



European Union  
European Social Fund  
Operational Programme Employment

# Linked Open Data in Czech Open Data Portal

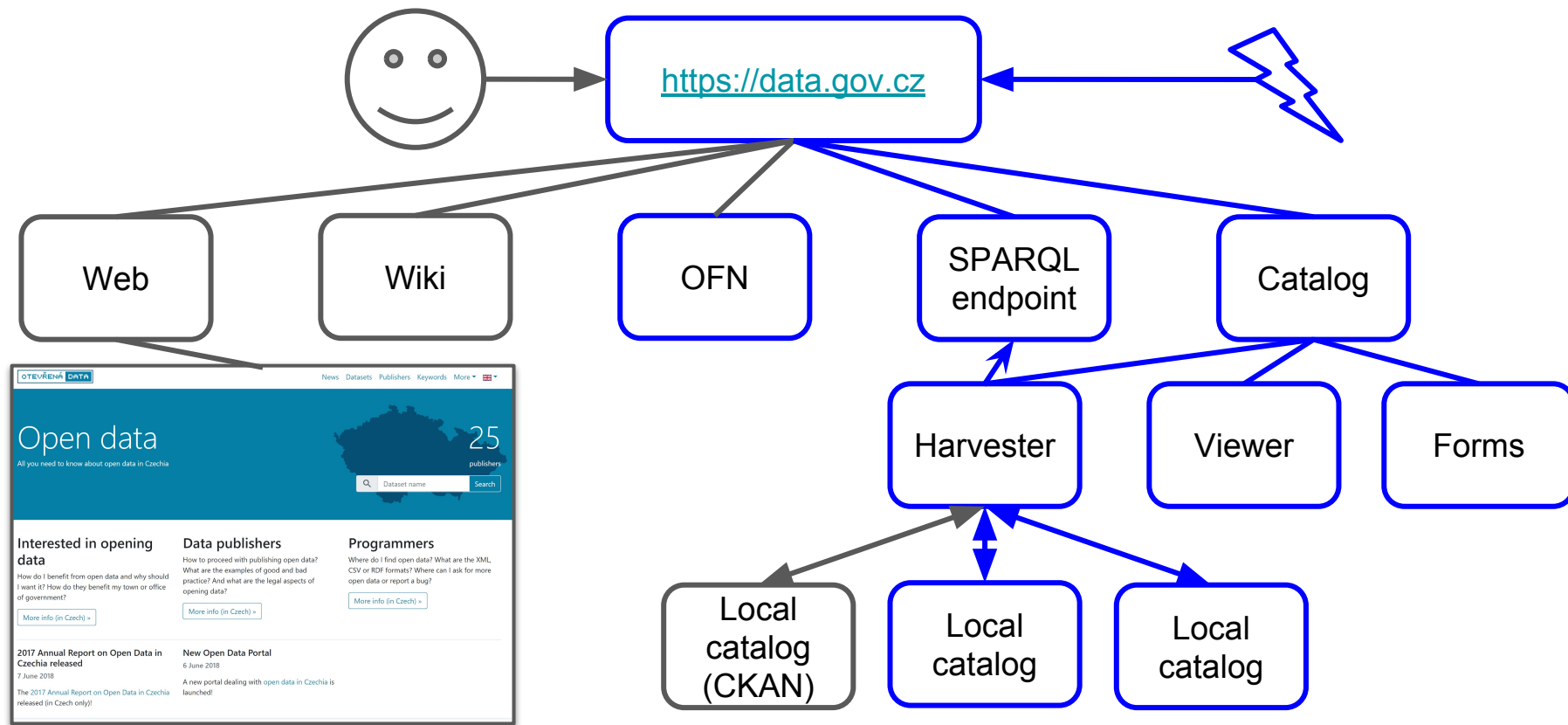
Jakub Klímek

**Project:** Implementace strategií v oblasti otevřených dat II  
**Reg. number:** CZ.03.4.74/0.0/0.0/15\_025/0004172

# Czech Open Data Portal (ODP) architecture and LOD



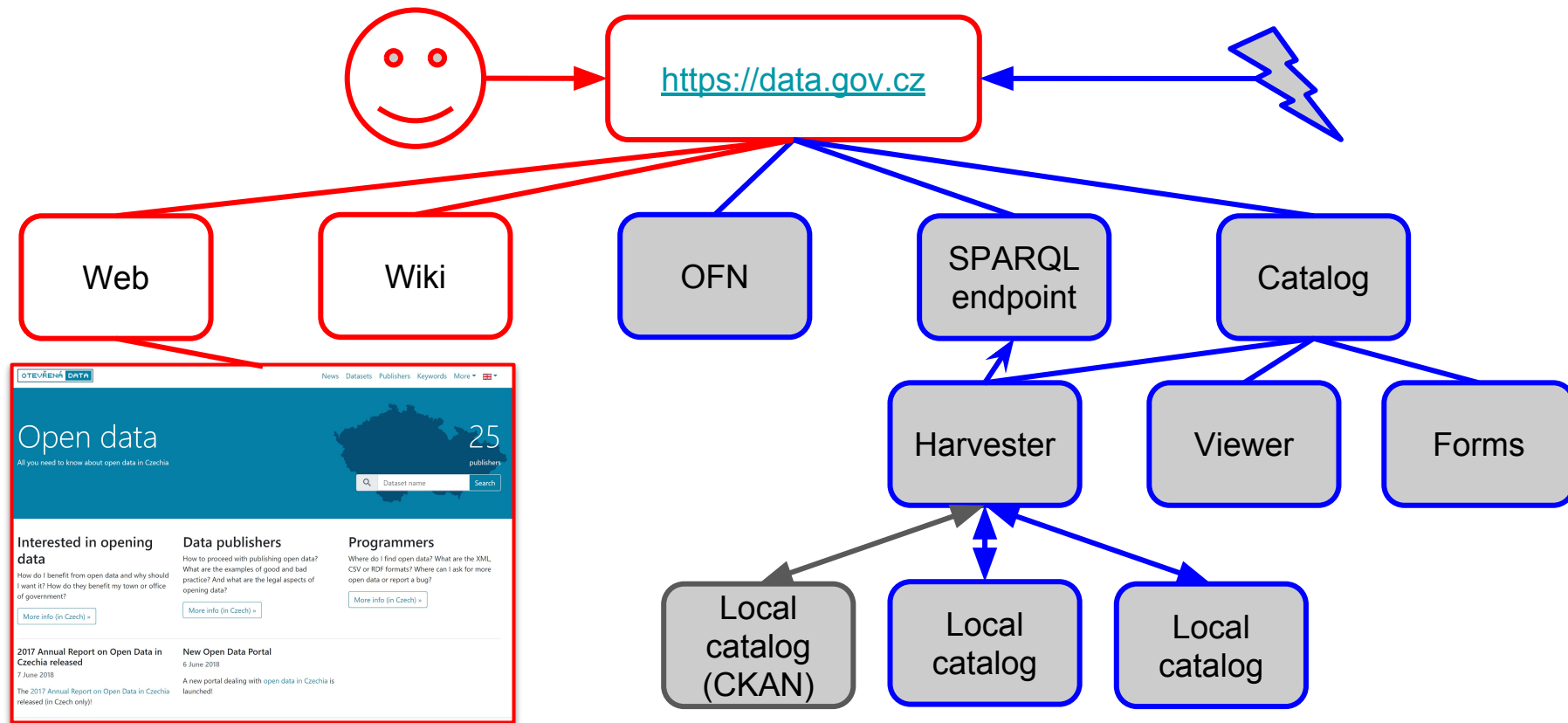
European Union  
European Social Fund  
Operational Programme Employment



# Czech Open Data Portal (ODP) architecture and LOD



European Union  
European Social Fund  
Operational Programme Employment



# Open Data Portal (ODP) Web and Wiki

Web for everyone interested in OD

OTEVŘENÁ DATA

News Datasets Publishers Keywords More

## Open data

All you need to know about open data in Czechia

25 publishers

Dataset name Search

### Interested in opening data

How do I benefit from open data and why should I want it? How do they benefit my town or office of government?

[More info \(in Czech\) »](#)

### Data publishers

How to proceed with publishing open data? What are the examples of good and bad practice? And what are the legal aspects of opening data?

[More info \(in Czech\) »](#)

### Programmers

Where do I find open data? What are the XML, CSV or RDF formats? Where can I ask for more open data or report a bug?

[More info \(in Czech\) »](#)

**2017 Annual Report on Open Data in Czechia released**  
7 June 2018  
The 2017 Annual Report on Open Data in Czechia released (in Czech only!)

**New Open Data Portal**  
6 June 2018  
A new portal dealing with [open data in Czechia](#) is launched!



European Union  
European Social Fund  
Operational Programme Employment

Wiki with info for organizations publishing OD

OTEVŘENÁ DATA Otevřená data

Hledat

Úvodní stránka / Postup otevírání dat / Vytvoření publikačního plánu

Historie / Otevřená data v ČR: Portal pro poskytovatele / Publikace datové sady v otevřené podobě / Vytvoření publikačního plánu

## Vytvoření publikačního plánu

Proces vytvoření publikačního plánu je zobrazen v diagramu níže. Ve zkratce zahrnuje:

- stanovení rolí pro otevírání dat
- ukotvení odpovědnosti a postupu publikace do interní legislativy
- výběr datových sad k otevření (posouzení přínosů, rizik publikace)
- stanovení harmonogramu publikace (s ohledem na technickou náročnost)
- schválení publikačního plánu vedením
- pravidelné vyhodnocování stavu publikace

Po kliknutí na činnost se dostanete na její popis. Jednotlivé převedené dráhy odpovídají rolím, které by se měly procesu otevírání dat účastnit. Po kliknutí na roli se dostanete na její popis.

**Diagram:**

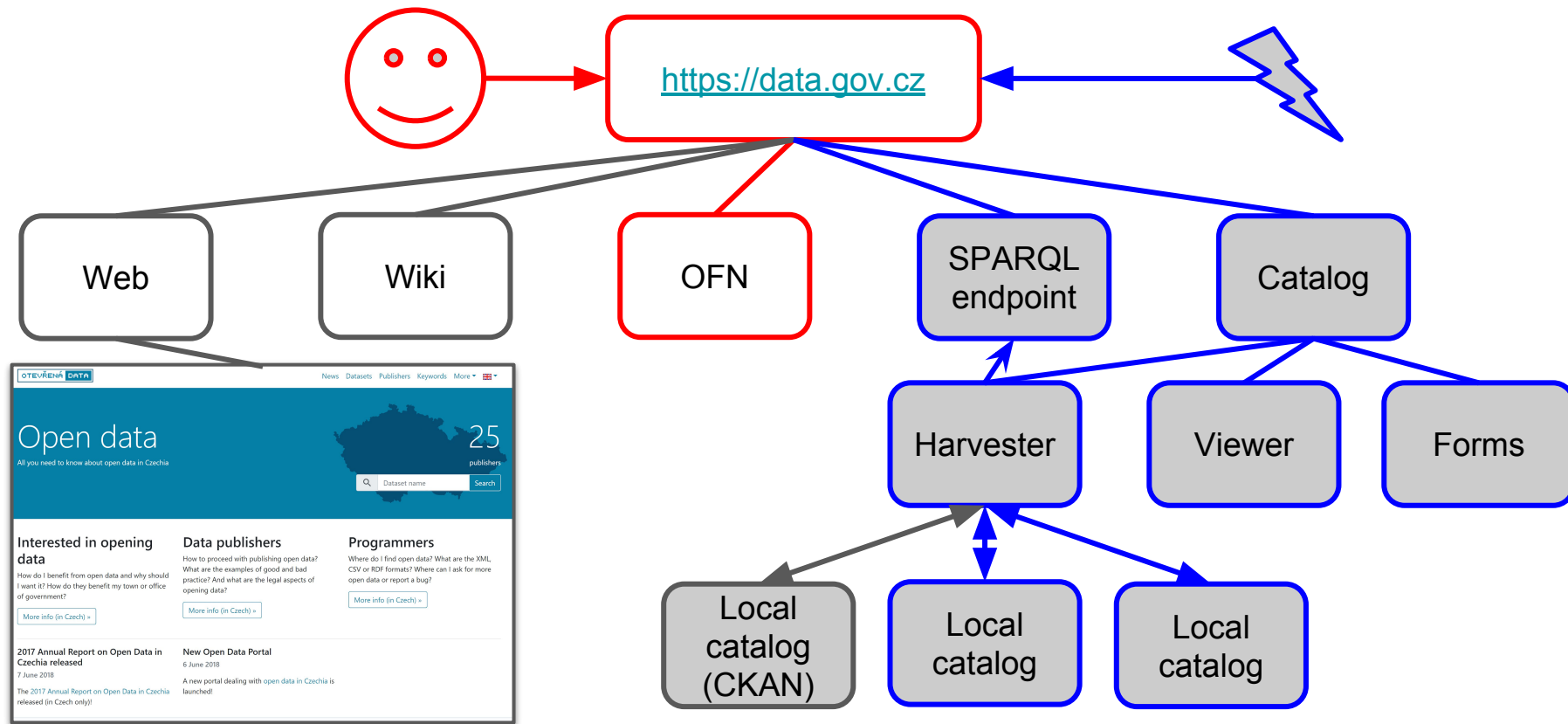
```
graph TD
    subgraph "Vedení poskytovatele dat"
        A[Rozhodnutí o otevírání dat] --> B[Obsazení role Koordinátora otevírání dat]
        B --> C[Schválení interní legislativy]
        C --> D[Obsazení dalších rolí]
        D --> E[Schválení publikačního plánu]
    end
    subgraph "Kurzor dat"
        F[Návrh datových sad ke zveřejnění] --> G[Příprava publikačního plánu]
        G --> H[Rozhodnutí o způsobu katalogizace]
        H --> I[Zveřejnění publikačního plánu]
        I --> J[Periodické vyhodnocování plnění publikačního plánu a informací vedení]
    end
    A --> F
    B --> G
    C --> H
    D --> I
    E --> J
```

On [GitHub](#), using [Jekyll](#)

# Czech Open Data Portal (ODP) architecture and LOD



European Union  
European Social Fund  
Operational Programme Employment



# Open Formal Norms (OFN)

## Standards Anchored in Czech FOIA (106/1999)



European Union  
European Social Fund  
Operational Programme Employment

- Standard for Linked Data (itself)
  - What is URI, IRI, URL
  - IRI patterns
  - IRIs of global entities
    - Addresses
    - Regions
    - Buildings
    - ...
  - RDF serializations
  - Common vocabularies
    - SKOS, The RDF Data Cube, DCAT, VoID
    - RDFS, OWL
- On GitHub, using W3C's ReSpec

Pracovní verze



## Propojená data

Draft 24. září 2018

**Tato verze:**  
<https://data.gov.cz/otevřené-formální-normy/propojená-data/draft/>

**Nejnovější verze:**  
<https://data.gov.cz/otevřené-formální-normy/propojená-data/>

**Předchozí verze:**  
<https://opendata.gov.cz/standardy:technicke-standardy-pro-datove-sady-na-stupni-5>

**Editor:**  
Jakub Klímek (Ministerstvo vnitra)

**Zúčastněte se:**  
[GitHub opendata-mvcr/otevrene-formalni-normy](#)  
[Nahlaste bug](#)

**Odkazy:**  
[Portál otevřených dat](#)

---

### Abstrakt

Cílem tohoto dokumentu je stanovit konvence, jejichž dodržování usnadní vybudování ekosystému propojených otevřených dat v rámci veřejné správy České republiky. Propojená data představují nejvyšší standard publikace dat na webu a jeho dosažení odstraní řadu překážek ve využívání dat veřejné správy.

Dokument začíná principy propojených dat, dále stanovuje pravidla, jak mají v rámci České republiky vypadat identifikátory IRI jednak obecně, a jednak pro vybrané referenční entity. Dále je popsán datový model RDF a jeho serializace a pravidla pro přenos RDF dat pomocí protokolu HTTP(S) přes Internet. Diskutovány jsou vhodné slovníky pro vybrané typy dat a následně také pravidla pro metadata propojených dat. Na závěr doporučujeme formulace do smluv s dodavateli systémů pracujících s propojenými daty.



Evropská unie  
Evropský sociální fond  
Operační program Zaměstnanost

ReSpec

# Open Formal Norms (OFN)

## Standards Anchored in Czech FOIA (106/1999)



European Union  
European Social Fund  
Operational Programme Employment

Pracovní verze

```
OPTIONAL { ?Agenda a-sgov-104-pojem:ma-platnost-agendy-do ?MaPlatnostAgendyDo . }  
OPTIONAL { ?Agenda a-sgov-104-pojem:ma-datum-posledni-zmeny-agendy ?MaDatumPosledniZmenyAgendy  
}  
LIMIT 100
```

ReSpec

Následující SPARQL dotaz vrací instance typu Agenda, pro které jejich vlastnost **Má název agendy** nabývá určité zadané hodnoty. Dotaz je typu SELECT, tudíž vrací tabulku.

**PŘÍKLAD 2:** Instance typu Agenda s danou hodnotou vlastnosti Má název agendy

[Spustit dotaz \(HTML\)](#) | [Spustit dotaz \(CSV\)](#)

```
PREFIX l-sgov-sbirka-111-2009-pojem: <https://slovník.gov.cz/legislativní/sbirka/111/2009/pojem/>  
PREFIX a-sgov-104-pojem: <https://slovník.gov.cz/agendový/104/pojem/>
```

```
SELECT *  
WHERE {  
  ?Agenda a l-sgov-sbirka-111-2009-pojem:agenda .  
  OPTIONAL { ?Agenda l-sgov-sbirka-111-2009-pojem:ma-nazev-agendy ?MaNazevAgendy . FILTER (LANG(?  
  OPTIONAL { ?Agenda l-sgov-sbirka-111-2009-pojem:ma-kod-agendy ?MaKodAgendy . }  
  OPTIONAL { ?OhlasujeAgendu l-sgov-sbirka-111-2009-pojem:ohlasuje-agendu ?Agenda . }  
  OPTIONAL { ?Agenda a-sgov-104-pojem:ma-datum-vzniku-agendy ?MaDatumVznikuAgendy . }  
  OPTIONAL { ?Agenda a-sgov-104-pojem:ma-datum-zaniku-agendy ?MaDatumZanikuAgendy . }  
  OPTIONAL { ?Agenda a-sgov-104-pojem:ma-platnost-agendy-od ?MaPlatnostAgendyOd . }  
  OPTIONAL { ?Agenda a-sgov-104-pojem:ma-platnost-agendy-do ?MaPlatnostAgendyDo . }  
  OPTIONAL { ?Agenda a-sgov-104-pojem:ma-datum-posledni-zmeny-agendy ?MaDatumPosledniZmenyAgendy  
  FILTER (STR(?MaNazevAgendy) = "Agenda řidičů")  
}  
LIMIT 100
```

Následující SPARQL dotaz vrací instance typu Agenda, pro které jejich vlastnost **Má kód agendy** nabývá určité

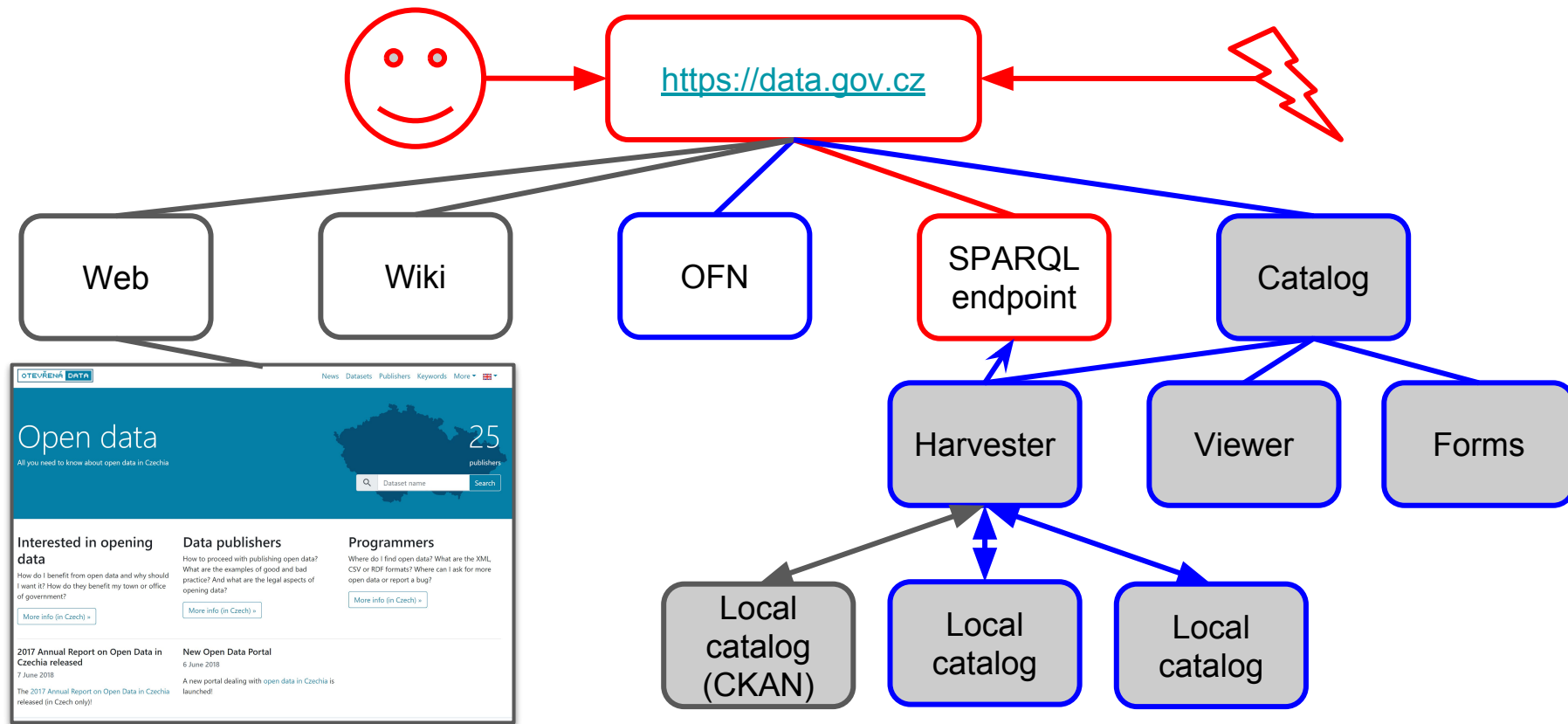
## Linked Data supported standards

- Standards for specific datasets
  - e.g. Addresses, Invoices, ...
- Conceptual overview
- Serializations in CSV, JSON, XML, RDF
- Examples of schemas
- Examples of SPARQL queries
- On GitHub, using W3C's ReSpec

# Czech Open Data Portal (ODP) architecture and LOD



European Union  
European Social Fund  
Operational Programme Employment





# SPARQL Endpoint

<https://data.gov.cz/sparql>

- [OpenLink Virtuoso Open-Source](#)
- [DCAT-AP v1.2](#) content of National Open Data Catalog
- Harvested by [EDP](#) via [SPARQL 1.1 HTTP Graph Store Protocol](#)
- Used for statistics and metadata quality measurements
- Used for dereference of IRIs



European Union  
European Social Fund  
Operational Programme Employment

### Virtuoso SPARQL Query Editor

Default Data Set Name (Graph IRI)

Query Text

```
PREFIX dcat: <http://www.w3.org/ns/dcat#>
PREFIX dct: <http://purl.org/dc/terms/>
PREFIX foaf: <http://xmlns.com/foaf/0.1/>

SELECT DISTINCT ?poskytovatel ?datova_sada ?nazev
WHERE {
  GRAPH ?g {
    ?datova_sada a dcat:Dataset;
    dct:title ?nazev;
    dct:publisher/foaf:name ?poskytovatel.
  }
}
```

LIMIT 100

*(Security restrictions of this server do not allow you to retrieve remote RDF data, see [details](#).)*

Results Format: 

HTML

Execution timeout: 

0

 milliseconds *(values less than 1000 are ignored)*

Options:

☒ Strict checking of void variables

☐ Log debug info at the end of output (has no effect on some queries and output formats)

☐ Generate SPARQL compilation report (instead of executing the query)

*(The result can only be sent back to browser, not saved on the server, see [details](#))*

Run Query

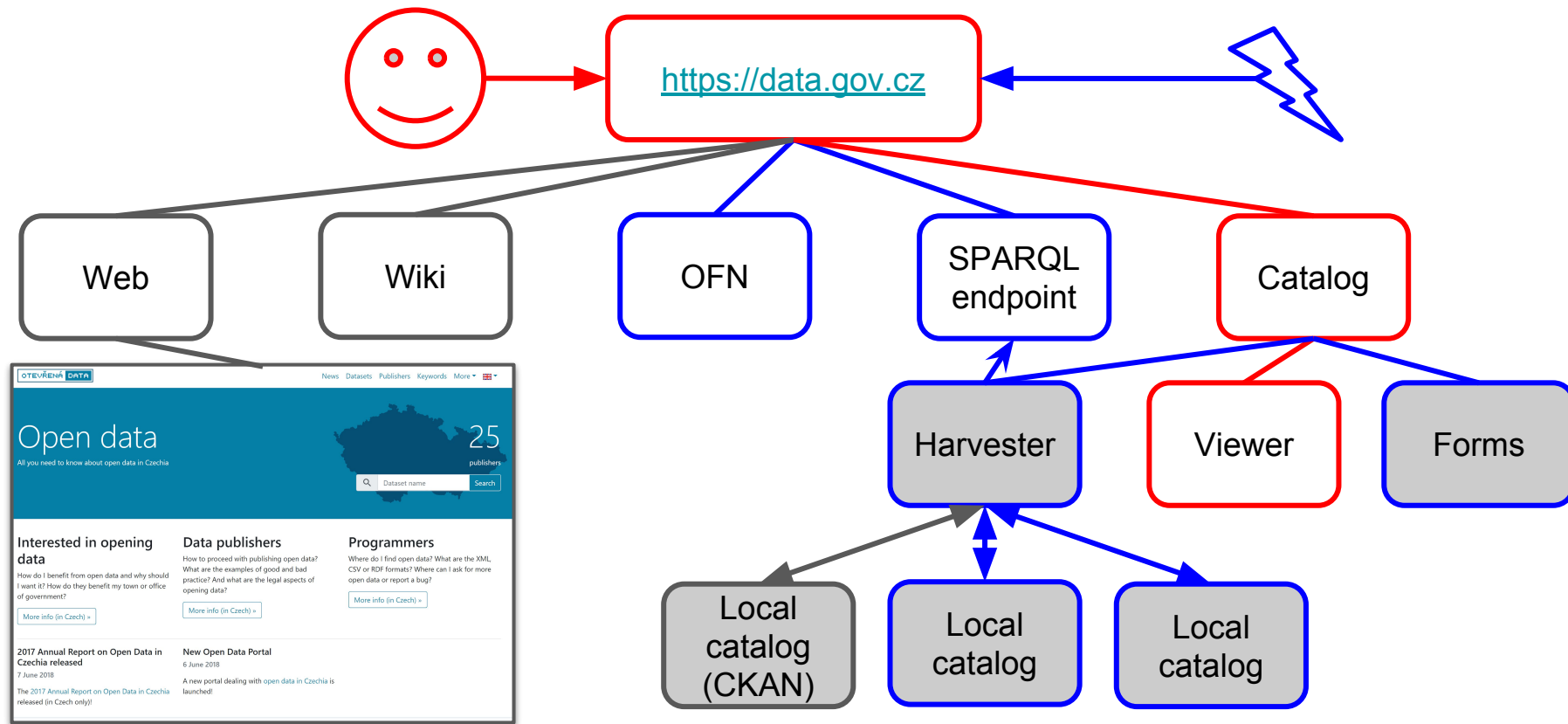
Reset

Copyright © 2018 [OpenLink Software](#)  
Virtuoso version 07.20.3230 on Linux (x86\_64-pc-linux-gnu), Single Server Edition

# Czech Open Data Portal (ODP) architecture and LOD



European Union  
European Social Fund  
Operational Programme Employment



# National Open Data Catalog

## LinkedPipes DCAT-AP Viewer



European Union  
European Social Fund  
Operational Programme Employment

- LP-DAV
- Viewer for people
- Open-source, [GitHub](#)
- Built around DCAT-AP
  - in contrast to CKAN, DKAN
- Czech and English UI
- Uses Apache Solr and Apache CouchDB for faster access

The screenshot displays the LinkedPipes DCAT-AP Viewer interface. At the top, there is a navigation bar with the text "OTEVŘENÁ DATA" and a search bar. Below the navigation bar, the "Publisher (25)" section lists several publishers with their names and the number of datasets they have published. The publishers listed are: Český úřad zeměměřický a katastrální (128 156), Český statistický úřad (528), HLAVNÍ MĚSTO PRAHA (238), Statutární město Plzeň (158), Český telekomunikační úřad (98), Statutární město Ostrava (91), and Česká správa sociálního zabezpečení (73). A "Show more" button is located at the bottom of this list.

The search results section shows "129 674 datasets found". The first result is titled "Žádosti o vydání občanského průkazu bez strojově čitelných údajů" by Statutární město Ostrava. The description states: "Datová sada obsahuje údaje o počtu podaných žádosti o vydání občanského průkazu bez strojově čitelných údajů v jednotlivých měsících evidovaných odborem vnitřních věcí Magistrátu města Ostravy". A "CSV" button is available for this dataset.

The second result is titled "Žádosti o vydání elektronického občanského průkazu se strojově čitelnými údaji a s kontaktním elektronickým čipem" by Statutární město Ostrava. The description states: "Datová sada obsahuje údaje o počtu podaných žádosti o vydání občanského průkazu se strojově čitelnými údaji a s kontaktním elektronickým čipem v jednotlivých měsících, evidovaných odborem vnitřních věcí Magistrátu města Ostravy". A "CSV" button is also available for this dataset.

# National Open Data Catalog

## LinkedPipes DCAT-AP Viewer



European Union  
European Social Fund  
Operational Programme Employment

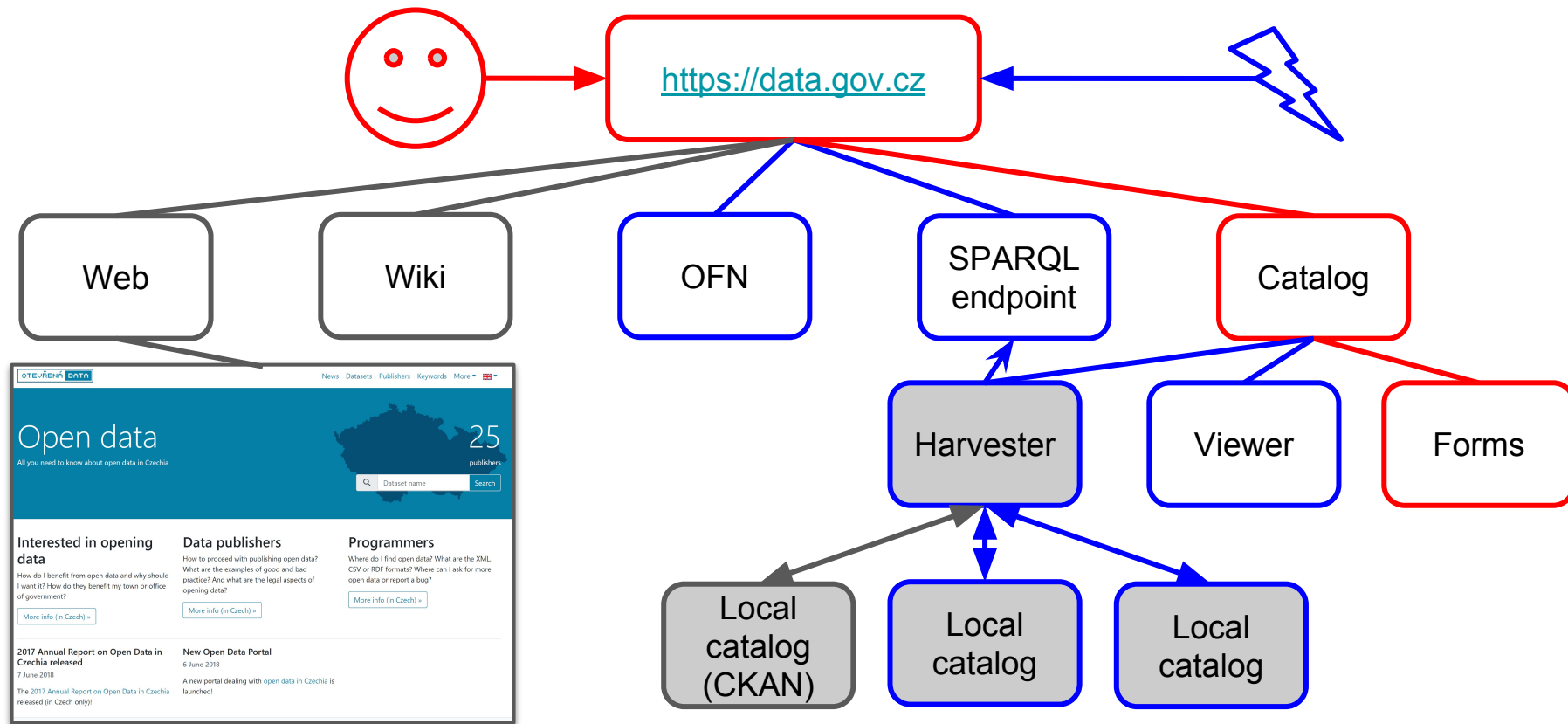
- LP-DAV
- Viewer for people
- Open-source, [GitHub](#)
- Built around DCAT-AP
  - in contrast to CKAN, DKAN
- Czech and English UI
- Uses Apache Solr and Apache CouchDB for faster access
- Some custom features
  - e.g. multilingual EuroVoc theme word cloud =>

The screenshot displays the 'OTEVŘENÁ DATA' (Open Data) interface. On the left, a list of publishers is shown, including 'Český úřad zeměměřický a katastrální (129 275)', 'Český statistický úřad (528)', 'HLAVNÍ MĚSTO PRAHA (238)', 'Statutární město Plzeň (158)', 'Český telekomunikační úřad (98)', 'Statutární město Ostrava (91)', and 'Česká správa sociálního zabezpečení (73)'. Below this is a 'Show more' button. The main section on the right features a search bar and filters. A large word cloud of terms is displayed, including 'parking area', 'public finance', 'pension scheme', 'http://eurovoc.europa.eu/131636', 'real estate market', 'financial aid', 'building plot', 'administrative control', 'supervisory body', 'retirement conditions', 'electronic government', 'Moravia-Silesia', 'incapacity for work', 'result of the vote', 'staff', 'demography', 'sport', 'road services department', 'local authority', 'EU financing', 'non-governmental organisation', 'health insurance', 'self-employed person', 'land register', 'ministry', 'regional and local authorities', 'zoo', 'enforcement of ruling', 'forest', 'budgetary control', 'budget policy', 'watercourse', 'land and buildings', 'rail network', 'road network', 'administrative unit', 'retired person', 'local legislation', and 'atlas'. The bottom section shows 'Themes (118)' with a list including 'parking area (138)'.

# Czech Open Data Portal (ODP) architecture and LOD



European Union  
European Social Fund  
Operational Programme Employment



# National Open Data Catalog

## LinkedPipes DCAT-AP Forms



European Union  
European Social Fund  
Operational Programme Employment

- LP-DAF
- Forms for people
- Open-source, [GitHub](#)
- Built around DCAT-AP
  - in contrast to CKAN, DKAN
- Czech and English UI
- Autocomplete for DCAT-AP codelists (Solr)
- Thorough help for each item
- Producing DCAT-AP v1.2 JSON-LD files

The screenshot displays the 'Dataset' tab of the LP-DAF form. The form is structured with a top navigation bar containing three tabs: 'Dataset' (active), 'Distribution', and 'Summary'. The main content area is divided into several sections, each with a title and a form field. The 'Dataset title' field contains 'Přístupy k agendám CzechPOINT - 2009'. The 'Dataset description' field contains a detailed text about the data. The 'Dataset theme' field is set to 'Government and public sector'. The 'Accrual periodicity' field is set to 'never'. The 'Keywords' field contains three tags: 'výpis', 'agendy', and 'CzechPOINT'. The 'Spatial coverage - RÚIAN territorial unit type' field is set to 'State territory'. The 'Spatial coverage - RÚIAN territorial unit code' field is set to 'Česká republika (1)'. The 'Temporal coverage from (optional)' field is set to '2009-12-31'. The 'Temporal coverage to (optional)' field is set to '2009-01-01'. The 'Name of the dataset curator (optional)' field is empty. The 'E-mail of the dataset curator (optional)' field contains 'otevrenadata@mvcz.cz'. The 'Link to the dataset documentation (optional)' field is empty. The 'EuroVoc concepts (optional)' field contains the tag 'elektronická správa'.

1 Dataset 2 Distribution 3 Summary

Dataset title  
Přístupy k agendám CzechPOINT - 2009

Dataset description  
Záznamy o jednotlivých výpisech z míst CzechPOINT. Pro každý výpis záznam obsahuje: číslo, datum, čas a typ výpisu a zkratku, název a typ místa CzechPOINT, na kterém byl výpis vydán. U advokátů, notářů a exekutorů byly údaje anonymizovány.

Dataset theme  
Government and public sector

Accrual periodicity  
never

Keywords  
výpis agendy CzechPOINT

Spatial coverage - RÚIAN territorial unit type  
State territory

Spatial coverage - RÚIAN territorial unit code  
Česká republika (1)

Temporal coverage from (optional)  
2009-12-31

Temporal coverage to (optional)  
2009-01-01

Name of the dataset curator (optional)

E-mail of the dataset curator (optional)  
otevrenadata@mvcz.cz

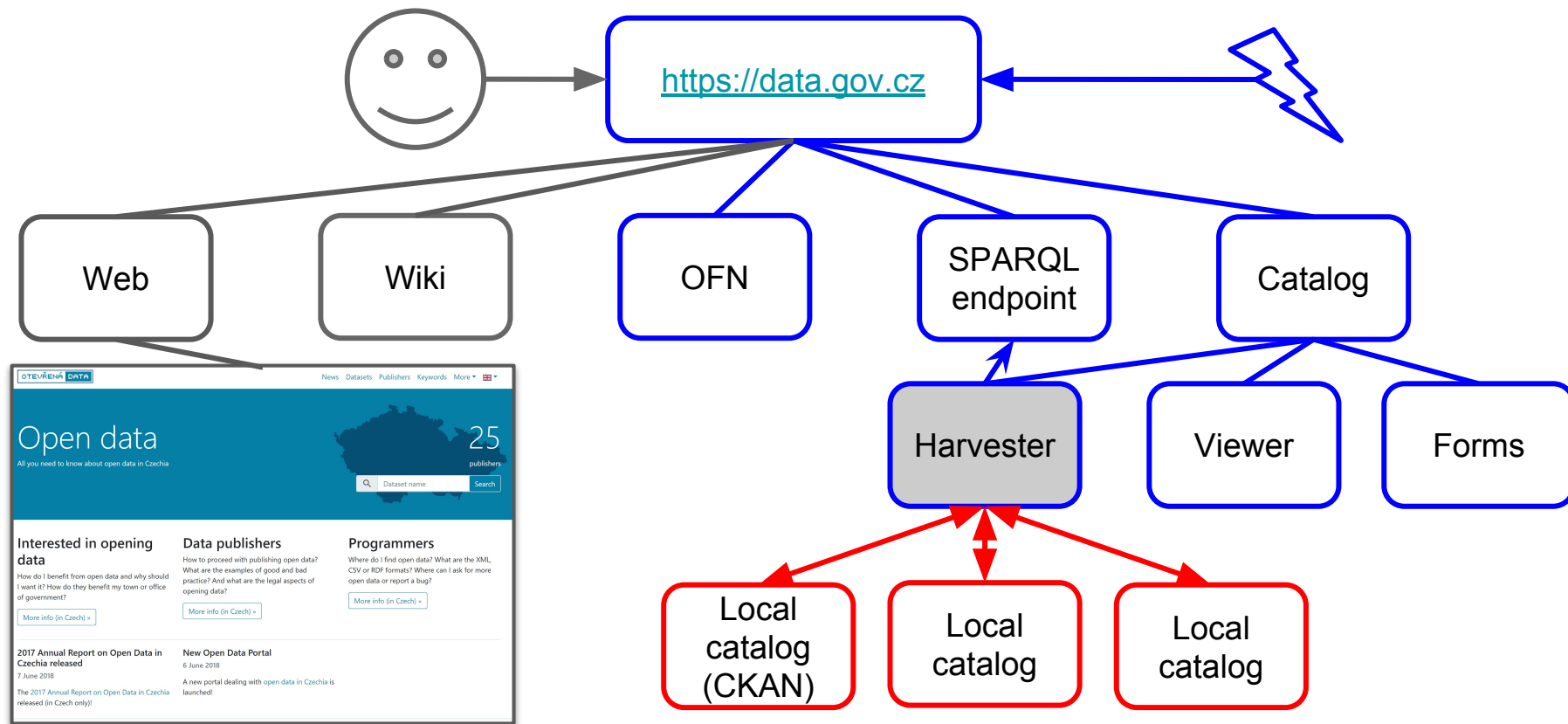
Link to the dataset documentation (optional)

EuroVoc concepts (optional)  
elektronická správa

# Czech Open Data Portal (ODP) architecture and LOD



European Union  
European Social Fund  
Operational Programme Employment



# National Open Data Catalog

## Local Catalog API Standard (DCAT-AP)



European Union  
European Social Fund  
Operational Programme Employment

### OFN “[Definition of data catalog API](#)”

- Generic data structure definition
  - Based on DCAT-AP v1.2
- 3 API kinds supported
  - (extended) CKAN API
    - JSON based
    - Mapping of JSON to DCAT-AP
  - DCAT-AP v1.2 static files
    - Dereferencable Catalog IRI
    - Dereferencable Datasets IRI
  - DCAT-AP v1.2 SPARQL endpoint

Pracovní verze



Evropská unie  
Evropský sociální fond  
Operační program Zaměstnanost

ReSpec

## Rozhraní katalogů otevřených dat

Draft 21. listopadu 2018

**Tato verze:**  
<https://data.gov.cz/otevrene-formalni-normy/rozhrani-katalogu-otevrenych-dat/draft/>

**Nejnovější verze:**  
<https://data.gov.cz/otevrene-formalni-normy/rozhrani-katalogu-otevrenych-dat/>

**Předchozí verze:**  
<https://opendata.gov.cz/nastroj:ckan-api>

**Editor:**  
Jakub Klímeček (Ministerstvo vnitra)

**Zúčastněte se:**  
[GitHub opendata-mvcr/otevrene-formalni-normy](#)  
[Nahlaste bug](#)

**Odkazy:**  
[Portál otevřených dat](#)

---

### Abstrakt

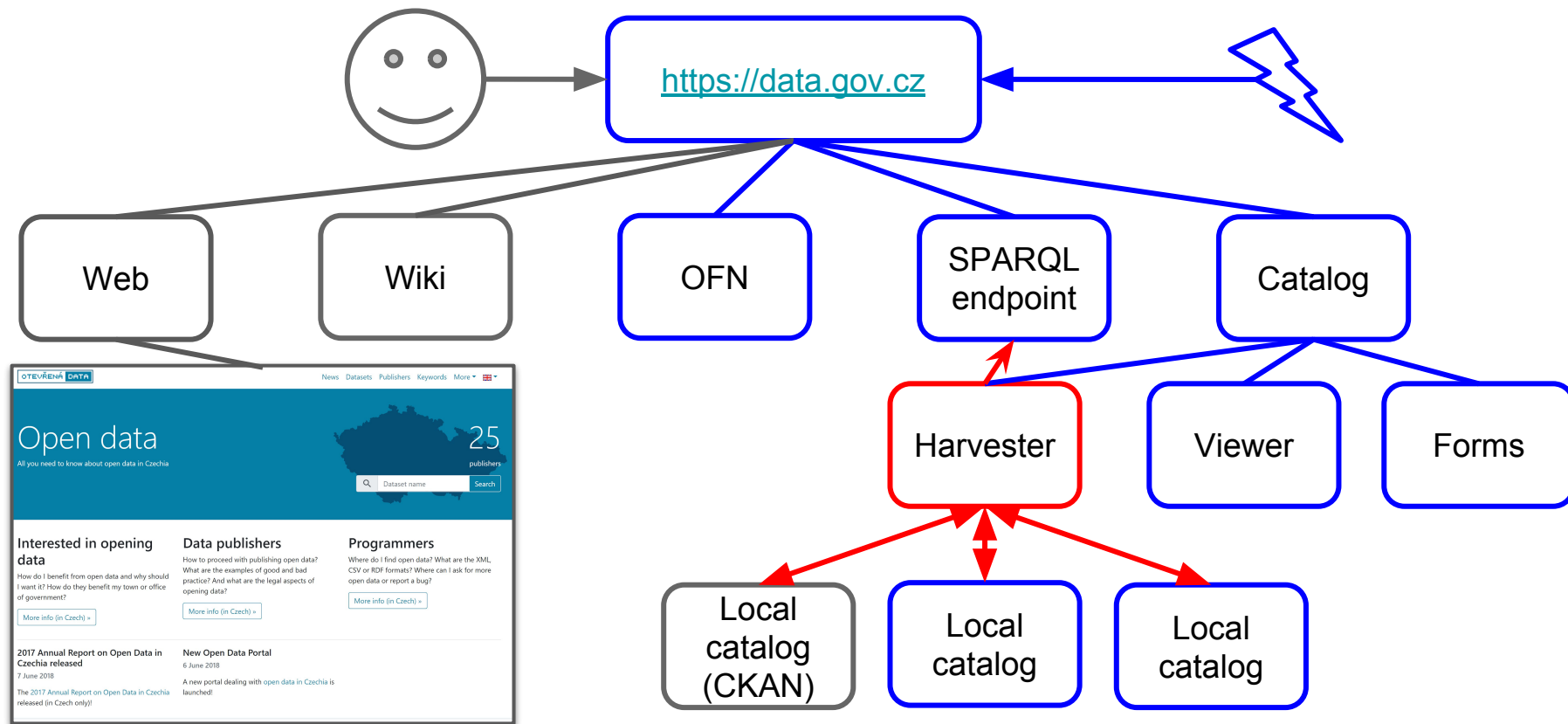
Cílem tohoto dokumentu je specifikovat rozhraní lokálních katalogů otevřených dat (LKOD) pro jejich harvestování Národním katalogem otevřených dat (NKOD). Stejná specifikace platí i pro datový model NKOD a data poskytovaná z NKOD v podobě souborů ke stažení či SPARQL endpointu. Nejprve je specifikován datový model katalogizačního záznamu. Poté jsou specifikovány 3 varianty rozhraní pro harvestaci. Varianta CKAN API



# Czech Open Data Portal (ODP) architecture and LOD



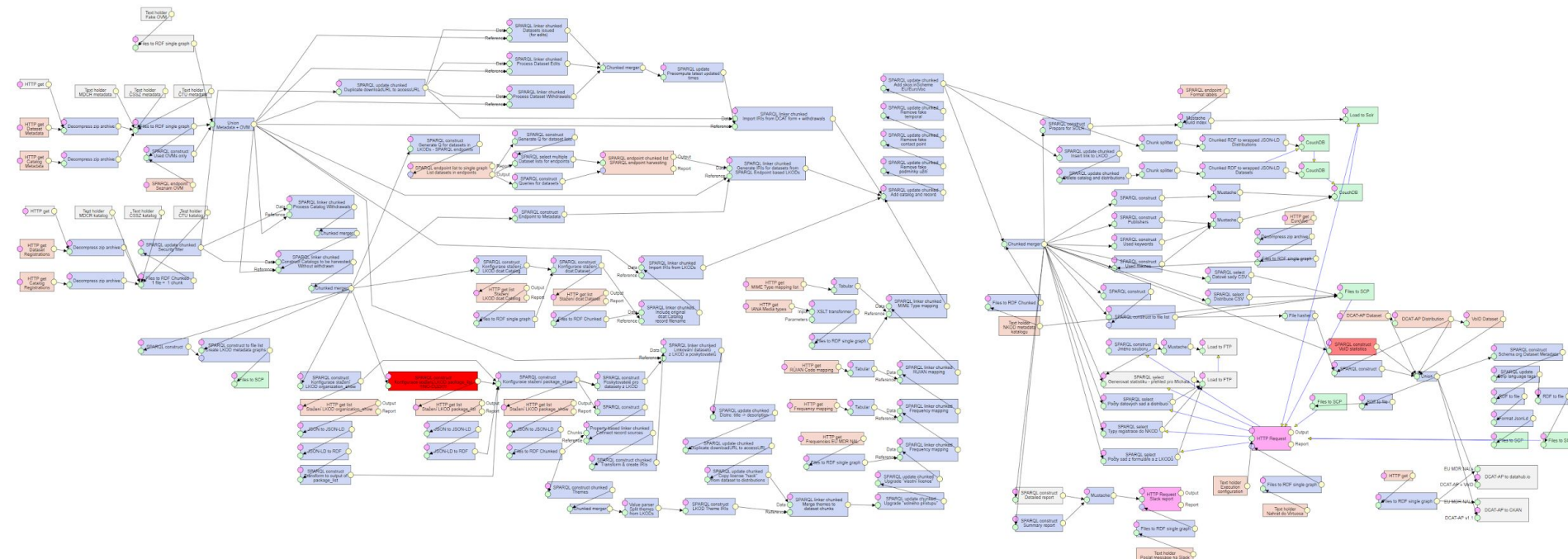
European Union  
European Social Fund  
Operational Programme Employment



# National Open Data Catalog LinkedPipes ETL - Overview



European Union  
European Social Fund  
Operational Programme Employment

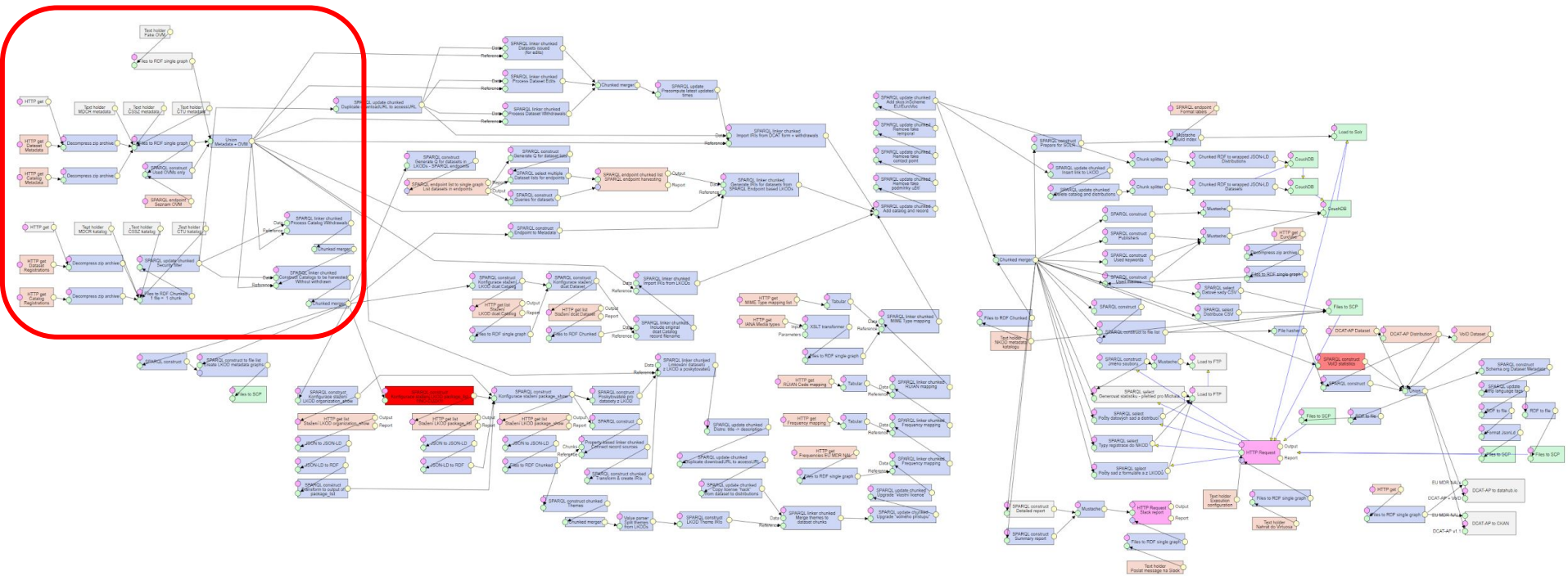


# National Open Data Catalog

## LP-ETL - Data input



European Union  
European Social Fund  
Operational Programme Employment

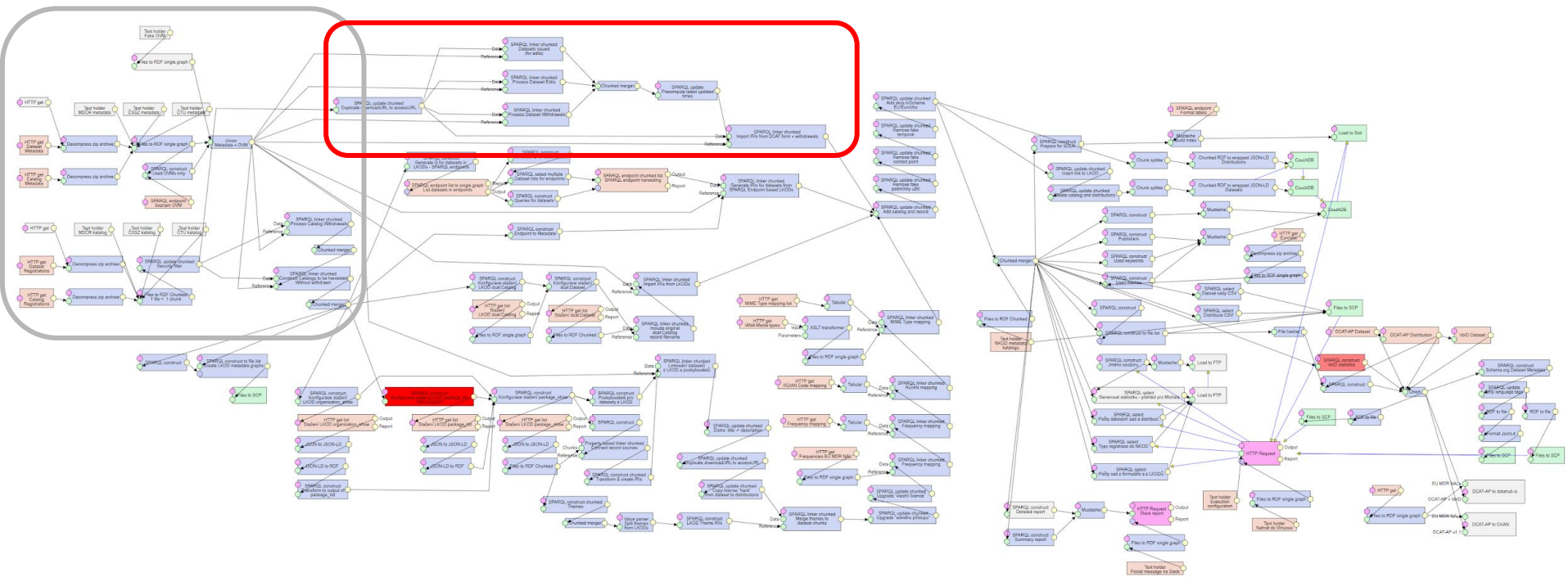


# National Open Data Catalog

## LP-ETL - input forms processing



European Union  
European Social Fund  
Operational Programme Employment



# LP-ETL - DCAT-AP SPARQL endpoint local catalogs



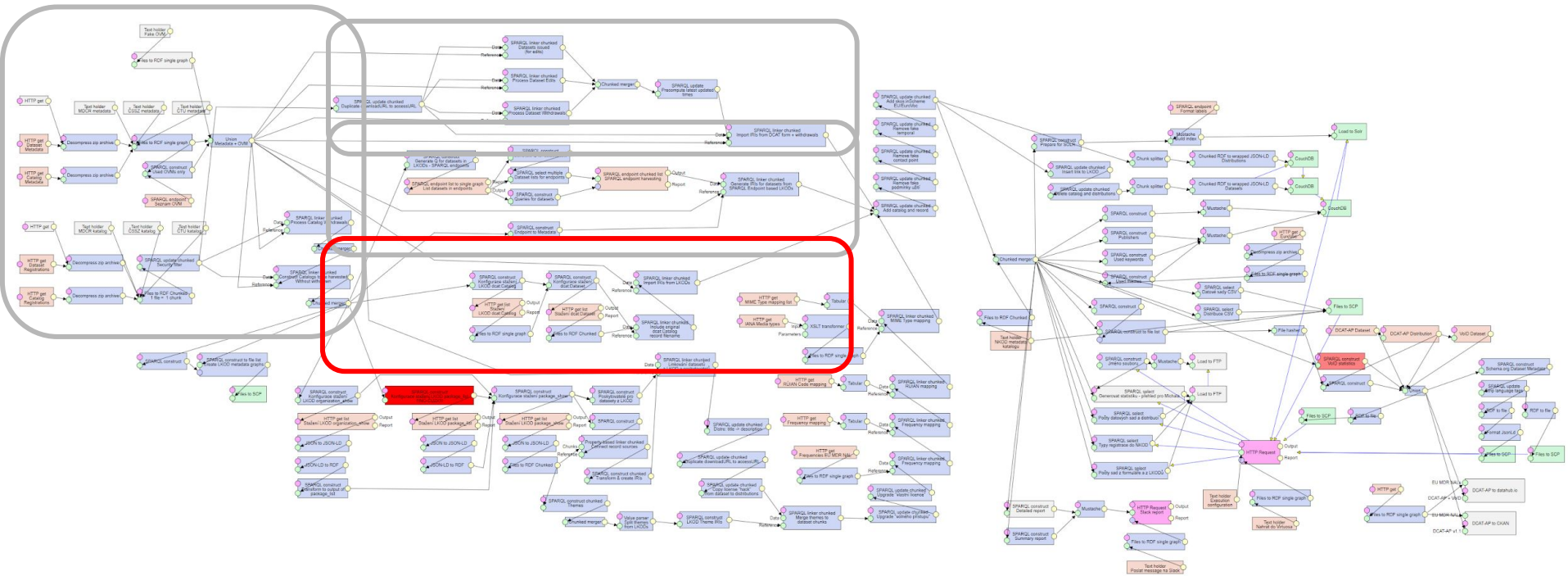


# National Open Data Catalog

## LP-ETL - DCAT-AP static files local catalogs



European Union  
European Social Fund  
Operational Programme Employment

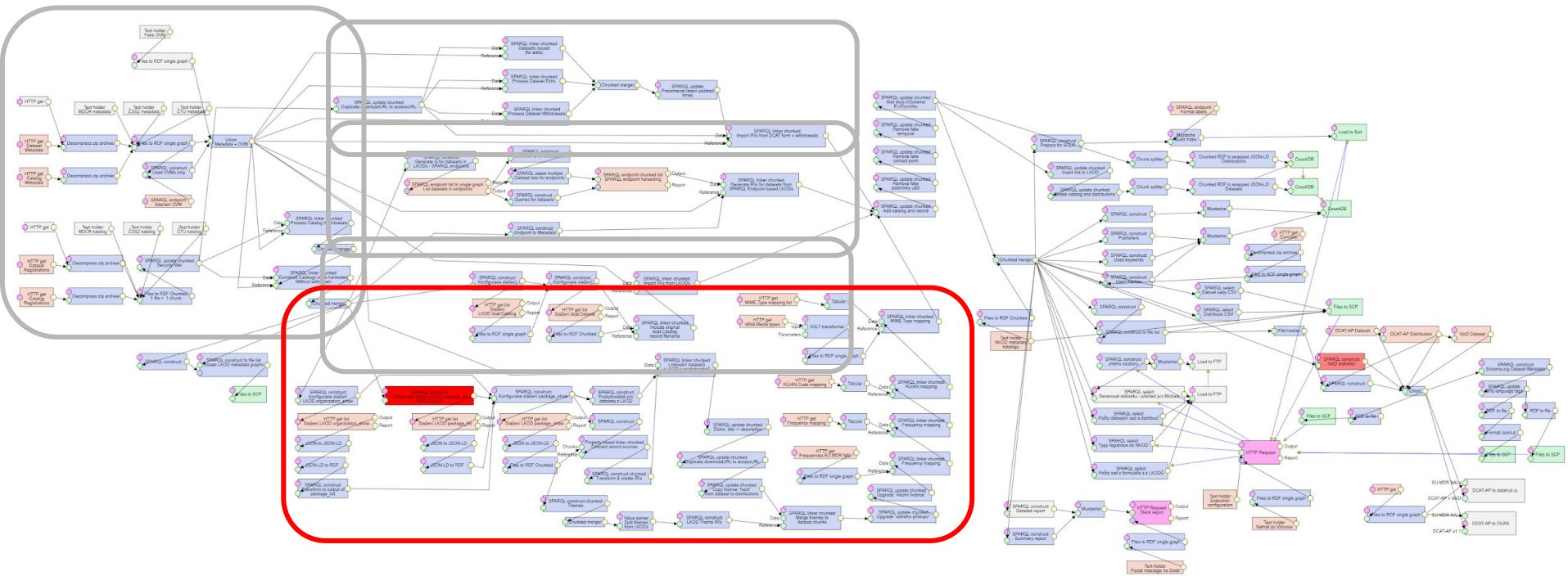


# National Open Data Catalog

## LP-ETL - CKAN APIs local catalogs



European Union  
European Social Fund  
Operational Programme Employment

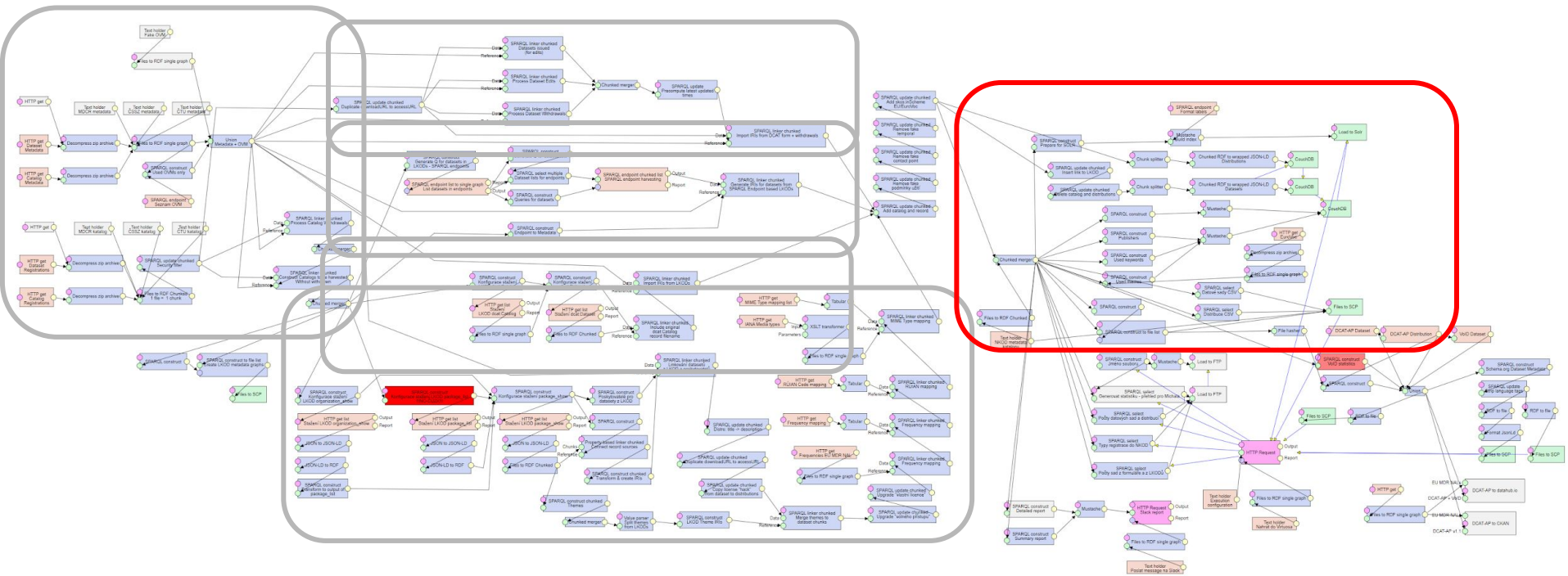


# National Open Data Catalog

## LP-ETL - Data output - dumps (RDF, CSV), endpoint



European Union  
European Social Fund  
Operational Programme Employment



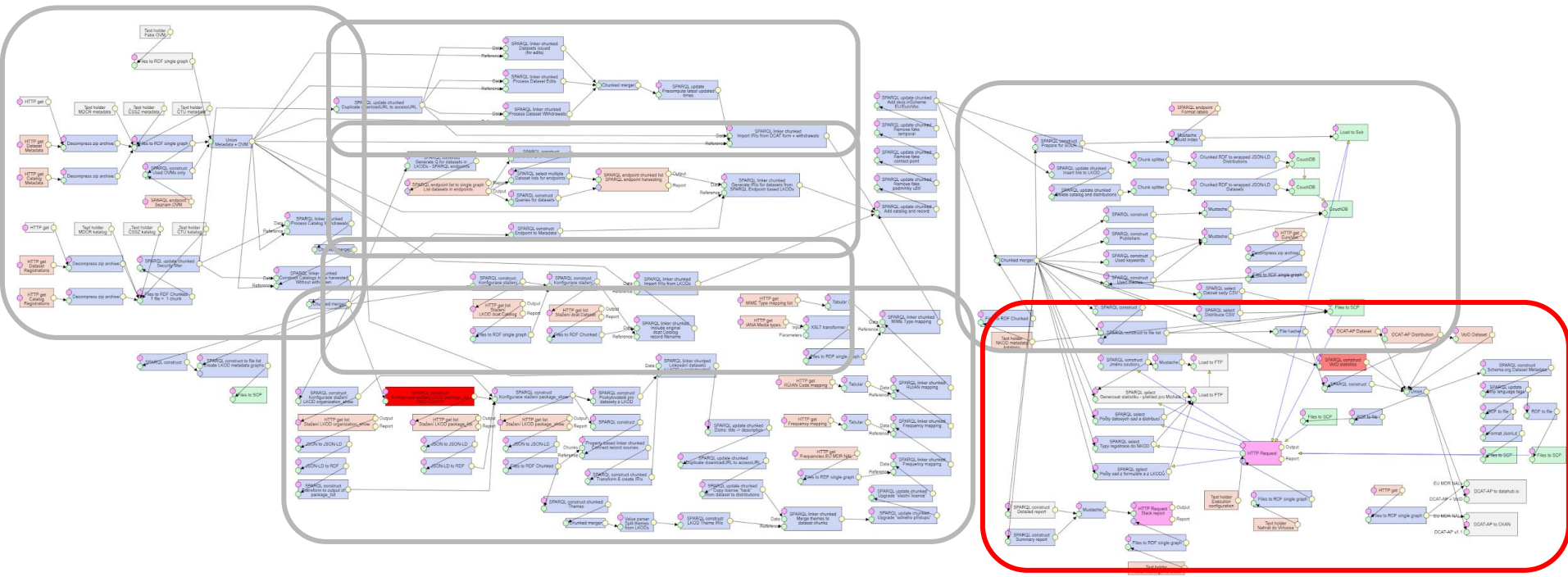


# National Open Data Catalog

## LP-ETL - Metadata



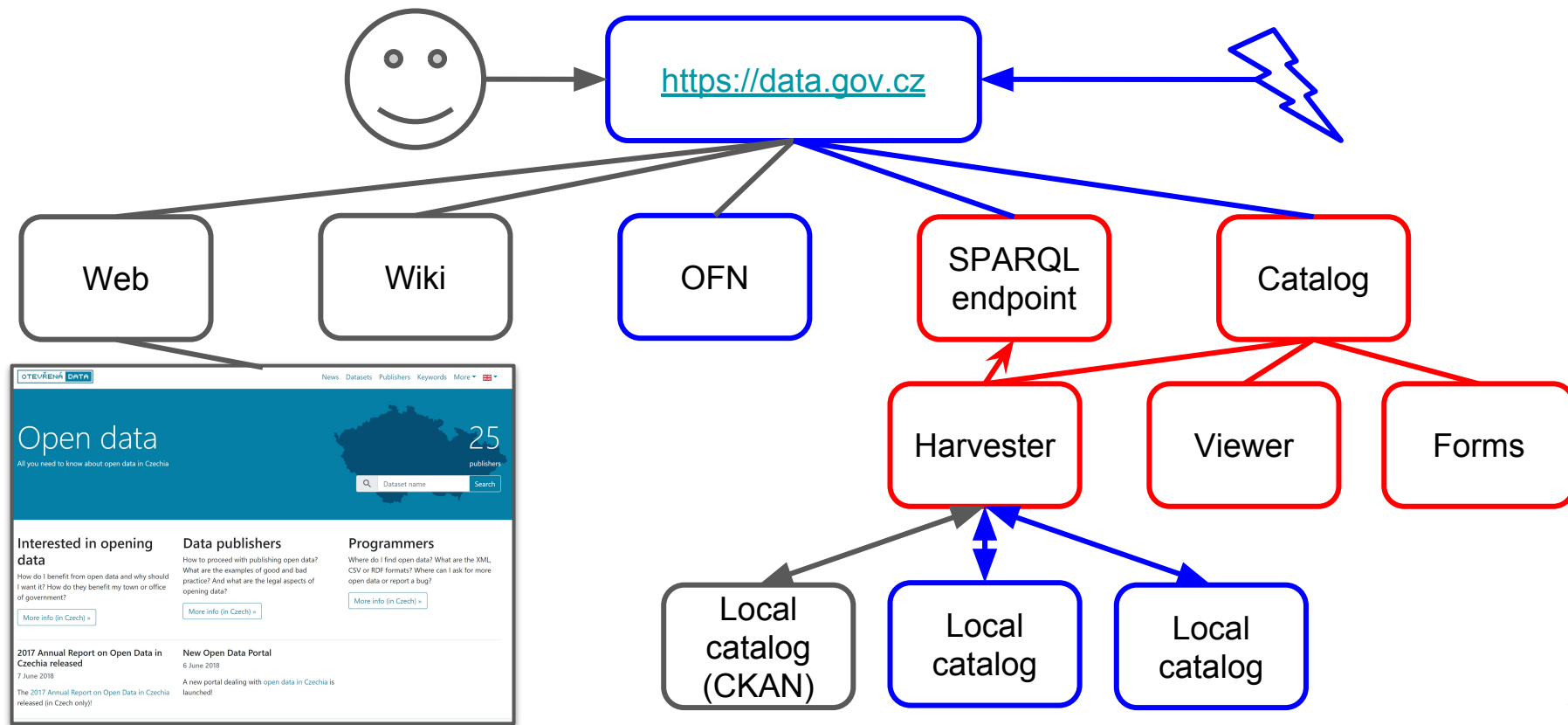
European Union  
European Social Fund  
Operational Programme Employment

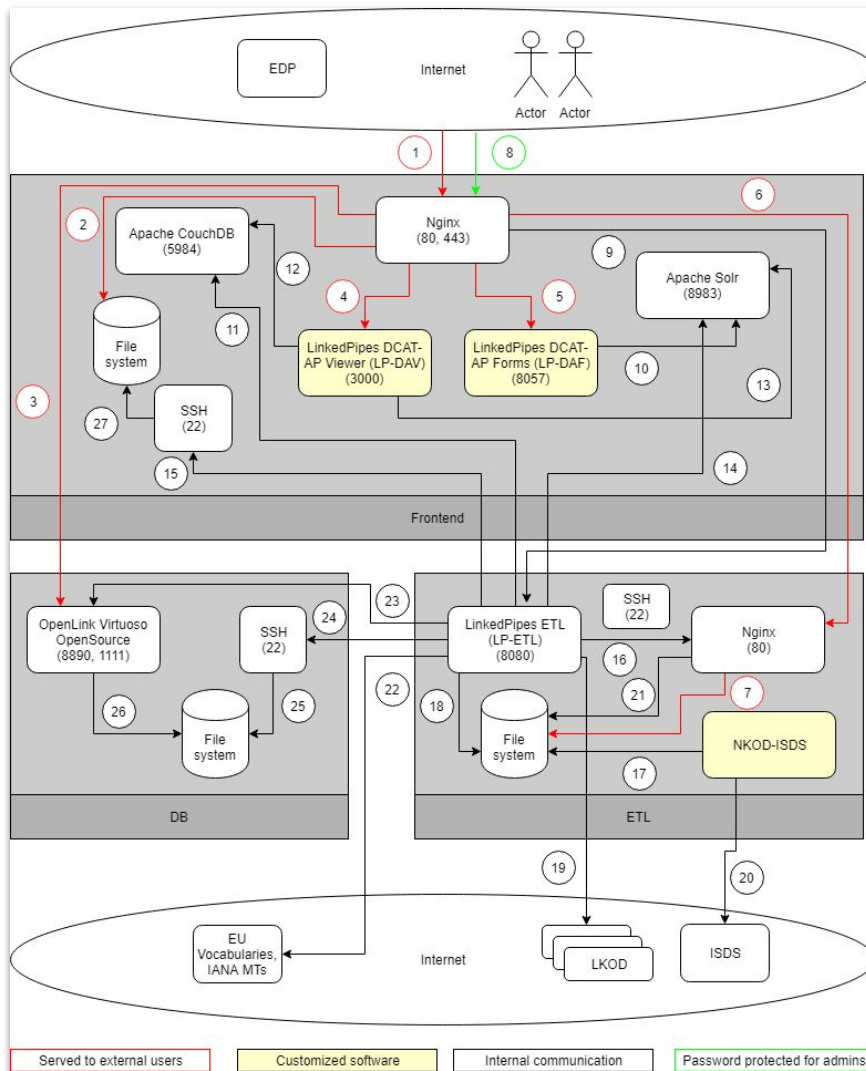


# Czech Open Data Portal (ODP) architecture and LOD



European Union  
European Social Fund  
Operational Programme Employment





# National Open Data Catalog Deployment and documentation

- Open-Source, on [GitHub](#)
- All software used for the catalog is open-source

# Czech Open Data Portal (ODP) architecture and LOD



European Union  
European Social Fund  
Operational Programme Employment

