

# Data management, kvalita

Otevřená propojená data  
Řízení kvality dat, role ve správě dat

INS\_2020\_6. přednáška

## Architektura – doplnění

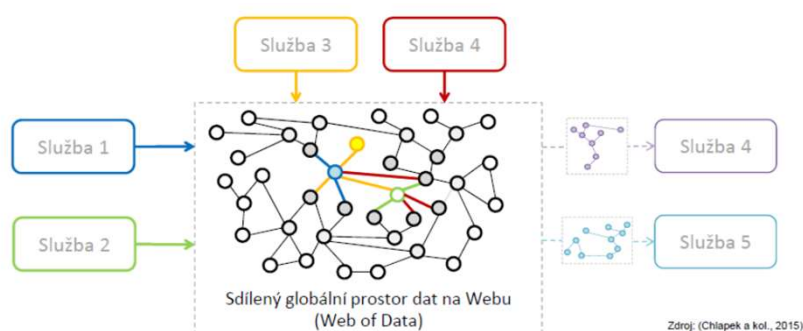
• **Architektura** (dle ISO/IEC 42010:2007, revidována normou ISO/IEC/IEEE 42010:2011) - fundamentální uspořádání systému, které tvoří komponenty a vztahy mezi nimi, včetně vztahu k prostředí, a principy, které řídí jeho návrh a rozvoj.

ISO / IEC / IEEE 42010: 2011 se zabývá tvorbou, analýzou a udržováním architektur systémů pomocí popisů architektury.

- Je vytvořen koncepční model popisu architektury.
- Je zadán požadovaný obsah popisu architektury.
- Jsou zavedena hlediska architektury, architekturní rámce a jazyky popisu architektury pro kodifikaci konvencí a běžných postupů popisu architektury.
- Je zadán požadovaný obsah hledisek architektury, architekturních rámců a jazyků popisu architektury.

## Propojená data – cíle

- Cílem propojených dat (angl. Linked Data) je vytvořit ekosystém webových služeb (aplikací), které publikují, obohacují a využívají data o entitách v jednom globálním sdíleném datovém prostoru (tzv. Web dat).



## Standardy Ministerstva vnitra ČR

- MV ČR garantuje standardy publikace a katalogizace otevřených dat VS ČR.

<http://opendata.gov.cz>

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

Portál otevřených dat

Vše, co potřebujete vědět o otevřených datech v České republice

Národní katalog otevřených dat

2 032 klíčových slov

Název datové sady Hledat

## Poskytovatelé (40)

Český úřad zeměměřický a katastrální (128 642)
Ministerstvo vnitra (3 676)
Úřad průmyslového vlastnictví (1 072)
Český statistický úřad (611)
HLAVNÍ MĚSTO PRAHA (321)
Statutární město Brno (171)
Statutární město Plzeň (164)

## Témata (169)

veřejné finance (137)
parkovací plocha (133)
Zemědělství, rybolov, lesnictví a výživa (84)
Vláda a veřejný sektor (60)
Životní prostředí (38)
veřejná správa (34)
důchodový plán (28)

## Klíčová slova (2032)

vektorová mapa (128 642)
Katastrální parcely (103 473)
Adresy (96 235)
katastrální parcela (90 955)
polygon (90 397)
pozemek (84 712)
hranice (84 699)

## Formáty (28)

CSV (17 328)
XML (4 285)
GNU zip (3 594)
ZIP (1 492)
GeoJSON (154)
Excel XLSX (119)
Esri Shape (110)

## Katalog otevřených dat

Ministerstvo zdravotnictví České republiky

Datové sady

Organizace

Skupiny

O nás

Vyhledat



## 🏠 / Datové sady

## 🔍 Organizace

ÚZIS ČR 36

Ministerstvo zdravo... 6

Hygienická stanice ... 2

Vyhledat datové sady...



65 datových sad nalezeno

Seřadit dle:

Relevance



**Číselník věkových skupin***Tento dataset nemá uveden žádný popis*[CSV](#)**Číselník typů hodnot***Tento dataset nemá uveden žádný popis*[CSV](#)**Číselník pohlaví***Tento dataset nemá uveden žádný popis*[CSV](#)**Číselník ORP***Tento dataset nemá uveden žádný popis*[CSV](#)**Číselník okresů***Tento dataset nemá uveden žádný popis*[CSV](#)**Číselník krajů***Tento dataset nemá uveden žádný popis*[CSV](#)

## Standardy Ministerstva vnitra ČR

Standardy jsou založeny na výše prezentovaných principech a dále je rozšiřují o:

- vzorové publikační plány pro jednotlivé typy orgánů veřejné moci a pro samosprávy
- doporučené postupy publikace a katalogizace otevřených dat
- povinnost katalogizovat otevřená data v Národním katalogu otevřených dat  
(<http://portal.gov.cz/portal/obcan/rejstrik/data/>)
- povinnost publikovat na stupni 3, volitelně na stupni 5
- povinnost dodržovat předpřipravená datová schémata pro doporučené datové sady na stupni 3
- povinnost dodržovat několik základních pravidel
  - <http://opendata.gov.cz/standardy:stupne-otevrenosti>
  - <http://opendata.gov.cz/standardy:technicke-standardy-pro-datove-sady-na-stupni-3>
  - <http://opendata.gov.cz/standardy:technicke-standardy-pro-vytvoreni-datovychschemat-na-stupni-3>
  - <http://opendata.gov.cz/standardy:technicke-standardy-pro-datove-sady-na-stupni-5>
  - <http://opendata.gov.cz/standardy:technicke-standardy-pro-vytvoreni-datovychschemat-na-stupni-5>

# Standardizace publikace a katalogizace otevřených dat <http://opendata.gov.cz>



## Vzorový publikační plán

- Vzorový publikační plán vymezuje doporučené oblasti či doporučené datové sady k publikaci otevřených dat včetně typických rizik a přínosů
- Vzorové publikační plány by měly být vytvořeny pro následující typy orgánů VS:
  - a) Ústřední orgán státní správy – ministerstvo
  - b) Ústřední orgán státní správy – ostatní
  - c) Krajský úřad
  - d) Obce s rozšířenou působností
  - e) Obecní úřad – ostatní
- Vzorový publikační plán by měl zejména obsahovat:
  - Seznam doporučených datových sad k otevření, příp. oblastí, ze kterých datové sady vybírat
  - Detailní popis doporučených datových sad
  - Vzory katalogizačních záznamů
  - Typická rizika spojená s publikací doporučených datových sad a doporučené kroky k jejich zmírnění
  - Typické přínosy publikace doporučených datových sad a doporučené kroky pro jejich dosažení (upozornit u jednotlivých datových sad možnost jak přes publikaci otevřených dat mohou subjekty splnit svou zákonnou povinnost - např. „dálkový přístup“ k určeným datům)
  - Příklad publikovaných otevřených dat u vybraného validátora
- Vzorové publikační plány jsou k dispozici <http://opendata.gov.cz> v části „Vzorové publikační plány“.

## Národní katalog otevřených dat VS ČR (NKOD)

- NKOD je dostupný prostřednictvím Portálu veřejné správy ČR (PVS) v záložce „Otevřená data“
- NKOD obsahuje:
  - katalogizační záznamy o datových sadách zveřejněných jednotlivými OVM včetně odkazů ke stažení dat
  - data ke stažení mohou být uložena:
    - na webu příslušného poskytovatele dat
    - přímo na PVS v Rejstříku datových zdrojů (součást NKOD)
- NKOD umožňuje katalogizaci následujícími způsoby:
  - vyplnění formuláře na PVS a jeho odeslání přes ISDS
  - automatické plnění metadat z lokálních katalogů OVM (po předchozí registraci lokálního katalogu v NKOD)
- Zahájení pilotního provozu 10. dubna 2015
- Plný ostrý provoz 15. května 2015

## Národní katalog otevřených dat VS ČR (NKOD)

**PORTÁL VEŘEJNÉ SPRÁVY** Na úřad přes internet

Informace pro občany České republiky | Informace pro podnikatele a živnostníky | Informace pro cizince sítí v ČR | Služby pro orgány veřejné moci

CzechPOINT@home | Zákony | Seznam datových schránek | Otevřená data | Životní situace | Věstníky | Zveřejněné informace | Užitečné odkazy

Hledat životní situaci, formulář, věstník nebo jinou informaci

Zadejte Hledaný výraz

Rozšířené vyhledávání [+](#)

INFORMACE PRO OBČANY ČESKÉ REPUBLIKY > OTEVŘENÁ DATA > NÁRODNÍ KATALOG OTEVŘENÝCH DAT > DATOVÉ SADY > DUBEN 2015

**Národní katalog otevřených dat - Datové sady**

Sekce **Datové sady** Národního katalogu otevřených dat umožňuje vyhledávat otevřené datové sady zveřejněné jednotlivými orgány veřejné správy v ČR. Zveřejněné datové sady lze procházet po jednotlivých letech a měsících. Pro vyhledávání datové sady využijte výše umístěné vyhledávací pole nebo odkaz "Rozšířené vyhledávání".

Chcete-li zaevidovat v katalogu novou datovou sadu, můžete ji vložit pomocí [webového formuláře \(návod\)](#). Dalšími posty, formát dat a získání dat

**2015**  
duben květen

ID záznamu	Publikující	Název
g.275074277	Český statistický úřad	Výsledky voleb do zastupitelstev obcí 2014
g.275072710	Český statistický úřad	Základní výsledky Sčítání lidu, domů a bytů 2011
g.274433531	Kraj Vysočina	Příspěvkové organizace Kraje Vysočina

## Úprava legislativy

- Návaznost na schválenou novelu zákona o svobodném přístupu k informacím (106/1999 Sb.), novelizován zákonem č. 222/2015 Sb. – účinnost od **10.9.2015**
- V souvislosti se standardy MV ČR existuje novelu zákona o informačních systémech (365/2000 Sb.)
- Novela zavádí pojem otevřených dat, významných datových sad a povinnost publikovat významné datové sady v podobě otevřených dat
- MV ČR upravuje prováděcí vyhláškou kritéria pro významné datové sady, minimální stupeň otevřenosti (3), postup publikace otevřených dat a Národní katalog otevřených dat
- MV ČR upravuje technické standardy pro otevřená data v podobě Technického manuálu
  - datové formáty pro stupně otevřenosti 3 a 5
  - obecné požadavky na strukturu dat
  - konkrétní datová schémata pro doporučené datové sady na stupni 3

## Správa dat

- Otevřená data jsou nedílnou součástí datové architektury příslušného subjektu.
- Správa dat musí být vykonávána na „uzavřených“ i otevřených datech.

## Cíle řízení kvality dat

- Zajistit, že data jsou uživatelům dostupná v kvalitě, kterou potřebují
- Průběžně vyhodnocovat a zvyšovat kvalitu dat

## Obsah řízení kvality dat

- Fáze zlepšování kvality dat (Batini a kol., 2009):
  1. Popis současného stavu
  2. Hodnocení/měření kvality dat
  3. Zvyšování kvality dat



## Obsah řízení kvality dat\*

- Aktivita řízení kvality dat dle DAMA-DMBOK (Mosley a kol., 2010):

1. Propagace konceptu datové kvality
2. Identifikace a popis požadavků na kvalitu dat
3. Profilování, analýza a hodnocení datové kvality
4. Definice metrik datové kvality
5. Definice byznys pravidel pro datovou kvalitu
6. Testování a ověřování požadavků na datovou kvalitu
7. Nastavení a ověření úrovně služeb datové kvality
8. Průběžné sledování a vyhodnocování datové kvality
9. Řešení zjištěných problémů v datové kvalitě
10. Čištění a oprava chyb v datech
11. Návrh a implementace operativních procedur řízení datové kvality
12. Sledování a vyhodnocování výkonnosti operativních procedur řízení datové kvality

## Obecné dimenze datové kvality

- **Správnost**

- Data musí věrně zachycovat reálné objekty a události.

- **Úplnost**

- Míra, do jaké jsou přítomna všechna požadovaná data.

- **Validita**

- Míra, do jaké data odpovídají nadefinovaným syntaktickým pravidlům.

- **Včasnost**

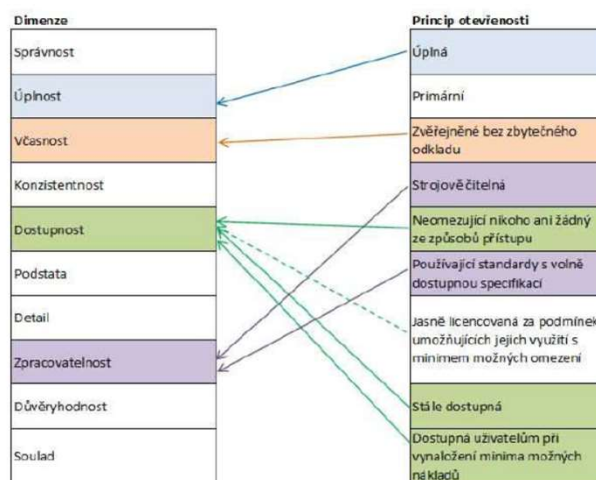
- Míra, do jaké data správně/aktuálně reprezentují realitu ke konkrétnímu časovému bodu.

## Obecné dimenze datové kvality

- **Konzistentnost datových hodnot**
  - Data vypovídající o stejném úseku reality/objektu musí být konzistentní napříč jednotlivými datovými zdroji.
- **Dostupnost**
  - Míra, do jaké je uživatel schopný přistupovat k datům.
- **Duplicitnost**
  - Žádný objekt či skutečnost není zaznamenána duplicitně.
- **Spolehlivost**
  - Data jsou spolehlivá, pokud víme, že jejich sběr a zpracování probíhá podle definovaného procesu.

Zdroj: upraveno dle (Tomčová, 2014)

## Mapování dimenzí kvality na principy OD



Zdroj: (Tomčová, 2014)

## Specifické dimenze kvality pro OD a LOD

Dimenze	Konsolidovaná definice	Příklad způsobu měření
Licencování	Zda jsou data opatřena vhodnou licencí upravující jejich užití a následné zpracování.	Vyhodnocení úrovně licence, která je datovému setu přidělena.
Primárnost	Míra do jaké jsou data zveřejněna tak, jak byla vytvořena.	Kontrolní součty u agregovaných statistických dat.
Zpracovatelnost	Míra do jaké jsou data srozumitelná a zpracovatelná v rámci automatických procesů.	Podíl dat, která jsou zapsána ve strojově čitelné formě.
Korektní zápis	Jsou uzly náležitě popsány podle toho, čím skutečnou jsou nebo označeny jen datovým typem? Zlepšuje schopnost vyhledávat a dotazovat se.	Podíl uzlů s korektním zápisem.

Zdroj: upraveno dle (Tomčová, 2014)

## Kategorie technik zvyšování kvality dat\*

### • **Techniky zaměřené na data**

- Zvýšení kvality dat změnou jejich hodnoty
- Příklady technik
  - Nahrazení nekvalitních dat jinými daty, která jsou kvalitnější
  - Standardizace – např. využívání číselníků
  - Identifikace a oprava chyb v datech
  - Atd.

### • **Techniky zaměřené na procesy**

- Zvýšení kvality dat úpravou procesů, v rámci kterých vznikají, aby bylo zamezeno vzniku chyb
- Příklady technik
  - Zavedení kroků do procesu, v rámci kterých jsou data kontrolována
  - Redesign procesu, aby v jednotlivých krocích již vznikala data vyšší kvality

Zdroj: (Batini a kol., 2009)