DIGITÁLNÍ A INFORMAČNÍ AGENTURA_

Export z Národní architektury eGovernmentu ČR

Obsah

Public Data Fund

Description of the Public Data Fund

The Public Data Fund (also referred to as "VDF") is the principle of creating and completing the image of the Interconnected data fund according to open data principles in order to support the sharing of data by OVS in the exercise of their public administration and beyond the scope of their rights and obligations captured in the RPP and outside the PPDF.

The VDF encapsulates and represents the actual physical open data sources scattered in the VS web space. The actual representation and delineation is implemented by tools designed to manage open data metadata, which not only concretizes the VDF concept itself, but also offers a means for further categorization across the entire CR data fund.

The importance and role of VDF is manifested in two areas of VS.

The first relates to the internal VS regime, where it is the main means of sharing public data between public bodies with each other, and also extends the mutual sharing of information to those data that are not in the PPDF at all. The importance of the VDF in this area is best captured by the wording of sub-objective 5.10 of Information Concept of the CR.

"*The Public Data Fund, consisting of published public data of public administration, is the basic method for sharing public information between public entities among themselves and for sharing public data between the public and private spheres in the Czech Republic. The Public Data Fund will move from the mere publication of machine-readable Open Data to the publication of legally binding, valid and regularly updated datasets, with clearly defined responsibility of the OSC for such datasets.*"

The second important area in which the VDF plays an important role is the area of moving the country towards open government and progressively making VS information available to the public. The VDF is made up of published VS information in the form of open data and thus becomes part of all publicly available open data in the country.

==== Data resources of the Czech Republic.

A data resource is a way of accessing data. The data of the CRS is accessible through data sources in the following categories:

PPDF - a interconnected data fund is described by here.

OD - open data - consists of all published VS data in open data format, which must meet the requirements for open data corresponding to the definition given in § 3(11) of Act No.106/1999 Coll..

For data in open data format, it is understood that they are:

- Freely available on the web as downloadable data files in machine-readable and open formats - CSV, XML, JSON, RDF and other open specification formats.
- Provided with terms of use that do not restrict their use.
- Listed in the National Catalogue of Open Data (NKOD) as datasets with direct links to the data files that make them up.
- Provided with complete, freely available and open documentation.
- Provided with contact details of the curator for feedback (errors, extension requests, etc.)
- Are published according to open formal standards within the meaning of § 4b(1) of Act No. 106/1999 Coll. on free access to information

VDF - public data fund - is a subset of open data. It consists of data sources that make publicly available data available in the PPDF and data sources that make publicly available data that the PPDF does not contain at all, in open data format. These are e.g. data necessary for the correct interpretation of PPDF data (e.g. codebooks) or data contained in agencies that are not legally obliged to publish their data to other agencies and are therefore not part of PPDF. The VDF is thus used to make PPDF data available to individual OSSs in those cases where the respective OSSs do not have the possibility to share this data via PPDF (they do not have rights assigned in RPP), or such data that are not part of PPDF at all and in both cases need such data for their work. All published data in the VDF must be in open data format and must meet the requirements for open data according to law no.106/1999 Coll. and the requirements of the legislation on the processing and protection of personal data. Beyond these rules, the provider must guarantee the timeliness, availability and accuracy of its data published in the VDF so that other OSSs can use it in their work.

The VDF also includes data with controlled access. These are datasets that fulfil all the conditions set out for open data, with the exception of the requirement that their provider does not restrict their subsequent use in any way. An example of such a practice is making the provision of data conditional on prior registration so that the provider can monitor how the data is handled. This is the case where the full opening of the data is prevented by objective reasons arising, for example, from legislation or the technical nature of the dataset in question.

Although controlled access data are not open data, they are considered as open data in the context of the VDF and must therefore fulfil all the required conditions for open data, including cataloguing in NKOD, except for the modified conditions of use. For the modification, the terms of use of the data must state and justify any restrictions imposed, specify the conditions for obtaining consent for use, and provide contact details of the publishing organisation for requesting access to the published datasets.

The granting of physical access to such datasets and the management and registration of authorised users is entirely the responsibility of the publishing organisation.

Access to VS data

In the performance of public administration, the OVM (or the SPUU) use data held in the basic registers system and data held on individual entities in other agency systems if they are obliged by law to publish their data to other agencies. These data constitute the PPDF and individual OVMs have access to them through the reference interface of public administration on the basis of their rights and obligations captured in the RPP.

The OVM may require data for its work beyond the permissions defined in RPP, or it may need data not contained in the PPDF. In such cases, it may access publicly available data via the VDF and use a suitable published open dataset via NKOD to obtain the necessary data.

In the absence of such a dataset, it shall ask the notifier of the corresponding agenda to add the open data to the VDF.

The public has open access to all open data and thus also to the data in the VDF via NKOD and the Open Data Portal (POD).

Basic characteristics of the VDF

The VDF contains data sources that serve as a supplementary data source for OVMs in such cases where OVMs do not have rights to use such data directly from PPDF in RPP, or do not contain such data in PPDF at all.

Ultimately, this extends VS data sharing beyond that provided by the reference interface in the interconnected data fund (see objective 5.10 of the Information Concept). An important and valuable effect of the VDF is also to

make it available to the public in an open data format within the limits and conditions defined in the laws for individual agencies.

The data in the VDF must be provided in open data format and thus meet all the requirements for open data according to Act No.106/1999 Coll. The VDF is the basic means and method for sharing public information between public bodies and for sharing public data between the public and private spheres.

If the OSC uses data from the VDF, it considers it to be correct and does not need to verify its accuracy.

The provider of data to the VDF must therefore:

- guarantee the accuracy of the published data,
- guarantee the quality of the published data,
- guarantee the timeliness of the published data,
- guarantee the regularity of updating the published data,
- ensure automatic notification of all changes to registered candidates using the POD functionality.

VDF Tools

The VDF, created on the basis of open data principles, creates a virtual distributed data space in which VCSs share public data registered in their managed ISVS with the public and with each other. The management of such a virtual environment and the maintenance of the quality, consistency and timeliness of published data is based on the management of the metadata of published datasets, the description of data sources and the semantic mapping of the elements of dataset structures to the semantic vocabulary of VS concepts using tools designed for this purpose.

The management of the VDF data will basically follow the same way as open VS data is already managed, with a few additional extensions. The basis will be the NKOD data catalogue, which will be modified and extended to distinguish between data categories, allow user registration and provide notification of changes.

An overview of all the tools used in working with VDF and open data is given in the following figure.

POD represents the tools and functional areas of NKOD required to catalog and manage all categories of data in PPDF.

Basic functions provided:

- open data cataloging (VDF and OD),
- automatic notification of changes in published datasets.

Basic POD logical units:

- NKOD - modified and extended to manage information about individual data categories and their inclusion in VDF.
  - NKOD is a tool in which individual VDCs catalogue the open data they publish and other VDCs and the public search and access the open data.
  - The existence of the NKOD and the obligation of individual OSSs to catalogue their open data in it is determined by Act No. 106/1999 Coll., on free access to information, which establishes the definition of open data and establishes the existence of the NKOD.
- Catalogue of Open Data Users (KUOD) - will record which OSS use which open data in the VDF.
- Catalog (registry) of data users requesting notification of changes to published datasets - a list of users with requests to send notification of changes for specific published datasets.
- Notification HUB - a tool that provides automatic distribution of notifications of changes in published datasets when a change is announced by the publishing OVS using KUOD with a notification request.

When registering a dataset, it passes the URI for sending change notifications to the publishing OVS.
- Register of requested VDF datasets - list of requests to add missing PPDF images to the VDF.

Semantic dictionary of terms - a dictionary of public administration terms, as a tool for harmonization of semantics of open data, will be created on the basis of lists of data maintained for agendas in the RPP according to Article 51(5)(g), (h) and (i) of Act No. 111/2009 Coll. On the basic registers and will gradually elaborate them into reusable and shared information models (ontologies), which will be linked to the data maintained for the agendas and also linked to the vocabularies and ontologies emerging from the EU initiative (e.g. ISA Core Vocabularies). The logical data schemas describing their machine (syntactic) expression at the logical level will be linked to the concepts of the semantic vocabulary of concepts, thus realising the linking of the semantics (meaning) of data across datasets and their sources. The semantic vocabulary of terms will be an integral part of the VDF and will become the basis for describing the semantics of published data and their interconnection.

The **Semantic Dictionary of Terms Management Tool (NSSSP)** - will enable the management of the semantic dictionary of terms by all VDFs and the description of their data for semantic harmonisation.

Portal of Open Data (POD) - gateway to the world of open VS data. Includes NKOD, educational information, publication standards, templates, recommended practices, related documents and information on open data published in the Open Data Publication and Cataloguing Standards

Local Catalogue of Open Data - LKOD - optionally implemented for the open data cataloguing needs of a specific data provider. The LKOD regularly automatically retrieves information about published datasets from individual local catalogues.

Application Catalogue over Open Data - a list of all services (applications) using published datasets with search options based on the datasets used or based on the life situations of the data consumer.

Content distribution of VDF data sources

VDF is composed of two types of data sources:

- data sources containing the public image of the PPDF,
- data sources extending the PPDF image with data that are not in the PPDF.

By the public image of a data source we mean the data from that source transformed into a machine-readable form that can be published. In the case that a data source can be published in its entirety in its original form, this transformation is an identity. That is, the data source is in this case a publicly accessible image of itself. For other data sources, the transformation preserves the details of the original data as much as possible and thus typically performs only the necessary anonymisation. This usually means either data projection (i.e. removing certain types of data from the dataset) or statistical aggregation.

The following simple general rules should apply to data sources producing a publicly available PPDF image.

- For every data source in PPDF, there must be an image in VDF, namely:
  - for data sources with legislated remote access, the subject of publication will always be the primary data,
  - for data sources without legislated remote access, primary data will be published if nothing prevents it, otherwise publication will be in the form of statistics and aggregations at the finest granularity.

List of data sources constituting the content of the VDF

- Primary registry images - statistics from the registry will be published as the registry image
  - ROS
  - RÚIAN
  - RPP
  - ROB
- Images of public registers (public registers of legal and natural persons according to Act No. 304/2013 Coll. (Act on Public Registers of Legal and Natural Persons.).
  - Federal Register,
  - foundation register,
  - register of institutes,
  - register of the association of unit owners,
  - Commercial Register,
  - register of public benefit societies.
- Images of data registered in the RPP pursuant to Section 51(5)(h) of Act No. 111/2009 Coll. on the Basic Registers.
- Images of all registers for which the law defines their remote access, if there are no legal obstacles or significant risks (in the case of registers without remote access defined in the law - publication of aggregations or statistics).
- Published (and local) codebooks, their semantic linking, possible consolidation and indication of contexts capturing their use.
- Published statistics and other data used by OVS and suitable for public access.

Simplified concept for publishing data in the VDF

The aim of publishing VS data in the VDF is to create a harmonised and consistent picture of the PPDF in the VDF, semantically linked also to the corresponding concepts of the legislation with the help of the Semantic Glossary of Terms, including captured inter-relationships. To ensure consistency of the VDF, and to produce an equivalent PPDF picture, it is therefore necessary that the published datasets of data sources in the PPDF are also published at the level of linked data. This is achieved by publishing specific dataset distributions from PPDF with a machine-readable link to a semantic vocabulary of terms.

- The basis of the publication to VDF is therefore a semantic dictionary of terms created on the basis of the data in the RPP, and which describes the PPDF and therefore future images in the VDF.
- The actual publication will be implemented by distributing the datasets using a format that allows for recording the relationships in the data and the links to the semantic dictionary of terms in a machine-readable form.

The published images of PPDF data sources in the VDF will create an infrastructure for all open VS data, which in practice means that publishing organizations will not have to publish data that are already published when publishing new datasets, but will only publish new extension information for individual VDF entities. However, their obligations will include the obligation to add links of the newly published data to existing entities in the VDF.

A published dataset from a PPDF data source must become part of the VDF in order to become part of the VDF:

- must meet all the conditions imposed on open data within the meaning of Act No. 106/1999 Coll. on free access to information,
- it must be published according to the prescribed open formal standards within the meaning of Section 4b(1) of Act No. 106/1999 Coll. on free access to information
- must have a context added that maps the structural elements of the dataset to a semantic dictionary of terms.

In addition, the following rules apply:

- In the case of duplication of published datasets in the VDF, or the addition of data by different publishers to the same published entity, the obligation to add links to already published data always falls on the

publisher who publishes the additional data.
- For each dataset published in the VDF, information on the [notification](#) mechanism of changes according to the corresponding OFN listed on the POD shall be provided.
- In order to ensure the legal validity of the published data, their quality, timeliness and prompt updating must be guaranteed through organisational and procedural support in the data provider's organisation.

Data Management of the Czech Government

The involvement of the VDF in the performance of public administration requires an assured guarantee of the quality of published datasets, which means ensuring the publication of truly legally binding, valid and regularly updated datasets with a clearly defined responsibility of the OVS for such datasets. It is clear that such a state of affairs will be difficult to achieve with the current approaches to open data publication in the PS, where publication is not very coordinated, often in non-standard ways and with the absence of the use of recommended standards.

In order to ensure a quality public data fund, it will be necessary to proceed to a planned and well-defined creation of VDFs and also to change the current paradigm of VS data management - to separate VS data from applications and move them to the centre of informatics management in accordance with the position of data in legislation. Governance and management of VS data seems to be necessary in the future also in the light of proclamations that data is the most valuable thing that organisations, and therefore public administration and ultimately the public, have.

To ensure the quality of published datasets, the following will be released:

- "Rules and procedures for publishing data in the VDF",
- "Rules and procedures for the publication of open data",
- "Data Policy of the SCS".

In addition, to ensure the ease of use of published datasets, existing and new OFNs (Open Formal Forms) will be continuously developed, defining conceptual and technical patterns for the actual publication of data. The OFNs are and will be published on the Open Data Portal and are binding for publishing organisations according to Act No. 106/1999 Coll.

Public Data Views

Public Data fund Rules

The VDF is made up of open data provided by individual VCSs for sharing with other VCSs and is a subset of all open VS data. Therefore, it is necessary to ensure that all functional units of each VS architecture respect and adhere to the rules applicable to:

- open data (supported by legislation under Act No. 106/1999 Coll.),
- VDF data - rules for open data, in addition to other specifics derived from the characteristics of VDF.

Open data

Open data are:

- Freely available on the web as downloadable data files in machine-readable and open formats - CSV, XML, JSON, RDF (JSON-LD, Turtle, ...) and other formats with open specifications.
- Provided with terms of use that do not restrict their use.
- Registered in the National Catalogue of Open Data (NKOD) as datasets with direct links to the data files that make them up.
- Provided with full documentation.
- Provided with contact details of the curator for feedback (errors, extension requests, etc.).
- Are published according to open formal standards within the meaning of Section 4b(1) of Act No. 106/1999 Coll. on free access to information.

Use of open data in public administration

There are no requirements on how to use the published open data, the data can be freely imported into appropriate applications and information systems. Published data is made available through NKOD (POD) and can be obtained as datasets. The data can also be accessed using third party applications that use the relevant data.

Open Data Publications

Preparing an ISVS for exporting open data

Providing access to IS data is essential. The IS in operation must therefore:

- allow access to the database or
- be able to download optionally structured data (tables) from the reporting module of the system, or
- offer an API from which complete data in the form of data files can be retrieved on a regular basis.

Suitable open formats:

- tabular data - CSV, XML, JSON, RDF (JSON-LD, Turtle, ...),
- hierarchical data - XML, JSON, RDF (JSON-LD, Turtle, ...),
- graph data - RDF (JSON-LD, Turtle, ...),
- geodata (spatial data) - GeoJSON, ESRI Shapefile, OGC GML, OGC GeoPackage.

The data structure must be documented with a human readable document, but also with a machine readable schema. Recommended languages for defining schemas Schemas:

- CSV - CSV on the Web schema,
- XML - XML Schema,
- JSON - JSON Schema,
- RDF - RDFS, OWL, SHACL

If the above points are followed, the complete data obtained from the IS is ready for publication in the form of open data.

Exporting data or API in proprietary format

If this is an extension of an existing IS that does not allow data export or does not offer an API in a machine-readable and open format and such modification cannot be done in the IS, the existing export or API already offered by the system (e.g. to MS Excel) is used and this output is further processed into an open format using other tools to achieve the same state as in the case of direct export to an open format.

Final preparation of data for publication as open data

Data extracted from the IS in one of the ways described above should then be published as open data. This means at a minimum:

1. In the case of an API, ensure that it is mined to obtain complete data for publication (i.e., any direct publication of the API does not satisfy the open data conditions)
2. Ensure regular updates of the extracted data (depending on the nature of the data, this may be at a frequency of e.g. daily, monthly or annually)
3. Publish the collected data on the web for download and subsequently publish each update
4. Provide documentation, terms of use and contact details for the curator
5. Catalogue them in the National Catalogue of Open Data (NKOD)

This can be done using open data preparation, publication and cataloguing tools such as LinkedPipes ETL. The publication of open data should be ensured conceptually at the level of the whole organisation. Full procedures are available on POD, including the required standards.

VDF Data ====

Open data in VDF

Open data is provided in the VDF for use, among other things, by other OSS in the exercise of public administration beyond the scope of their rights and obligations captured in the RPP. Beyond open data, the following applies to open data in the VDF:

- If an OSP uses data from the VDF, it considers it to be correct and does not need to verify its correctness.
- The provider of the data in the VDF guarantees the accuracy, quality, timeliness and regular updating of the data published in the VDF.
- The VDF data provider ensures automatic notification of all changes in the data published in the VDF to registered stakeholders using the Open Data Portal functionality.

Use of open data from the VDF in public administration

In the performance of public administration, the VDF may need data from other VDFs to which it does not have access within the scope set out in the RPP. If it is public data, it accesses it through the VDF. If the data is available in the VDF, no other means of data access and sharing is allowed. Typically, VDF data is used in the following way:

- manually searching for the required datasets in the NKOD and finding links to download the data,
- setting up scripts to periodically import the found datasets into the IS itself from the found links,
- importing datasets into the VS IS,
- registration in the Notification Hub for regular and machine retrieval of updates,
- setting up scripts to import changes obtained from the Notification Hub into the VS IS.

Publishing open data to VDF

For publishing open data to VDF, the same rules apply as for publishing open data mentioned above. In addition, the following rules must be observed:

- The published data is described by a semantic dictionary of terms that is created based on the data in the RPP. The description of the data with a semantic dictionary of terms is created and published according to the open formal standard "Description of data with a semantic dictionary of terms". The semantic

dictionary of terms is created and published according to the open formal standard 'Semantic dictionary of terms'.

- IRIs are used to identify the entities about which data are published in the VDF according to the open formal standard 'Linked Data'.
- No duplication with data already published in the VDF is published in the published data. In the case that an LDC publishes data on an entity for which another LDC already publishes data in the VDF, the LDC shall only publish new supplementary data on that entity. In case it introduces its own IRI to identify an entity than another OVS, it shall link its own IRI to the original one according to the open formal standard 'Linked Data'.
- Relationships between entities in data from the same provider and different providers are represented according to the open formal standard Linked Data. The data provider in the VDF tries as much as possible to link the entities about which it publishes data to the entities about which other VCSs publish data.

Data mandatorily published in the VDF

The following data are mandatorily published as open data in the VDF:

Provider publishing data in the VDF	Data published	Method of publication
Czech Statistical Office	Numerators introduced by a communication in the Collection of Laws	According to OFN Numerators
Declarant of the agenda within the meaning of Section 48(f) of Act No. 111/2009 Coll., on the Basic Registers		

Common rules for open data and VDF data

Organisational and procedural arrangements for data publication

The involvement of the VDF in the exercise of public administration requires the publication of truly legally binding, valid and regularly updated datasets, with clearly defined responsibilities of the OVS for such datasets. In order to meet these requirements, the publishing organisation must implement appropriate organisational measures, assign staff to appropriate process roles and implement publication process activities into staff job descriptions. As a minimum, the following key process roles are required to be assigned:

Data Opening Coordinator, whose responsibilities and duties include:

- Ensuring synergy and control of the output of all other roles involved in data opening,
- communicating with all staff involved in data release,
- external communication with users of VS open data,
- communication and cooperation with the National Open Data Coordinator,
- communication with the Data Office and the relevant Chief Data Officer (CDO).

Data Curator - a key role for:

- ensuring the quality, accuracy, timeliness and thus legal bindingness of data for a particular agenda,
- publication of datasets in accordance with the applicable legislation of the Czech Republic and the Standards for the Publication and Cataloguing of Open Data of the Czech Republic.

The complete recommendations, appropriate internal document templates, all suggested procedural roles and standard publishing processes are listed on the Open Data Portal.

Privacy

If personal data within the meaning of Act No. 110/2019 Coll., on the processing of personal data and Regulation (EU) No. 2016/679, the General Data Protection Regulation (GDPR), is the subject of the information system's records, this does not mean that open data cannot be published from the system. In these cases, the following recommendations apply.

1. In the case of a public record or register and a specific legal regulation mandates the publication of information, personal data can be published in the form of open data.
2. The protection of personal data can be ensured by Anonymisation or Pseudonymisation. The personal data is removed from the data and, where appropriate, replaced by a meaningless artificial identifier. The data without personal data can then be published as open data. However, depending on the nature of the data, it is necessary to check that the data in combination does not allow the identification of a specific person even after the removal of apparent personal data. This may include combinations such as city, age and gender and similar.
3. Data that cannot or should not be disclosed under the previous point may be disclosed in aggregate form. That is, in the form of statistics. However, in the case of publication of statistics, it is desirable to use the finest possible granularity of data and temporal breakdown.

Legal aspects

The legislative framework of open data in the Czech Republic is its regulation contained in Act No. 106/1999 Coll., on free access to information and in Government Regulation No. 425/2016 Coll., on the list of information published as open data, which sets out the obligation for selected public administration bodies to publish data from specific information systems managed by them in the form of open data. More detailed information on strategic documents, action plans and related Czech and EU regulations is available in an updated form on the Open Data Portal (POD).

Open Data, Open Data, OD, VDF, Open Data Portal, NKOD, thematic area

From:
https://archi.gov.cz./ - **Architektura eGovernmentu ČR**

Permanent link:
https://archi.gov.cz./en:nap:verejny_datovy_fond

Last update: **2021/07/01 10:29**