



Evropská unie  
Evropský sociální fond  
Operační program Zaměstnanost

Martin Černý

# Datová kvalita – metodika

31. 3. 2023



Evropská unie  
Evropský sociální fond  
Operační program Zaměstnanost



Česká informační agentura životního prostředí







## Dopady nekvalitních dat

- Problémy s rozhodováním
- Ztráta času
- Ztráta důvěry
- Ztráta příjmů
- Snížení kvality služeb pro zákazníky
- Dodatečné náklady





## Datová kvalita

Datová kvalita je definovaná jako míra požadovaných vlastností.

(přesnost, platnost, včasnost, úplnost, jedinečnost a konzistenci)



Data



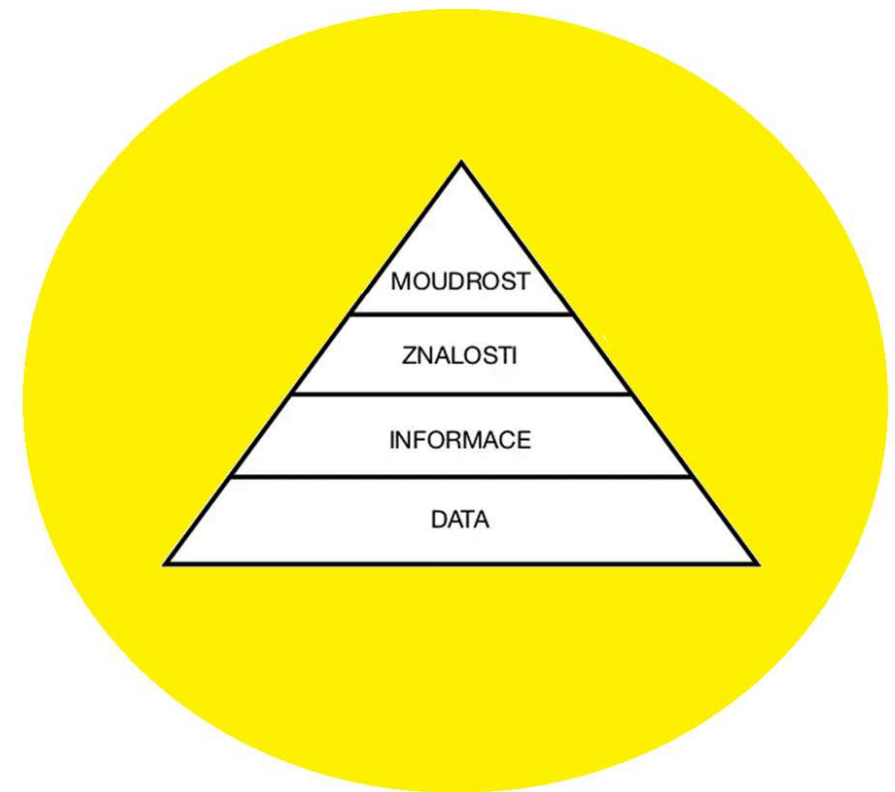
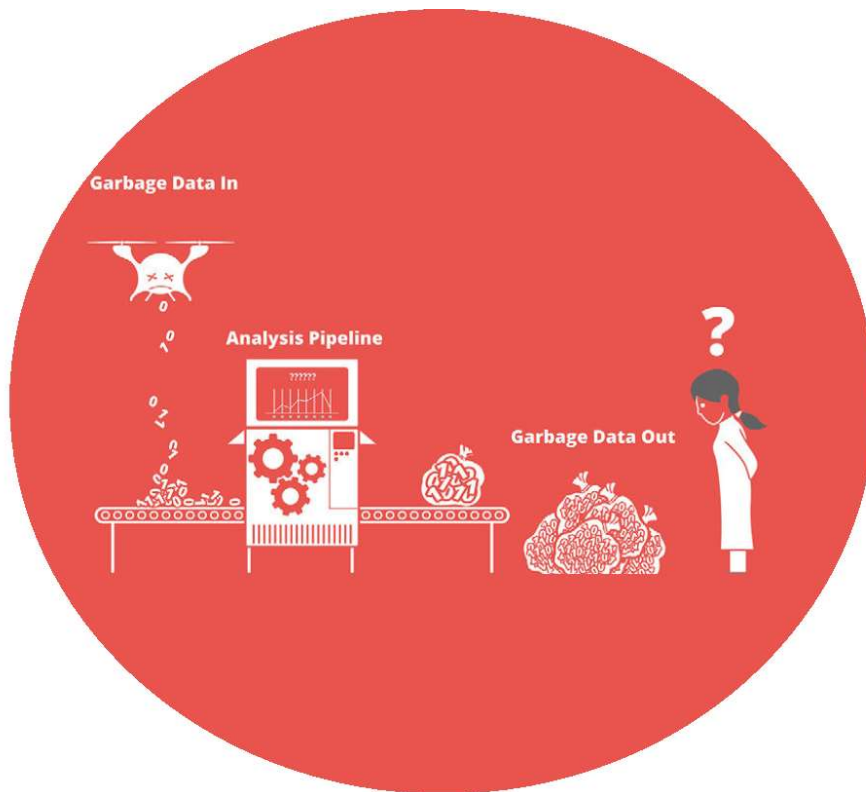
Metadata



Procesy



# GIGO - Pyramida znalostí



# Cílová skupina CENIA publikací

Zpráva o životním prostředí  
Krajské zprávy o životním prostředí  
jsou zpracovány na základě zákona č. 123/1998 Sb.

- MŽP
- Vláda, které je Zpráva předkládána ke schválení
- Senát
- Výboru pro životní prostředí Poslanecké sněmovny
- hodnocení indikátorů Státní politiky živ. prostředí pro podporu státní správy
- tisk, krajské úřady, odborná veřejnost, studenti, ...



Zpráva  
o životním prostředí  
České republiky

2020



Zpráva  
o životním prostředí  
v kraji Hl. m. Praha



Evropská unie  
Evropský sociální fond  
Operační program Zaměstnanost



## Datový citát

**„ Naše data jsou na 100 % spolehlivá”**

*Neřekl Nikdo*



Evropská unie  
Evropský sociální fond  
Operační program Zaměstnanost

**NERP**



## První kroky datové kvality

- Analýza stávajícího stavu
- Cíle – budoucí stav
- Politika
- Procesy
- Role a odpovědnosti
- Dimenze dat
- Kontrola



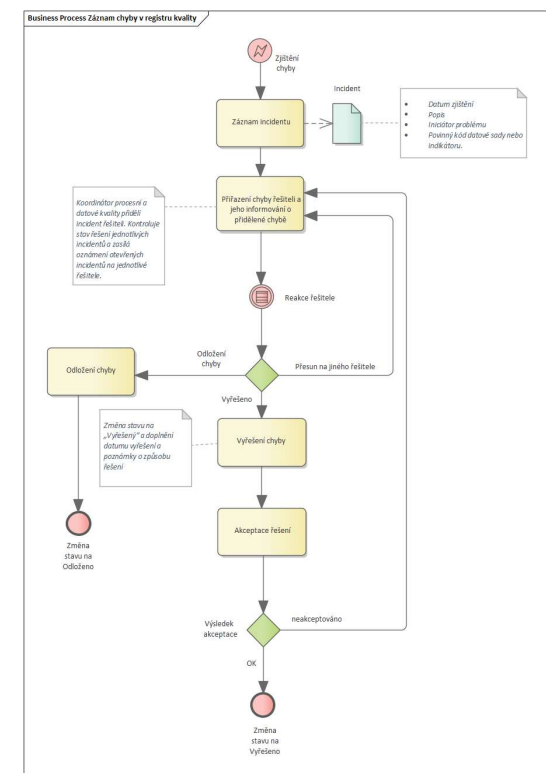
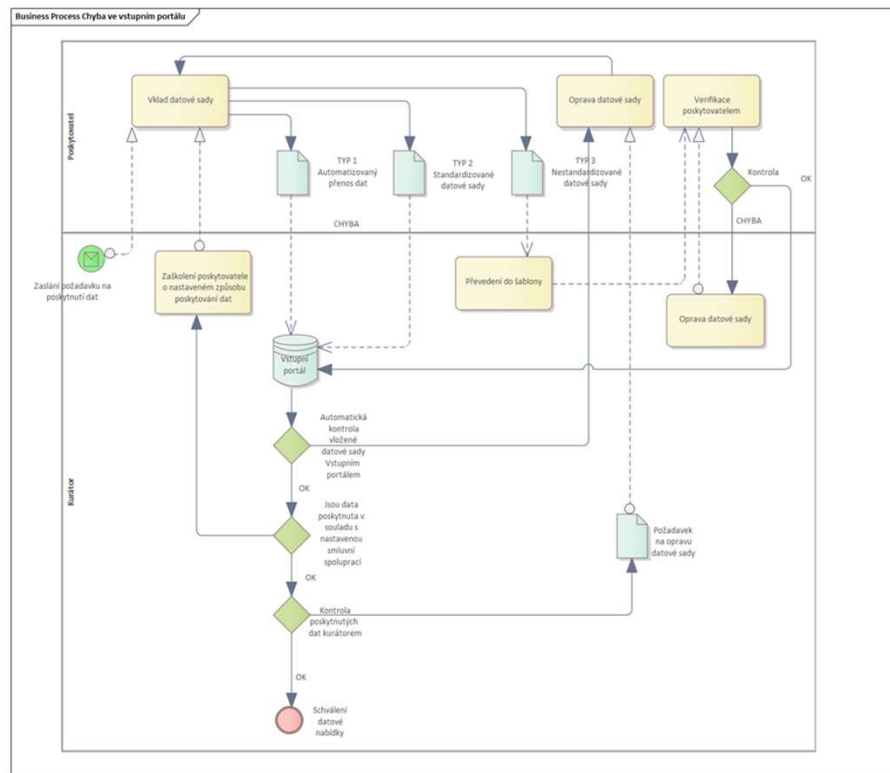


# Procesní katalog

ID	Název procesu	Stručný popis	Vlastník	Iniciační bussines událost	Vstupy	Hlavní kroky	Koncová bussines událost	Výstupy
	Kontrola, měření, audit procesů a 1 metodik	<ul style="list-style-type: none"> <li>Postup pro audit a měření nastavených Metrik</li> <li>Kontrola aktuálnosti dokumentace a procesů, shody s právními požadavky, normami a bezpečnosti informací</li> </ul>	koordinátor procesní a datové kvality	Termín pro audit a měření - 1 x rok	Výsledky měření, kontroly, dotazníkového šetření a záznamů v registru kvality	<ul style="list-style-type: none"> <li>Kontrola aktuálnosti Metodiky národního reportingu dat o životním prostředí</li> <li>Audit procesů</li> <li>Měření datové kvality a vyhodnocení</li> <li>Dotazníkové šetření</li> <li>Záznam registru nápravných opatření</li> </ul>	Identifikace neshody	Záznam v registru nápravných opatření
	Kontrola chyb ve 2 vstupním portálu	Postup pro zachycení chyby ve vstupním portálu	Kurátor dat	Zaslání požadavku na poskytnutí dat	Požadovaná data	<ul style="list-style-type: none"> <li>Nahrání datové sady do vstupního portálu</li> <li>Automatizované ověření vložených dat na vstupním portálu</li> <li>Verifikace šablony s daty poskytovatelem</li> <li>Oprava datové sady kurátorem</li> <li>Oprava datové sady poskytovatelem</li> <li>Schválení datové sady</li> </ul>	Schválení datové sady ve vstupním portálu	Počet datových sad vrácených poskytovateli s chybou. Počet datových sad po termínu.
	Záznam chyby v 3 registru kvality	Postup pro řešení chyb vzniklých po schválení ve vstupním portálu v průběhu životního cyklu dat.	Kurátor dat, Správce dat, Manager kvality dat	Zjištění chyby	Záznam incidentu v registru chyb	<ul style="list-style-type: none"> <li>Záznam v registru chyb</li> <li>Přiřazení incidentu řešiteli a jeho informování</li> <li>Řešení incidentu</li> <li>Odložení chyby</li> <li>Akceptace vyřešení problému</li> </ul>	Akceptace opravy incidentu	Počet incidentů a doba řešení incidentů v registru datové chyby



# Diagramy BPMN





## Role a odpovědnosti

- Kontaktní osoba poskytovatele dat
- Kurátor dat
- Správce dat
- Koordinátor kvality dat
- Uživatel dat



# Vlastnosti dat – dimenze dat

- Dostupnost
- Přesnost
- Úplnost
- Konzistence
- Duplicita
- Validita
- Relevance
- Spolehlivost
- Užitečnost
- Aj.

***Datový prvek  
Předmět datové  
kvality:***

**Text  
Číslo**





## Pravidla kvality dat

- Komplettnost – počet kompletně vyplněných datových šablon
- Chyba v hodnotách – počet přepsaných datových sad
- Přepsané hodnoty číselníku – syntaktická správnost počet přepsaných dimenzí číselníku/Podíl hodnot, které neodpovídají formálním pravidlům viz. syntaxe číselníku
- Špatný formát hodnot – použití nepovolených znaků v číselné hodnotě – např. text místo čísla
- Strojová zpracovatelnost – Počet datových sad v strojově nezpracovatelném formátu
- Včasnost



# Chyby ve vstupním portálu

Data portál

Agentura ochrany přírody a krajiny ČR

DATOVÉ SADY

- Seznam datových sad
- Datová sada
- Metadata
- Podmínky užití dat
- Přílohy
- Odevzdání dat

## Rybí přechody (2021) k odevzdání

Agentura ochrany přírody a krajiny ČR

Domů / Datové sady

Datová sada

### Rybí přechody (2021)

Kód sady: aopk-rybi-prechody-2021

Popis datové sady \*

Zadejte popis...

Zdroj dat

Agentura ochrany přírody a krajiny ČR (aopk) X Zadejte zdroje dat...

Poznámka: Lze vkládat pouze zdroje nabízené v seznamu. V případě zdroje zde neuvedeného prosím kontaktujte kurátora o jeho založení.

Uložením potvrzují správnost údajů **Před pokračováním potvrďte správnost údajů!**

Datové soubory (CSV nebo Excel)

Validace je Ok!

aopk-rybi-prechody.xlsx (aopk-rybi-prechody-2021-202208101500.xlsx)	XLSX	10.8.2022 15:00	
---	------	-----------------	--

[Uložit datovou sadu](#) [Nahrát datový soubor](#)

Rybí přechody (2021) **aktivní**

Rybí přechody (2022) **aktivní**

Excelová šablona ke stažení [Sloupce \(pro CSV\)](#)

Technická podpora  
no-reply-dataport@mzp.cz

Komentář pro zpracování  
Zpráva pro kurátora ...



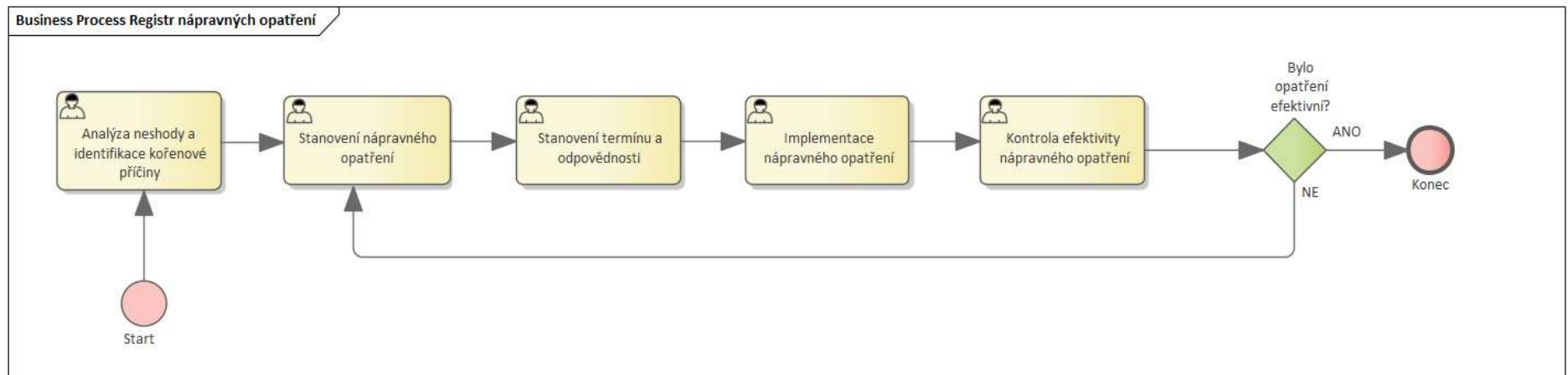
# Registr chyb

Evropská unie Evropský sociální fond Operační program Zaměstnanost		cenia		Registr ne/kvality				
Projekt:				NERP - Národní environmentální platforma reportingu, CZ.03.4.74/0.0/0.0/15_025/0016059				
ID	Datum zjištění	Kód DS/ indikátoru	Popis problému	Téma	Iniciátor	Aktuální řešitel	Stav řešení	Datum vyřešen
1	21.01.2022	cenia-indikator-podil-objemu-produkce-druhonych-surovin-na-primem-materialovem-vstupu-2018	Datum pořízení není 2018		Miroslav Kukuc	Jan Fránek	Vyřešený	24.01.2022
2	21.01.2022	cenia-indikator-podil-jednotlivych-kategorii-nakladani-s-odpady-2018	Datum pořízení není 2018		Miroslav Kukuc	Jan Fránek	Vyřešený	24.01.2022
3	21.01.2022	cenia-indikator-podil-evropsky-vyznamnych-druhu-nachazejicich-se-v-nedostatecnem-a-nepriznivem-2018	Datum pořízení není 2018		Miroslav Kukuc	Jan Fránek	Vyřešený	24.01.2022
4	21.01.2022	cenia-indikator-kvalita-zemedelske-pudy-2018	Datum pořízení není 2018		Miroslav Kukuc	Jan Fránek	Vyřešený	24.01.2022
5	21.01.2022	cenia-indikator-kvalita-koupacich-vod-2018	Datum pořízení není 2018		Miroslav Kukuc	Jan Fránek	Vyřešený	24.01.2022
6	21.01.2022	data-uzei-eko-zemedelstvi-2018	Nesprávná jednotka v DS data-uzei-eko-zemedelstvi-2018		Miroslav Kukuc	Andrea Danelová	Vyřešený	24.01.2022
7	25.01.2022	md-preprava-osoby-vzdalenost-celkova	Kód DS přesahuje max. povolený počet znaků (musíme opravit vše: šablonu, mds, ckan, model...)		Miroslav Elko	Jiří Přech	Vyřešený	07.04.2022
8	25.01.2022	md-preprava-veci-vzdalenost-celkova	Kód DS přesahuje max. povolený počet znaků (musíme opravit vše: šablonu, mds, ckan,		Miroslav Elko	Jiří Přech	Vyřešený	07.04.2022



# Registr nápravných opatření

- Postup pro řešení opakovaných chyb, závažných chyb, neshod a potenciálních neshod v oblasti datové kvality včetně bezpečnostních incidentů.







## Kontrola

- Audit procesů a metodik
- Měření a vyhodnocení metrik
- Zpětná vazba
- Přezkoumání rizik
- Registr nápravných opatření
- Zpráva o kvalitě dat
- Přezkoumání vedením





## Závěr

- Proaktivní přístup
- I minimální řešení datové kvality má význam pro zlepšení
- Zapojte kolegy do řešení zlepšování dat
- Naše metodika je volně dostupná

<https://nerp.cenia.cz/wp-content/uploads/2023/02/Methodika-NERP-Datova-kvalita.pdf>

