 2015
```

### Umístění souborů

#### Logovací soubory

```
// LOGging to File
LOG_FilePath=C:\\_Jason_2015-12-15\\_log\\
```

#### Umístění dat pro import

```
// Data location
DATA_BaseDir=/local/data/eap/import
DATA_DirSeparator=/
DATA_NumberOfIndexes=2
```

Stanovuje počet indexů k importu

#### Umístění uzlu elasticsearch

```
// ElasticSearch
ELK_URL=host.example.com
ELK_PORT=9300
ELK_CLUSTER=eap1
```

Hostname uzlu  
Název clusteru, do kterého patří daný uzel

#### Indexy elasticsearch

```
// 1st Index
IND_01_Name=IPPC
IND_01_Index=ippc-test
IND_01_Evaluated=TRUE
IND_01_NumberOfTypes=3
IND_01_Type_01=expert
IND_01_Type_02=company
IND_01_Type_03=appliance
IND_01_Type_04=n/a
IND_01_Type_05=n/a
IND_01_Type_06=n/a
IND_01_Type_07=n/a
IND_01_Type_08=n/a
IND_01_Type_09=n/a
IND_01_Type_10=n/a
IND_01_Type_11=n/a
IND_01_Type_12=n/a
IND_01_Schedule=TBD Later
```

Název konektoru  
Název indexu  
Zpracovat importní službou?  
Počet typů v indexu  
Název prvního typu  
Název druhého typu  
Název třetího typu  
Názvy dalších typů  
je-li uvedeno n/a, parametr se nevyhodnocuje

Doba spouštění (bude použito v dalších verzích)

```
// 2nd Index
```

IND\_02\_Name=CITES  
IND\_02\_Index=cites  
IND\_02\_Evaluated=TRUE  
IND\_02\_NumberOfTypes=2  
IND\_02\_Type\_01=global  
IND\_02\_Type\_02=n/a  
IND\_02\_Type\_03=n/a  
IND\_02\_Schedule=TBD Later



## Datové soubory

Soubory pro import je nutno v souborovém systému uspořádat do předně dané struktury.

Kořenovým adresářem je `DATA_BaseDir`. Na něj navazuje strom adresářů ve dvou úrovních. Na první úrovni je adresář pojmenovaný po indexu `IND_nn_Index`. Na druhé úrovni je adresář pojmenovaný po typu `IND_nn_Type_mm`.

Vlastní soubory s daty formátovanými podle standardu JSON se umísťují **vždy** do adresářů druhé úrovně.

Jméno datového souboru se řídí touto jmennou konvencí:

`index_type_id_timestamp.json`

- index** je tvořen hodnotou parametru `IND_nn_Index`
- type** je tvořen hodnotou parametru `IND_nn_Type_mm`
- id** je jednoznačný identifikátor datového objektu. Tento identifikátor pochází z datového zdroje.
- timestamp** je časová značka platnosti nebo poslední změny datového záznamu. Vzhledem k vlastnostem různých souborových systémů je tvar časové značky zjednodušen.  
Příklad: `2015-12-07T12-17-54-000Z`

Příklad celého názvu souboru: `ippc_expert_mzpj5fig159g_2014-04-04t10-01-29-000z.json`

Příklad adresářového stromu

```
DATA_BaseDir
|
|---_config
|   |---config.properties
|---_log
|   |---logImport_<DateTime>
|   |---...
|   |---logImport_<DateTimeX>
|
|---ippc-test
|   |---expert
|       |---ippc_expert_<IDcisloJedna>.json
|       |---...
|       |---ippc_expert_<IDcisloXXX>.json
|   |---company
|       |---ippc_company_<IDcisloJedna>.json
|       |---...
|       |---ippc_company_<IDcisloYYY>.json
|   |---appliance
|       |---ippc_appliance_<IDcisloJedna>.json
```

```
|      |      |---...  
|      |      |---ippc_appliance_<IDcisloZZZ>.json  
|      |---...  
|      |---...  
|---<LastType>  
|      |---ippc_<LastType>_<IDcisloJedna>.json  
|      |---...  
|      |---ippc_<LastType>_<IDcisloNNN>.json  
|      |...  
|---<system_2>  
|      |---<typ_1>  
|      |      |---<system_2>_<typ_1>_<IDcisloJedna>.json  
|      |      |...  
|      |      |---<system_2>_<typ_1>_<IDcisloMMM>.json  
|      |---...  
|
```

## Spouštění importní služby

Importní služba je implementována jar modulem `ImportModule-1.0.0.CI-SNAPSHOT.jar` uloženým v `DATA_BaseDir`

Pro spuštění je vyžadováno běhové prostředí Java verze 1.8.

Služba se spouští příkazem `java -jar ImportModule-1.0.0.CI-SNAPSHOT.jar`

Pokud se konfigurační soubor nachází v adresáři `DATA_BaseDir /_config/config.properties`, není třeba používat žádné parametry v příkazové řádce.

## Doporučení

Importní službu není nutno spouštět na stejném stroji, na kterém se nachází uzel elasticsearch, ale je to výhodné.

V závislosti na platformě, na které se importní služba spouští lze nastavit systémovými prostředky automatické spouštění nejlépe jedenkrát denně.

## Logy

Každý běh služby je podrobně logován. Soubory logů jsou označeny časovou značkou běhu. Např.

`logImport_2015-12-15_10-19`.

## Konektory

Vlastní importní služba importuje pouze data ve formátu JSON připravená do adresářových struktur uvedených výše. Do této podoby je třeba data přímo získaná z datových zdrojů nějak připravit.

K tomu slouží **konektory importní služby**.

Konektor může data přímo stáhnout z datového zdroje, pokud je tento zdroj vybaven službou pro on-line předávání dat, nebo je najde v souborové podobě, pokud datový zdroj disponuje pouze jednoduchými možnostmi exportu dat.

Z toho vyplývá, že pro každý datový zdroj je třeba samostatný konektor, zatímco importní služba zůstává stále stejná.

Úkolem konektoru je

1. Konvertovat data z datového zdroje (databáze/databází) do indexu/ů v EAP
2. Každý datový objekt (tabulku databáze) datového zdroje převést na typ EAP
3. Každý záznam datového objektu (řádek databáze) identifikovat a na základě jeho atributů (sloupců) převést do dokumentu formátovaného JSON.
4. Konvertovaná data uložit do definovaných adresářů.

## Doporučení

1. Za účelem zvýšení vypovídací hodnoty je vhodné v rámci konverze vytvářet co nejplošší dokumenty a nepoužívat normalizaci obvyklou v relačních databázích. Data jsou při ukládání komprimována a i při značné redundanci nezaberou tolik místa jako data uložená v relačních tabulkách. Kromě toho EAP je distribuovaný systém z obozu veledat (big data), takže nehrozí jeho zahlcení.
2. Je žádoucí v jednotlivých dokumentech doplnit číselníkové kódy o jejich hodnoty a do EAP předávat obojí. Usnadní se tím křížové prohledávání napříč indexy.

## Datové zdroje

Každý datový zdroj se připojuje zvláštním importním konektorem, který co nejjednodušším způsobem převádí data ze zdroje do podoby analyzovatelné v rámci EAP. Importní konektory se liší podle typu datového zdroje.

Importní konektory navrhuje správce EAP v kontextu aktuálního datového mapování EAP.

Existující datové zdroje obvykle nedisponují sofistikovanými exportními službami pro předávání dat do EAP, a proto se importní konektory v takovém případě napojují přímo na exportovaná data, na ručně konsolidovaná exportovaná data nebo na stávající rozhraní (např. stávající webové služby).

Modernizované nebo nově budované datové zdroje musí disponovat **Exportní službou EAP**.

## Typy datových zdrojů

Existují různé typy datových zdrojů

- Databáze
- Jednoduché soubory
- Webové služby
- Jiné zdroje jako např. RSS kanály

Každý typ datového zdroje vyžaduje poněkud jiný přístup pro import do EAP. Importní konektor pro každý typ tedy bude mít poněkud jiné vlastnosti.

### Databáze

Databáze jsou datové zdroje sestávající se z tabulek a jejich vztahů. Importní konektor k nim může přistupovat pomocí ovladače (JDBC, ODBC, proprietární, ...) nebo pomocí knihoven k vlastním datovým souborům obsahujícím databáze.

Importní konektor v takovém případě by měl obsahovat služby

1. **Identifikace** zdrojových datových objektů (tabulek), časových značek platnosti nebo úpravy záznamů a identifikátorů záznamů
2. **Export** dat z datového zdroje do souborové podoby
3. **Konsolidace** exportovaných dat, která „zploští“ data do cílových datových objektů (typů) tak, aby se dala snadněji prohledávat a analyzovat. Při konsolidaci dat se zároveň doplní do dokumentů číselníkové hodnoty.
4. **Převod** konsolidovaných dat do formátu JSON při použití kódování UTF-8
5. Korektní **zařazení** exportovaných dat do adresářové struktury a korektní pojmenování výsledných souborů.

Importní konektor pak správce EAP zaregistruje do EAP a o provedení vlastního importu se pak postará vlastní univerzální importní služba.

Importní konektor je vhodné v případě, kdy existuje on-line připojení k databázi navrhovat jako plně automatický. Aktualizace dat ze zdroje pak probíhá bez lidského zásahu.

### Jednoduché soubory

Jednoduché soubory jsou datové zdroje vzniklé obvykle exportem z nějakého informačního systému, který neumožňuje on-line přístup.

Importní konektor v takovém případě by měl obsahovat služby

1. **Identifikace** zdrojových datových objektů (tabulek), časových značek platnosti nebo úpravy záznamů a identifikátorů záznamů
2. **Konsolidace** exportovaných dat, která „zploští“ data do cílových datových objektů (typů) tak, aby se dala snadněji prohledávat a analyzovat. Při konsolidaci dat se zároveň doplní do dokumentů číselníkové hodnoty. Tato služba se provádí pouze v případě, že ji charakter datového zdroje vyžaduje
3. **Převod** konsolidovaných dat do formátu JSON při použití kódování UTF-8
4. Korektní **zařazení** exportovaných dat do adresářové struktury a korektní pojmenování výsledných souborů.

Importní konektor pak správce EAP zaregistruje do EAP a o provedení vlastního importu se pak postará vlastní univerzální importní služba.

Importní konektor v takovém případě bude navrhován jako poloautomatický, přičemž služby konektoru (viz výše) budou probíhat za účasti operátora.

### Webové služby

Webové služby jsou součástí moderních IT služeb a zpřístupňují datové zdroje on-line. Pokud stávající datové zdroje jsou schopny poskytovat data za účelem dalšího využití prostřednictvím webových služeb, je možné s výhodou tyto služby využít bez nutnosti úpravy datového zdroje.

V případě, že stávající webové služby neumožňují získávat potřebná data, je vhodné nejprve posoudit, zda data tohoto zdroje nejsou přístupná jinou formou (například jako on-line databáze nebo exportované soubory).

Pokud bude zjištěno, že potřebná environmentální data nejsou dostupná žádným z uvedených způsobů, je třeba přistoupit k doplnění některé z možností přístupu k datům.

V případě, že stávající webové služby zpřístupňují potřebná data, postupuje se stejně jako v případě, že datový zdroj je typu Databáze.

1. **Identifikace** zdrojových datových objektů, časových značek platnosti nebo úpravy záznamů a identifikátorů záznamů
2. **Export** dat z datového zdroje do souborové podoby
3. **Konsolidace** exportovaných dat, která „zploští“ data do cílových datových objektů (typů) tak, aby se dala snadněji prohledávat a analyzovat. Při konsolidaci dat se zároveň doplní do dokumentů číselníkové hodnoty.

4. **Převod** konsolidovaných dat do formátu JSON při použití kódování UTF-8
5. Korektní **zařazení** exportovaných dat do adresářové struktury a korektní pojmenování výsledných souborů.

Importní konektor pak správce EAP zaregistruje do EAP a o provedení vlastního importu se pak postará vlastní univerzální importní služba.

Importní konektor je třeba v tomto případě vždy navrhovat jako plně automatický. Aktualizace dat ze zdroje pak probíhá bez lidského zásahu.

### **Jiné zdroje**

Jiné zdroje mohou být přístupné on-li nebo off-line. Podle toho se na ně aplikuje některý z návrhů uvedených výše.



## Mapování dat

EAP dokáže automaticky rozpoznat řadu datových typů. Pokud datové zdroje obsahují běžné typy dat, není bezpodmínečně nutno vytvářet datové mapy.

Pro fulltextové indexování souborových příloh nebo v případě, kdy je třeba lépe ošetřit vypovídací schopnost importovaných dat, je vhodné v rámci návrhu importního konektoru vytvořit ve spolupráci se správcem EAP relevantní datové mapování.

Pro představu analytických možností na základě vhodného využití datových typů a mapování je uveden výčet většiny datových typů, se kterými dokáže EAP pracovat.

## Základní datové typy

### Textový datový typ

`string`

### Numerické datové typy

`long, integer, short, byte, double, float`

### Datový typ datum a čas

`date`

### Logický datový typ

`boolean`

### Binární datový typ

`binary`

## Komplexní datové typy

### Datový typ pole

Podpora polí nevyžaduje žádný specializovaný datový typ

### Objektový datový typ

`object` pro jednoduché objekty JSON

### Vnořené datové typy

`nested` pro pole objektů JSON

## Geografické datové typy

### Datový typ geografický bod

`geo_point` pro body dané vlastnostmi `lat/lon`

### Datový typ geografický tvar

`geo_shape` pro komplexní tvary jako jsou například polygony

## Speciální datové typy

### Datový typ pro IP adresu v4

`ip` pro adresu IPv4

### Kompletační datový typ

`completion` pro poskytování návrhů samočinného dokončování

### Datový typ počet výskytů

`token_count` pro zjištění počtu výskytů určitých příznaků v textu

### Hašovací datový typ

`murmur3` k výpočtu hašů hodnot během indexace a jejich uložení v indexu

### Datový typ příloha

`attachment` podporuje indexování přiložených souborů jako například formáty PDF, Microsoft Office, Open Document, ePub, HTML, atd.

## Vícenásobná pole

Používají se k indexování stejných datových polí různými způsoby pro různé účely. Například pole typu `string` může být indexováno jako analyzované pole pro fulltextové hledání a jako neanalyzované pole pro třídění nebo agregaci. Alternativně lze zároveň indexovat pole typu `string` standardním analyzátozem, českým analyzátozem a anglickým analyzátozem.

To je základním účelem vícenásobných polí. Většina datových typů podporuje vícenásobná pole.

## Využití mapování

Mapování dat a možnost využití aliasů pro indexy výrazně zesiluje analytickou sílu EAP.

Záleží jen na tvůrcích importních konektorů, nakolik budou schopnosti a možnosti analytické platformy využity.

## Datový slovník

Z dokumentu je patrné, že koncepce importní služby a společných analytických služeb EAP umožňuje vytvoření jednotného datového slovníku všech environmentálních dat. Tento slovník není nutno aplikovat na datové zdroje, ale lze jej vytvořit v rámci agregace dat na úrovni EAP.

Datový slovník EAP je jedním z výstupů EAP a může sloužit jako základní vodítko při tvorbě nových a modernizaci existujících datových zdrojů.

## Požadavky předávání dat na nové a modernizované datové zdroje

Za **datové zdroje EAP** jsou považovány veškeré **informační systémy**, které shromažďují environmentální data. Zpravidla se jedná o agendové informační systémy státní správy v oblasti ochrany životního prostředí.

V případě návrhu nového takového systému nebo modernizaci stávajícího, je nutno zajistit předávání environmentálních dat do EAP za účelem jejich analytického zkoumání a vytěžování.

### Identifikace environmentálních dat

V rámci **analytické fáze** projektu nového systému nebo modernizace stávajícího budou ve spolupráci Zadavatele a Řešitele projektu identifikovány datové objekty (třídy), jejichž data budou předávána do EAP. Zpravidla budou předávány časové snímky dat, které budou vytvářeny při každé změně obsahu nebo stavu životního cyklu.

Součástí analytické dokumentace budovaného nebo modernizovaného systému bude **katalog datových tříd, jejichž časové snímky budou předávány do EAP.**

### Předávání environmentálních dat

Environmentální data se do EAP předávají **dávkově**, aby nebyl omezován provoz datového zdroje.

Předávání probíhá zpravidla **jedenkrát denně** v nočních hodinách.

K předávání slouží na straně datového zdroje Exportní služba EAP, což je webová služba realizující poskytující operace uvedené níže.

K Exportní službě daného systému se pomocí specifického konektoru aktivně připojuje Importní služba EAP, která přebírá připravená data a označuje, která data již byla úspěšně převzata. To umožní agendovému informačnímu systému průběžně odmazávat časové snímky předané EAP.

### Exportní služba datového zdroje EAP

Každý nový nebo modernizovaný datový zdroj (zdroj environmentálních dat) musí disponovat Exportní službou datového zdroje EAP. Podoba exportní služby se liší podle toho, zda systém spravuje životní cykly datových objektů, případně zda by napojení exportní služby na životní cykly u modernizovaných systémů nevyvolalo přílišné náklady.

Volba metody exportní služby se provede v rámci **analytické fáze** projektu nového systému nebo modernizace stávajícího budou ve spolupráci Zadavatele a Řešitele a zanesse se do analytické dokumentace projektu.

### Metoda „Export změn“

V případě modernizace datového zdroje obsahujícího environmentální data nebo v případě nového systému, který nespravuje životní cykly datových objektů, bude za účelem předávání dat EAP do zdroje doplněna webová služba **wsEap** která bude mít pro každý předávaný datový objekt operaci **getObjectByFilter(filter)**, kde **Object** je název datového objektu (v EAP typ) a **filter** je struktura umožňující filtrování výstupů podle různých kritérií, nejméně však v tomto rozsahu:

```

<filter>
  <id?></id>          Jednoznačný identifikátor záznamu
  <dateFrom?></dateFrom> Datum platnosti/poslední úpravy od
  <dateTo?></dateTo>   Datum platnosti/poslední úpravy do
  <maxRec?></maxRec>   Maximální počet vrácených záznamů (0 vrací vše)
</filter>

```

Tato služba už vrací **konsolidovaná** data, takže není nutno v importním konektoru EAP provádět služby **export** a **konsolidace**, které se provádějí u historických datových zdrojů. Vede to ke zjednodušení celého procesu předávání dat do EAP. Importní konektor EAP pouze sjednotí datové typy tak, aby nedocházelo ke kolizím při vyhledávání.

V případě modernizace stávajících datových zdrojů je vhodné zvážit požadavek na tvorbu **časových snímků** datových objektů a logování podle standardu **syslog**. Pokud totiž datový zdroj slouží ke správě složitějších **životních cyklů** datových objektů, jejich **časové snímkování** při každé **procesní události** výrazně zvýší vypovídací schopnost předávaných dat a navíc umožní využití EAP jako služby správy verzí.

### Metoda „Export životních cyklů“

Pokud budou vytvářeny nové datové zdroje, je třeba zvážit, zda tyto zdroje budou spravovat životní cykly datových objektů či nikoliv.

Pokud v rámci datového zdroje nebude prováděna správa životního cyklu datových objektů, je třeba v těchto zdrojích zajistit časové snímkování těchto objektů.

V opačném případě postačí vytvoření webové služby **wsEap** v rozsahu uvedeném výše.

### Časové snímkování

Časové snímkování je metoda, kterou se při každé změně stavu nebo úpravě obsahu nějakého datového objektů vytvoří „otisk“ datového objektu, který se opatří běžnou systémovou časovou značkou a logem obsahujícím identifikaci události, která změnu způsobila a identifikaci původce změny.

Časová značka se zaznamená v objektu časového snímku do vlastnosti **dateSnapshot**.

Vlastnosti logu se odvíjejí od definice **syslog** viz [RFC 5424](#)

```
<snapshotLog>... Obsah logu podle RFC 5424 </snapshotLog>
```

Případná redundance obsahu není na škodu. Smyslem není úsporné ukládání dat, ale co největší vypovídací schopnost pro analytické účely.

Časové snímky se předávají do EAP pomocí stejné webové služby **wsEap** uvedené výše s tímto rozšířením:

Každý objekt časového snímku má jednoznačný textový identifikátor **idSnapshot**, který je tvořen jednoznačným identifikátorem datového záznamu doplněného o časovou značku časového snímku.

Příklad:

1. Časový snímek záznamu ABCD123465 ze 7.12.2015 13:17:54 středoevropského času:  
<idSnapshot>ABCD123465\_2015-12-07T12:17:54:000Z</idSnapshot>
2. Časový snímek stejného záznamu ABCD123465 ze 13.12.2015 8:12:33 středoevropského času: <idSnapshot>ABCD123465\_2015-12-13T07:12:33:000Z</idSnapshot>

Každá operace webové služby, vracející časové snímky má název a obsah podle vzoru `getObjectSnapshotByFilter(filter)`, kde `Object` je název datového objektu (v EAP typ) a `filter` je struktura umožňující filtrování výstupů podle různých kritérií, nejméně však v tomto rozsahu:

<filter>	
<idSnapshot>?</idSnapshot>	Jednoznačný identifikátor časového snímku
<id>?</id>	Jednoznačný identifikátor datového záznamu
<dateFrom>?</dateFrom>	Datum platnosti/poslední úpravy od
<dateTo>?</dateTo>	Datum platnosti/poslední úpravy do
<maxRec>?</maxRec>	Maximální počet vrácených záznamů (0 - vše)
</filter>	

Navíc webová služba obsahuje operaci umožňující verifikaci předaných časových snímků `storedObjectSnapshot(snapshotList)`, kde `Object` je název datového objektu (v EAP typ) a `fsnapshotList` je struktura umožňující předání zpětné informace o časových snímcích úspěšně zapsaných do EAP.

<fsnapshotList>	
<idSnapshot>?</idSnapshot>	Jednoznačný identifikátor časového snímku
<idSnapshot>?</idSnapshot>	Jednoznačný identifikátor časového snímku
<idSnapshot>?</idSnapshot>	Jednoznačný identifikátor časového snímku
<idSnapshot>?</idSnapshot>	Jednoznačný identifikátor časového snímku
...	
</fsnapshotList>	

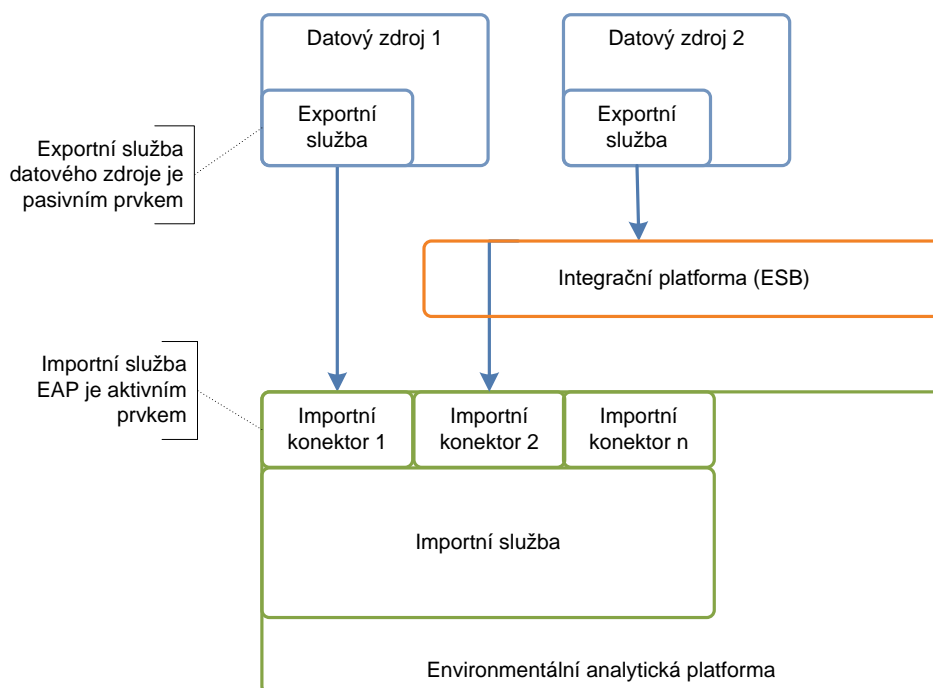
Tato operace umožní datovému zdroji prostřednictvím exportní služby a importního konektoru importní služby EAP zjistit, které časové snímky byly úspěšně zapsány do EAP. Tuto informaci pak může použít k redukci množství uložených dat v produkčním úložišti.

## Schéma komunikace

Předávání dat do EAP u nových a modernizovaných systémů je velmi jednoduché a nenákladné. Každý takový systém je vybaven standardní exportní službou, která připraví data k předání a pak pasivně očekává přístup importní služby EAP, která data aktivně převezme.

Importní služba EAP přistupuje k Exportní službě datového zdroje buď přímo, nebo prostřednictvím integrační platformy, která v případě většího množství datových zdrojů výrazně zjednoduší řízení komunikace. Z toho důvodu musí být exportní služby datových zdrojů navrženy tak, aby je bylo možno snadno a bez vynaložení dalších nákladů zaregistrovat v integrační platformě a dále využívat jejím prostřednictvím.

## Environmentální analytická platforma



Obrázek 1

## Katalog požadavků

Souhrn požadavků na nový nebo modernizovaný agendový informační systém

ID	Požadavek	Akceptační kritérium
EAP1	Existuje katalog datových objektů (třídy), jejichž data budou předávána do EAP	Analytická dokumentace nového nebo modernizovaného IS obsahuje katalog datových objektů s detailním popisem atributů předávaných do EAP
EAP2	Byla zvolena a popsána metoda předávání dat do EAP	Analytická dokumentace nového nebo modernizovaného IS obsahuje detailní rozpracování vybrané metody předávání dat do EAP včetně zdůvodnění <sup>1</sup> .
EAP3	Existuje Exportní služba datového zdroje pro EAP včetně WSDL.	Systém poskytuje webovou službu <b>wsEAP</b> , která implementuje všechny operace vybrané metody předávání dat.
EAP4	Každý předávaný datový objekt je identifikovatelný jednoznačným identifikátorem	Data jsou zaznamenávána tak, aby mohl být každý datový objekt (záznam) předávaný do EAP jednoznačně identifikován jediným a jednoznačným identifikátorem.
EAP5	Existuje detailní dokumentace Exportní služby datového zdroje pro EAP	Součástí dokumentace nového nebo modernizovaného IS obsahuje detailní popis implementace Exportní služby včetně návodu k použití a deskriptivního popisu datových prvků.
EAP6	Exportní služba datového zdroje pro	Služba musí být navržena tak, aby ji bylo možno

<sup>1</sup> Zásadně nebude akceptován výběr metody „Export změn“, pokud nový IS spravuje životní cyklus některých datových objektů (tříd)

ID	Požadavek	Akceptační kritérium
	EAP je dostupná prostřednictvím integrační platformy	zaregistrovat a využívat v rámci budoucí integrační platformy resortu MŽP

---

Zpracoval: *Ing. Radim Jäger*