



FINANČNÍ SPRÁVA

# OPEN data Centrálního registru dotací - IS CEDR III

# OPEN data Centrálního registru dotací

## OBSAH

- obecně o IS CEDR III (5 min)
- konkrétní řešení IS CEDR III (10 min)
- dotazy a diskuse

# Centrální registr dotací

**Spravuje:**

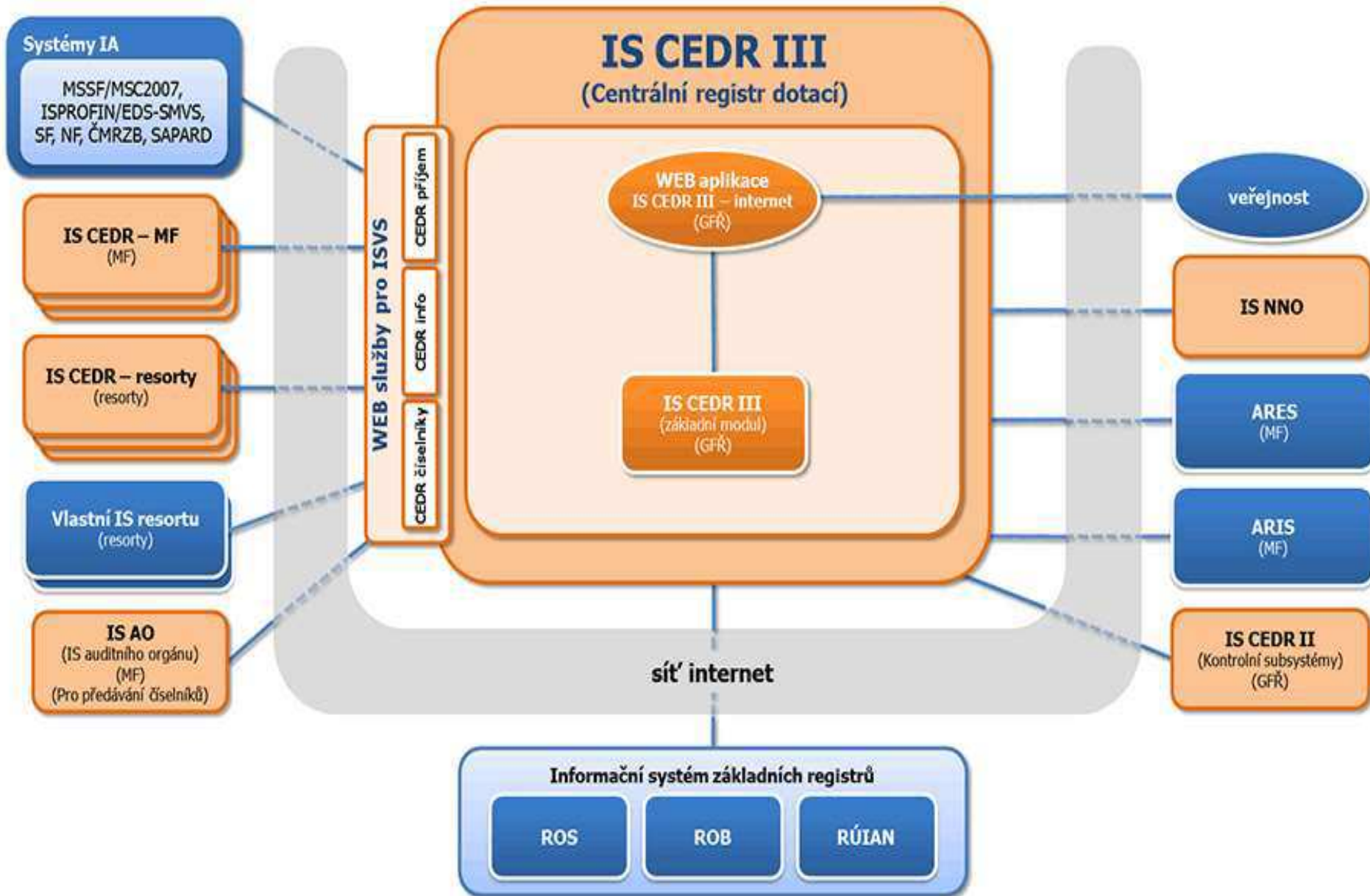
**Generální finanční ředitelství**

nábř. Kpt. Jaroše 1000/7, 170 00 Praha 7

Věcným gestorem systému je **Odbor nefiskálních agend GFŘ**

**Informace o systému** včetně návazných systémů a subsystémů:

<http://cedr.mfcr.cz/cedr3internetv419/CedrInfoISPAGES/CedrInfoPage.aspx>



# OPEN data Centrálního registru dotací

## 3.1. Stručná informace o datových sadách v IS CEDR III a použité standardy

- Stupně otevřenosti datových sad a použité standardy publikace dat IS CEDR
- Princip propojitelných dat
- Základní a pomocné datové sady
- Popis a dostupnost datových sad

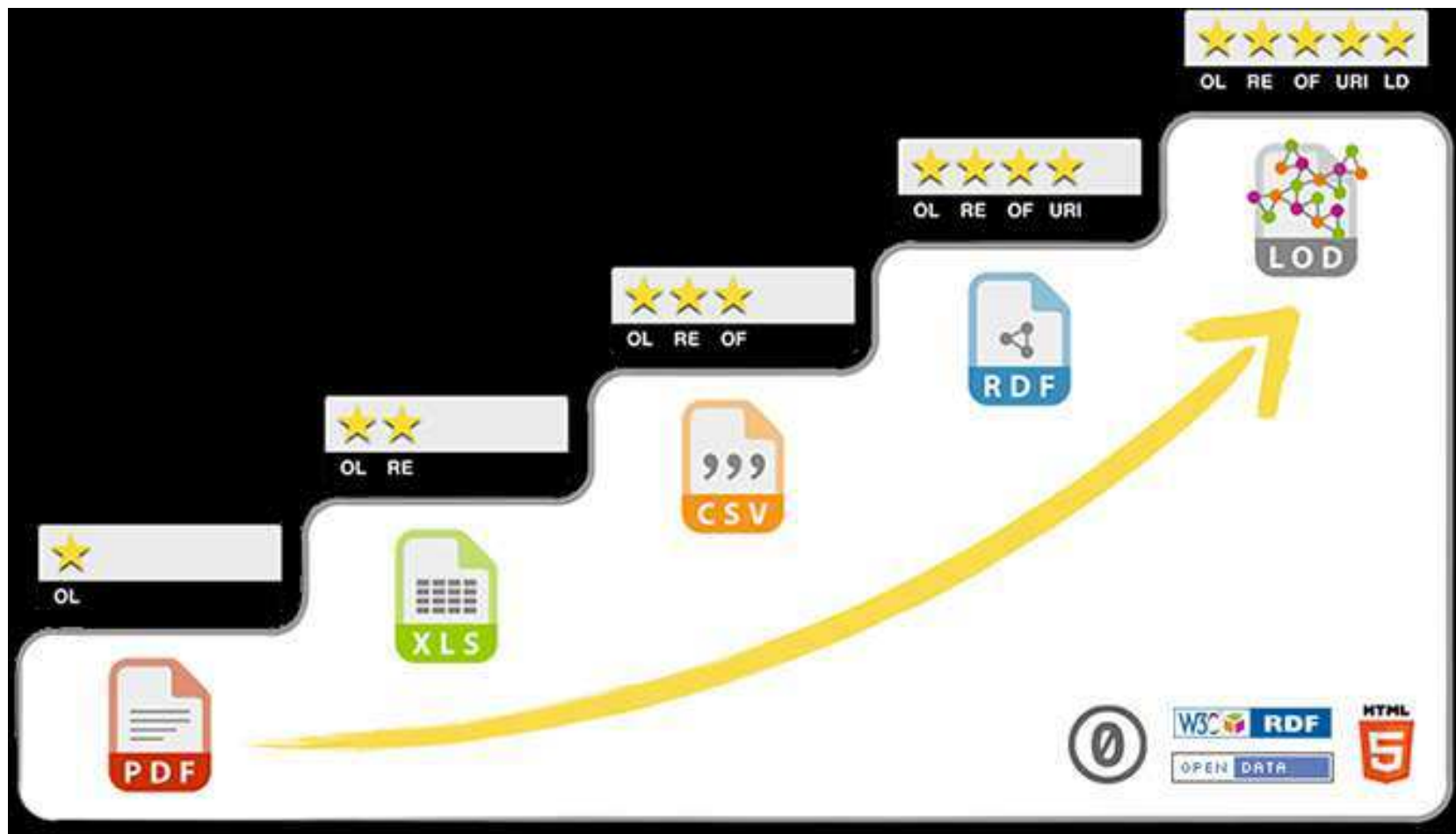
## 3.2. Umístění datových sad v katalogu datových sad

## 3.3. Využívání dat z datových sad

## 3.5. Použité technologie

## 3.6. Technická specifikace

# Stupně otevřenosti datových sad



# Použité standardy publikace dat IS CEDRIII

- **Všechny publikované datové sady splňují nejvyšší stupeň otevřenosti datových sad dle dokumentu Standardy publikace a katalogizace otevřených dat dle doporučení Ministerstvo vnitra České republiky - Standardy publikace a katalogizace otevřených dat VS ČR**
- **Viz.**  
**[http://opendata.gov.cz/ media/standardy publikace a katalogizace otevrenych dat vs cr.pdf](http://opendata.gov.cz/media/standardy_publicace_a_katalogizace_otevrenych_dat_vs_cr.pdf)**
- **Důvodem použití nejvyššího stupně 5, byl požadavek na samopopisnost dat, strojové zpracování data automatickým způsobem a jednoznačná identifikace entit v rámci internetu pro jednoduché a jednoznačné použití dat externími aplikacemi.**

# Použité standardy publikace dat IS CEDRIII

**Nejvyšším stupněm otevřenosti 5 se rozumí splnění následujících podmínek:**

- **datová sada je dostupná v síti WWW s vhodnými podmínkami užití**
- **data jsou poskytována ve strojově čitelném formátu, který umožňuje automatizované zpracování,**
- **formát dat musí být otevřený s volně dostupnou specifikací,**
- **k identifikaci entit datové sady jsou použita IRI,**
- **datová sada splňuje standard propojených dat.**



# Princip propojitelných dat

## IRI PrijemcePomoci

<http://cedropendata.mfcr.cz/c3lod/cedr/resource/PrijemcePomoci/CBE70417ABA6E27FE9E938E28DBEE71A5623DC47>

<a href="#">dcterms:title</a>	Jihomoravský kraj
<a href="#">bus:obchodniJmeno</a>	Jihomoravský kraj
<a href="#">bus:ico</a>	70888337
<a href="#">cedr:sidliNaAdrese</a>	<a href="http://cedropendata.mfcr.cz/c3lod/cedr/resource/Adresa/E0C78A9073B188DF72C4AB54FE87E236ADD570CA">http://cedropendata.mfcr.cz/c3lod/cedr/resource/Adresa/E0C78A9073B188DF72C4AB54FE87E236ADD570CA</a>
<a href="#">owl:sameAs</a>	<a href="http://cedropendata.mfcr.cz/c3lod/ares/resource/EkonomikaSubjekt/70888337">http://cedropendata.mfcr.cz/c3lod/ares/resource/EkonomikaSubjekt/70888337</a>

## IRI Adresa

<http://cedropendata.mfcr.cz/c3lod/cedr/resource/Adresa/E0C78A9073B188DF72C4AB54FE87E236ADD570CA>

<a href="#">dcterms:title</a>	Brno, Žerotínovo náměstí 449
<a href="#">spa:mestskyObvodMestskaCastNazev</a>	Brno-střed
<a href="#">spa:uliceNazev</a>	Žerotínovo náměstí
<a href="#">spa:mestskyObvodMestskaCastKod</a>	550973 (xsd:integer)
<a href="#">ruian:mestskyObvodMestskaCast</a>	<a href="http://cedropendata.mfcr.cz/c3lod/ruian/resource/ciselnik/MestskyObvodMestskaCast/v01/550973/20110701">http://cedropendata.mfcr.cz/c3lod/ruian/resource/ciselnik/MestskyObvodMestskaCast/v01/550973/20110701</a>
<a href="#">spa:adresniMistoKod</a>	19170424 (xsd:integer)

<a href="#">bus:ico</a>	70888337
<a href="#">dcterms:title</a>	Jihomoravský kraj

## IRI ARES :

<http://cedropendata.mfcr.cz/c3lod/ares/resource/EkonomikaSubjekt/70888337>

# Princip propojitelných dat

**Připravenost na napojení na jiné datové sady v ČR:**

- **Seznam aktuálních IČ z registru ARES – vazba pomocí položky IČO.**
- **Registr územní identifikace, adres a nemovitostí (RÚIAN) České republiky – vazba pomocí kódu adresního místa.**

ARES

RÚIAN

# Základní a pomocné datové sady

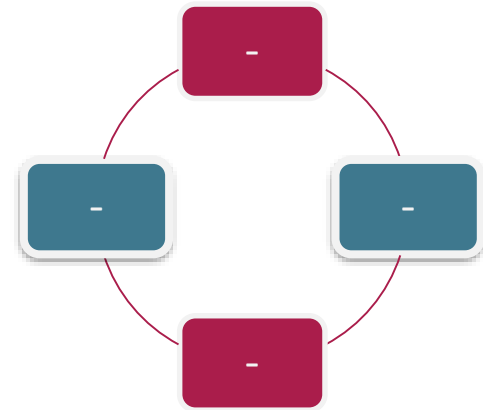
**Otevřená data IS CEDR III se skládají ze :**

- **základní datová sady IS CEDR III**

- příjemcích dotací
- adresách příjemců dotací
- přidělených dotacích jednotlivým příjemcům
- částkách rozhodnutých
- částkách uvolněných, čerpaných, vrácených, v jednotlivých rozpočtových obdobích

- **pomocných datových sad**

- číselníky (RUIAN, ČSU, ISZR, MMR, EDSSMVS)
- informace o subjektech a fyzických osobách (ARES, ROB)
- slovníky ( vazby, prvky a jejich typy odvozené z ISDP )



# Popis a dostupnost datových sad

- **Popis datových sad je dle specifikace VOID - Vocabulary of Interlinked Datasets [2], která je určena pro strojové automatické zpracování s jednoznačnou identifikací datové sady pomocí IRI.**
- **Popis struktury datových sad (slovník datové sady) dle specifikace OWL - Web Ontology Language [3]. Pomocí OWL se definují jednotlivé třídy, vlastnosti tříd, vztahy mezi třídami, popřípadě datové typy vlastností.**
- **K popisu číselníků, byla použita specifikace SKOS - Simple Knowledge Organization System, tak jak je doporučeno v Standardy publikace a katalogizace otevřených dat.**

# Popis a dostupnost datových sad

**Data z datový sad jsou dostupná :**

- **prostřednictvím LOD serveru (formát: RDF, JSON-LD, TURTLE, N-TRIPLES, CSV),**
- **v dumpech souborech rozdělených po třídách ve formátu RDF,**
- **pomocí rozhraní SPARQL – umožňuje uživateli data filtrovat a seskupovat, obdoba SQL jazyka,**
- **na základě připomínek odborné veřejnosti jsme zapracovali publikaci dat ve formátu CSV.**

# Popis a dostupnost datových sad

- **Celkový objem dat datové sady IS CEDR III v RDF formátu cca 35 GB v nekomprimovaném formátu, po komprimaci je objem redukován na 1,2 GB.**
- **Počet trojic v triplestore databázi 129 537 941**
- **Celkový objem dat datové sady IS CEDR III v CSV formátu cca 2,7GB po komprimaci 0,5 GB**
- **Počty řádků dle tříd**

▪ Příjemce pomoci	1 024 482	
▪ Dotace	1 577 672	
▪ Rozhodnutí	1 975 973	rozhodnutá částka 4 826 (v mld.)
▪ Částky za období	2 104 574	uvolněná částka 4 371 (v mld.)
- **Datová sada IS CEDR III je aktualizována jednou za čtvrt roku.**

# Umístění datových sad v katalogu datových sad

**Datová sada IS CEDR III je zveřejněna v Národním katalogu otevřených dat (NKOD)**

**[http://portal.gov.cz/app/RejData/rec.jsp?id=100621008&id\\_rej=97898&y=2016&m=12&doctype=idx](http://portal.gov.cz/app/RejData/rec.jsp?id=100621008&id_rej=97898&y=2016&m=12&doctype=idx) prostřednictvím lokálního katalogu Ministerstva  
Financí (MF) <http://data.mfcr.cz/dataset/dotace-cedr-iii> .**



# Využívání dat z datových sad

## Data z datové sady lze využívat různými způsoby:

### ▪ **online**

- Jednoduché prohlížení vystavených dat pomocí webového prohlížeče
  - **Fulltextové prohledávání**
- aplikacemi zpřístupňující open data IS CEDR III uživatelům přívětivým způsobem. Lze využít online SPARQL rozhraní.

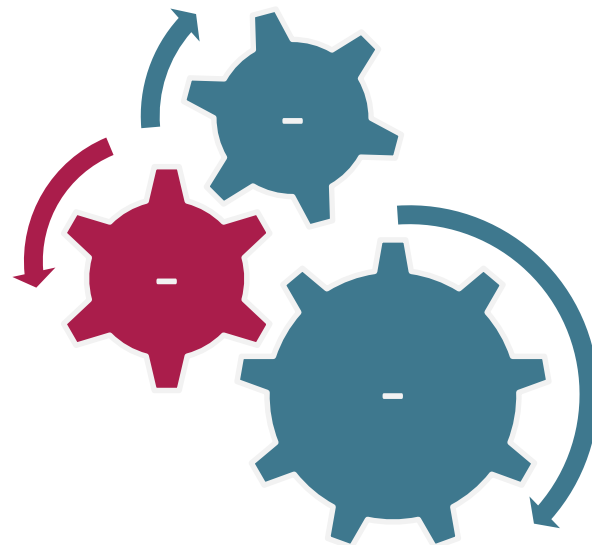
### ▪ **offline**

- po stažení dump souborů uživatel může data transformovat, slučovat a zapojit do vlastních aplikací nebo informačních systémů.

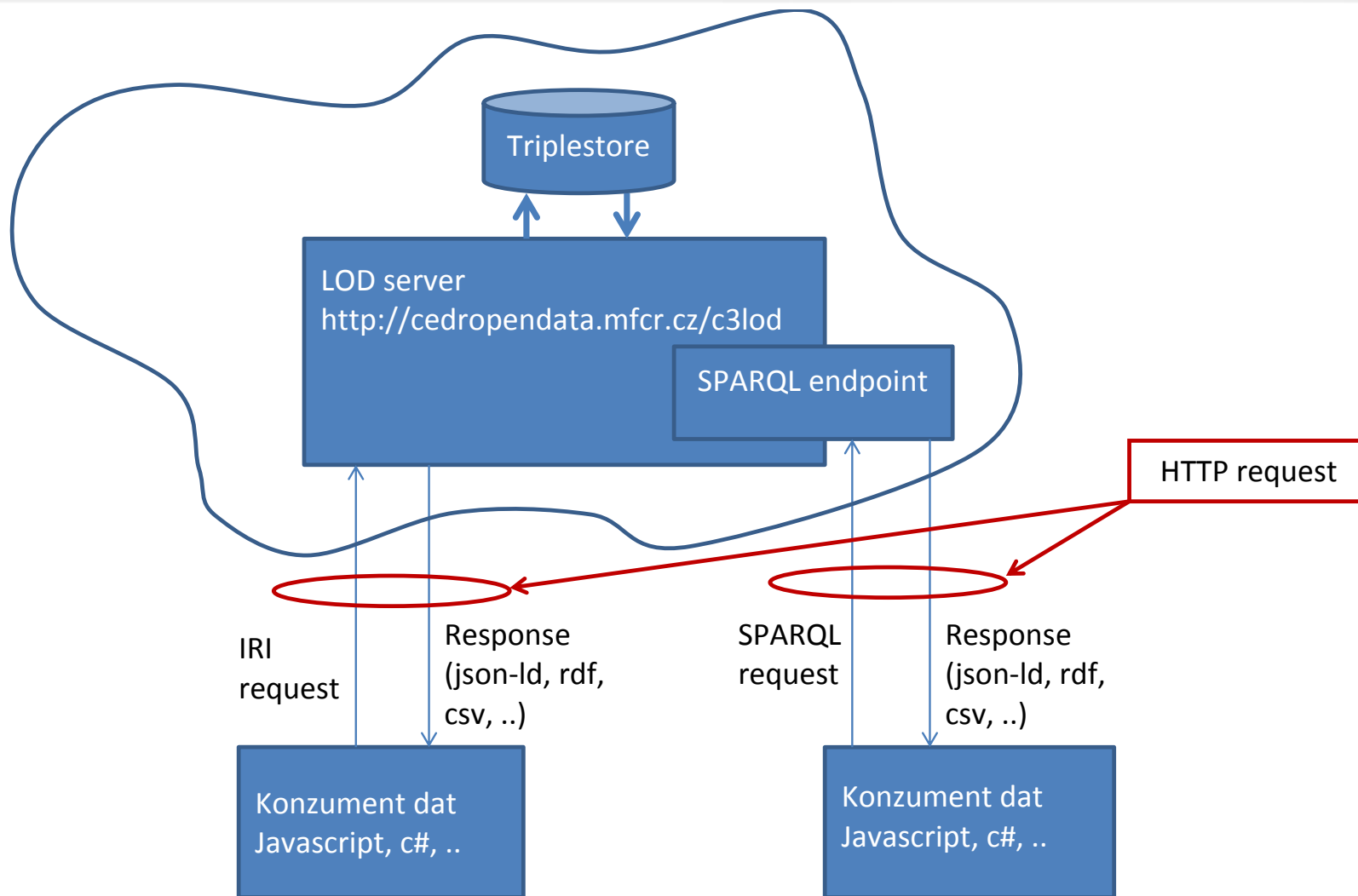


# Použitá technologie

- **Otevřená data IS CEDR III jsou publikována pomocí Linked open data serveru (LOD) impelentovaném technologií .Net provozovaném na IIS 7.5 jehož jádrem je opensource výkonná škálovatelná grafově orientovaná databáze BlazeGraph.**



# Použitá technologie



# Technická specifikace otevřených dat IS CEDR III

- **Kompletní technická specifikace otevřených dat IS CEDR III je popsána v dokumentu C3\_OpenData - datová sada IS CEDR III v2 02.pdf na stránkách <http://cedr.mfcr.cz/cedr3internetv419/OpenData/DocumentationPage.aspx#zalozka3>**



# Technická specifikace otevřených dat IS CEDR III

## ▪ Přístupové body datové sady

- VOID - Datová sada IS CEDR III je popsána [VOID slovníkem](#).
- OWL - Struktura data IS CEDR III je popsána [vlastním slovníkem](#) dle specifikace OWL.
- SPARQL REST API - Rozhraní pro dotazování na data v datové sadě IS CEDR III <http://cedropendata.mfcr.cz/c3lod/cedr/sparql>
- FULLTEXT SEARCH REST API - Rozhraní pro prohledávání datové sady IS CEDR III <http://cedropendata.mfcr.cz/c3lod/cedr/fct>

# Odkazy

**[1] Standardy publikace a katalogizace otevřených dat VS ČR**

**Ministerstvo vnitra České republiky 2015**

**[http://opendata.gov.cz/ media/standardy\\_publicace\\_a\\_katalogizace\\_otevrenych\\_dat\\_vs\\_cr.pdf](http://opendata.gov.cz/media/standardy_publicace_a_katalogizace_otevrenych_dat_vs_cr.pdf)**

**[2] VOID - Vocabulary of Interlinked Datasets (VOID)**

**<http://vocab.deri.ie/void>**

**[3] OWL - Web Ontology Language**

**<https://www.w3.org/TR/owl-ref/>**

**[4] SPARQL - SPARQL 1.1 Query Language**

**<https://www.w3.org/TR/2013/REC-sparql11-query-20130321/>**

# Odkazy

## [5] SKOS - Simple Knowledge Organization System

<https://www.w3.org/2004/02/skos/>



# Dotazy a diskuze

**Zodpovězení dotazů, diskuze.**

