



Evropská unie
Evropský sociální fond
Operační program Zaměstnanost

Miroslav Kukuc

Seminář NERP II.

13.1.2022, CENIA



Evropská unie
Evropský sociální fond
Operační program Zaměstnanost



NERP

Národní Enviromentální Reportingová Platforma

Optimalizace systému řízení příjmu, validace, zpracování a reportingu
datových sad v rezortu životního prostředí

CZ.03.4.74/0.0/0.0/15_025/0016059



Evropská unie
Evropský sociální fond
Operační program Zaměstnanost





Účel a obsah vzdělávání

Účelem školení je vzdělávání organizací státní správy k vyšší digitalizaci, zpracování, využívání a zpřístupnění dat o životním prostředí

- **Data**
- Legislativa
- **Otevřená data**
- **Procesy**
- Životní prostředí





Program

- Proces akvizice dat
- Příprava vizualizací
- Envirometr
- Požadavky na datové sady





Programové prohlášení vlády

~~například, v oblasti veřejného nadobutí přírody~~

- Zajistíme kvalitní a veřejně přístupný monitoring stavu všech klíčových složek životního prostředí.
- Zajistíme podporu výzkumu, vývoje a inovací v oblasti životního prostředí pokračováním programu Prostředí pro život.



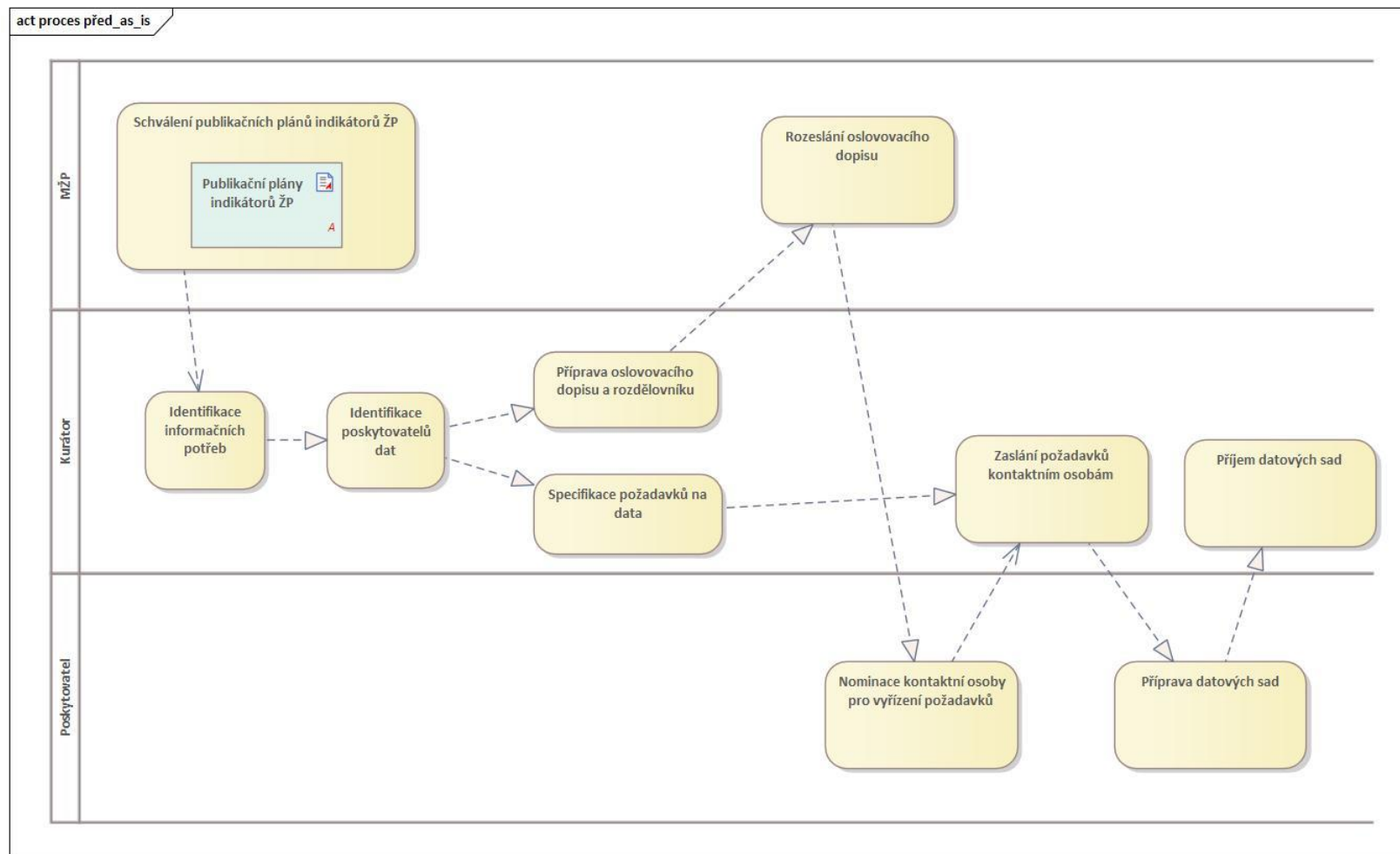
Procesy zpracování dat



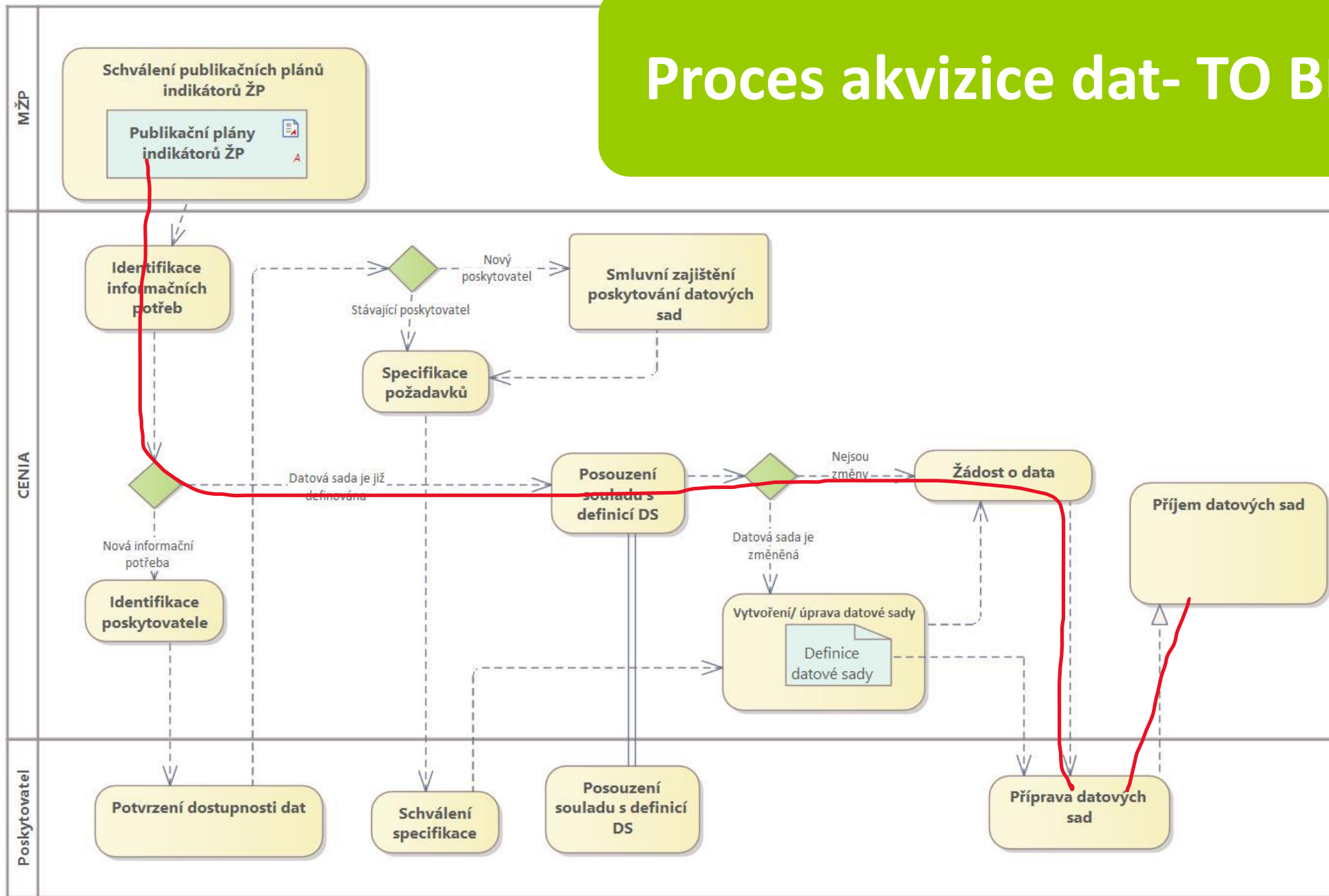
Evropská unie
Evropský sociální fond
Operační program Zaměstnanost



Proces akvizice dat – AS IS

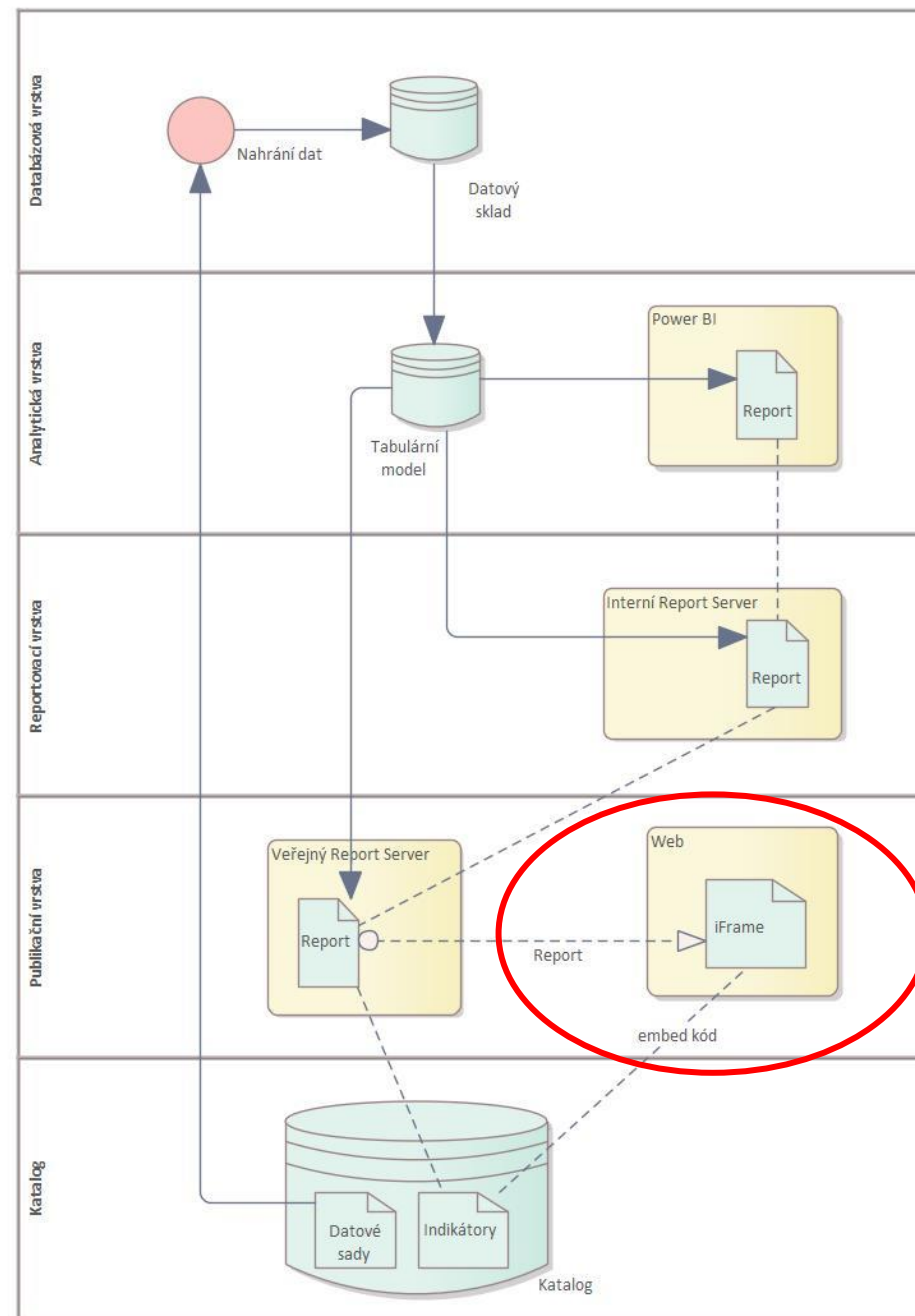


Proces akvizice dat- TO BE



Publikace na webu

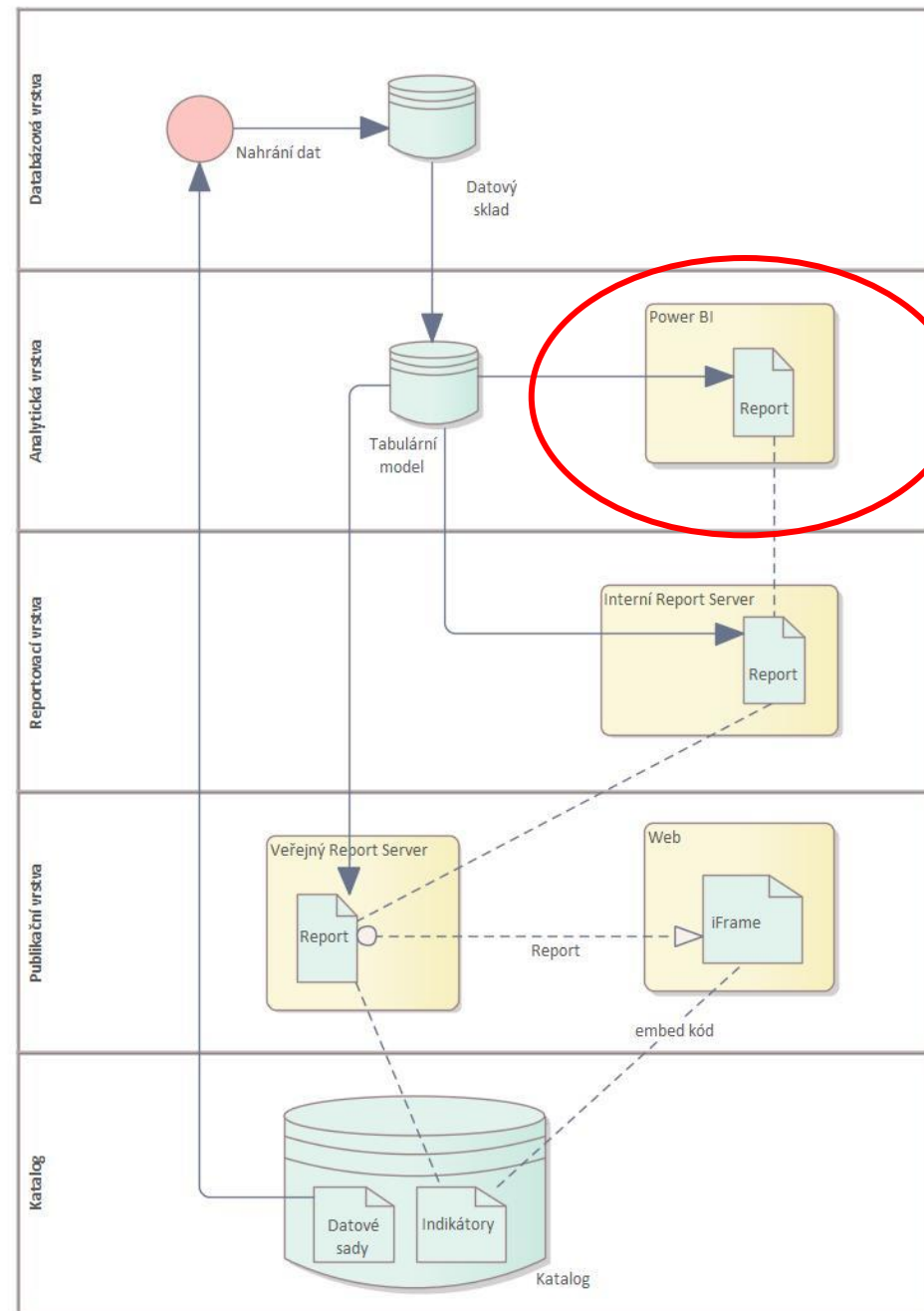
- Webový portál přebírá obsah z katalogu
- Vizualizace je zobrazována prostřednictvím embed kódu – v i-frame se zobrazuje report publikovaný na veřejném report serveru
- Data jsou živě připojena na databázi



Příprava reportu

- V Power BI
- Data jsou živě připojena na databázi

Business Process Report flow





Envirometr

- Web pro data o životním prostředí

www.envirometr.cz



Workshop

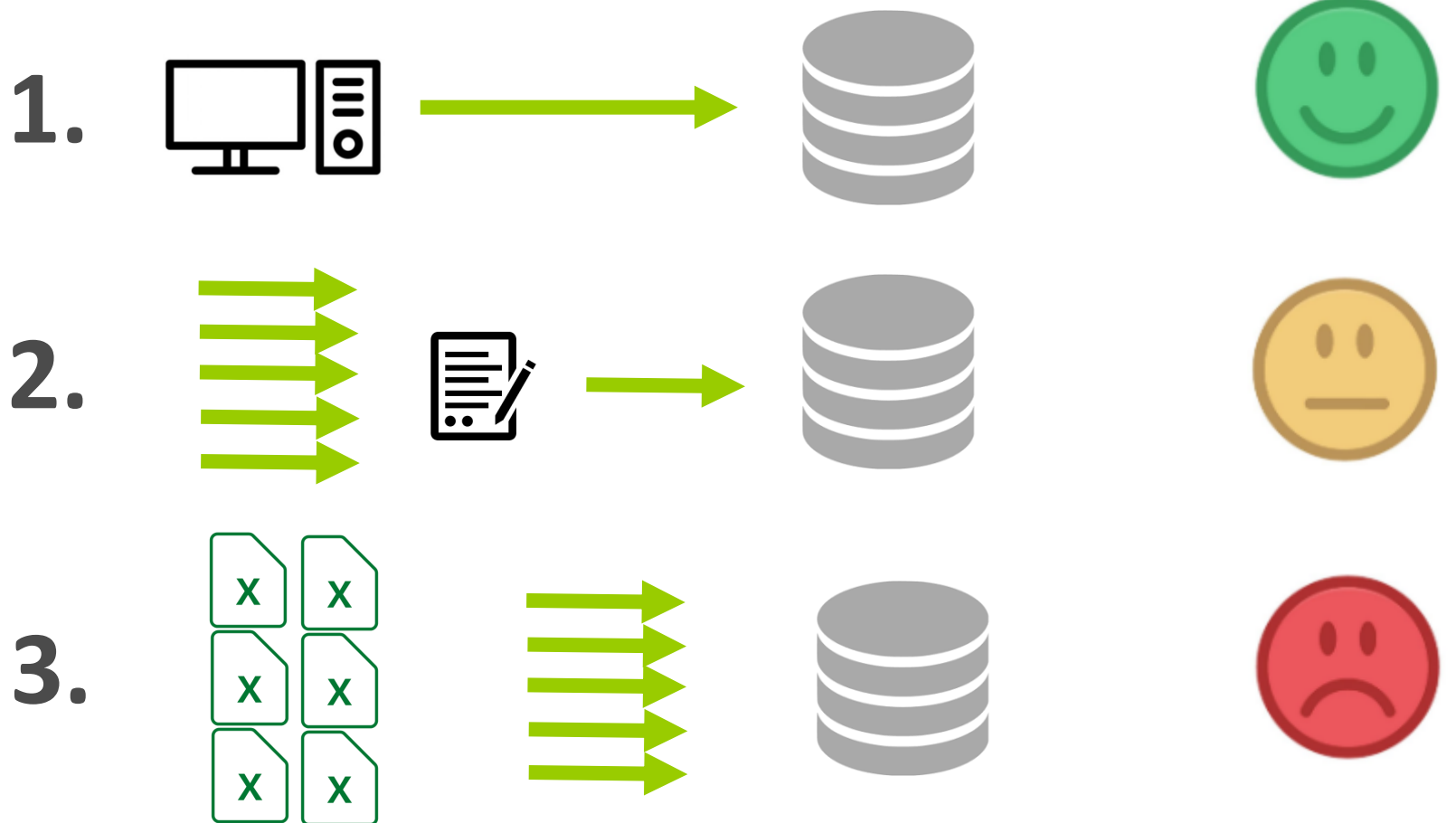
- Otevřete si www.envirometr.cz
- Zkuste se zorientovat v navigaci na stránce
 - Napište nám zpětnou vazbu **10 min**
- Projděte si indikátory, které Vás zajímají, nebo které jsou vytvořeny nad Vašimi daty
 - Napište nám zpětnou vazbu **15 min**



Metodika, formáty dat



Kterým způsobem probíhá zpracování dat o životním prostředí?



Komplexnost validace

Pracnost validace

Každý soubor je jiný

Základní validace

Stejný soubor od více poskytovatelů

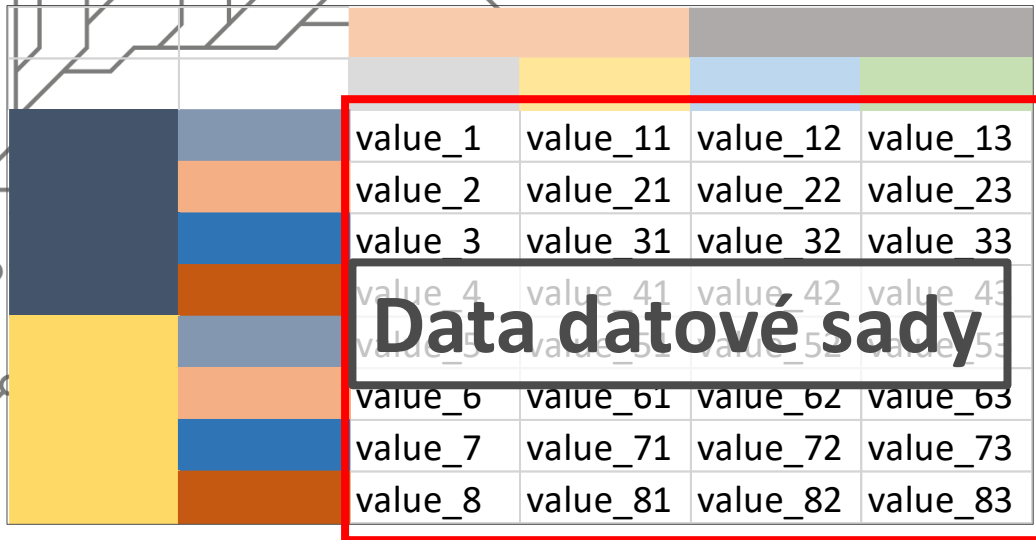
Komplexní validace

1

n

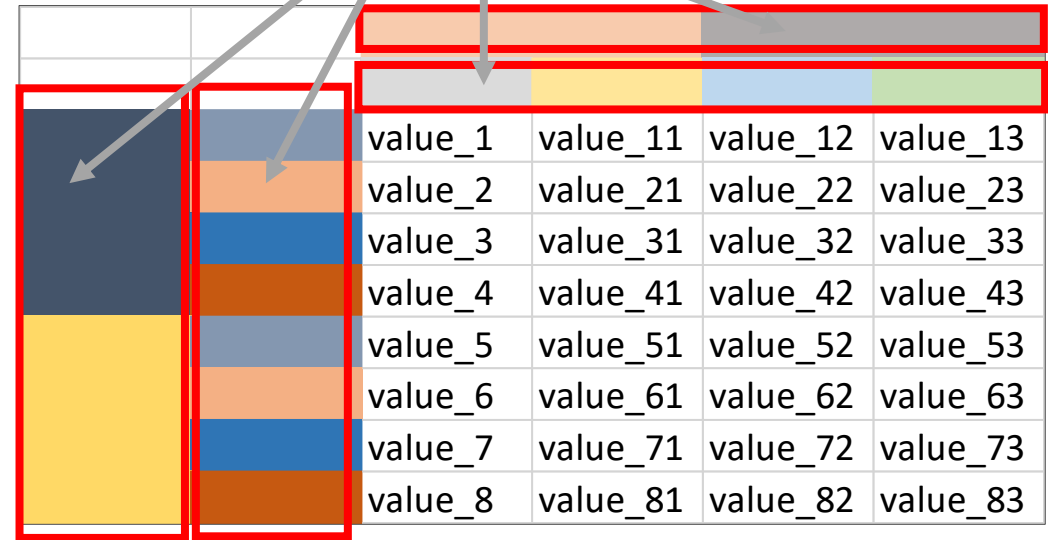
Počet souborů jednoho typu

Dimenze datové sady



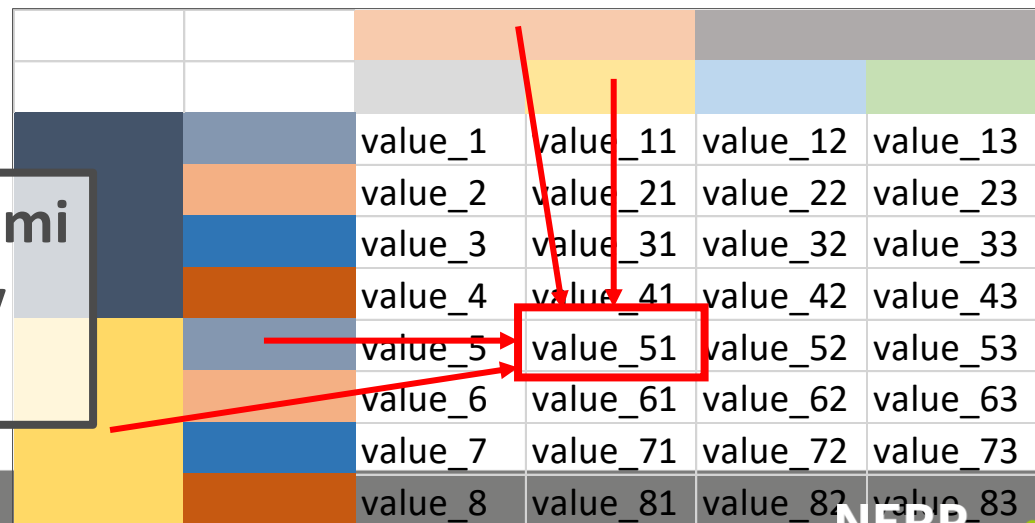
value_1	value_11	value_12	value_13
value_2	value_21	value_22	value_23
value_3	value_31	value_32	value_33
value_4	value_41	value_42	value_43
value_5	value_51	value_52	value_53
value_6	value_61	value_62	value_63
value_7	value_71	value_72	value_73
value_8	value_81	value_82	value_83

Data datové sady



value_1	value_11	value_12	value_13
value_2	value_21	value_22	value_23
value_3	value_31	value_32	value_33
value_4	value_41	value_42	value_43
value_5	value_51	value_52	value_53
value_6	value_61	value_62	value_63
value_7	value_71	value_72	value_73
value_8	value_81	value_82	value_83

Číslo je identifikováno hodnotami na dimenzích a dalšími atributy (např. jednotka)



value_1	value_11	value_12	value_13
value_2	value_21	value_22	value_23
value_3	value_31	value_32	value_33
value_4	value_41	value_42	value_43
value_5	value_51	value_52	value_53
value_6	value_61	value_62	value_63
value_7	value_71	value_72	value_73
value_8	value_81	value_82	value_83



Číselník = výčet všech hodnot dimenze

Umožňuje analyzovat data napříč různými datovými sadami

Dimenze – Dny v týdnu

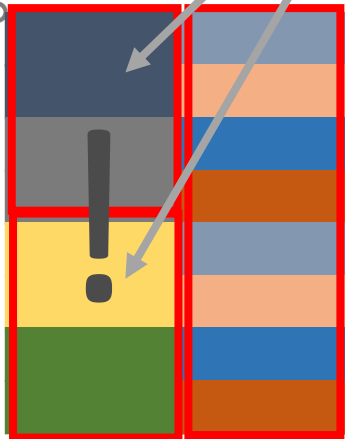
Číselník:

		value_1	value_11	value_12	value_13
		value_2	value_21	value_22	value_23
		value_3	value_31	value_32	value_33
		value_4	value_41	value_42	value_43
		value_5	value_51	value_52	value_53
		value_6	value_61	value_62	value_63
		value_7	value_71	value_72	value_73
		value_8	value_81	value_82	value_83

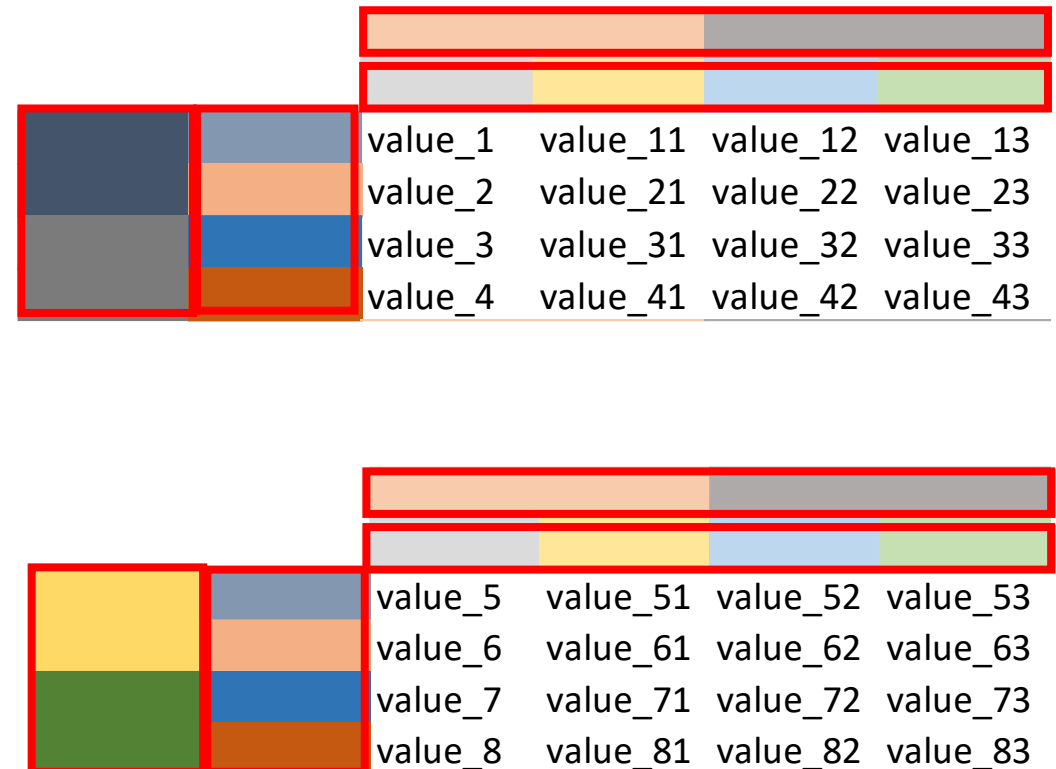
- pondělí
- úterý
- středa
- čtvrtek
- pátek
- sobota
- neděle



2 dimenze datové sady



value_1	value_11	value_12	value_13
value_2	value_21	value_22	value_23
value_3	value_31	value_32	value_33
value_4	value_41	value_42	value_43
value_5	value_51	value_52	value_53
value_6	value_61	value_62	value_63
value_7	value_71	value_72	value_73
value_8	value_81	value_82	value_83



Nehomogenní tabulka



Homogenní tabulky

Těžba dřeva, 1970–2019

Dřeviny	2013	2014	2015	2016	2017	2018	2019
	tis. m ³ b. k.						
Těžba dřeva celkem	15,331	15,476	16,163	17,617	19,387	25,689	32,586
v tom:							
jehličnaté	13,229	13,472	14,385	15,924	17,735	24,213	31,313
z toho:							
smrk	10,667	10,984	12,230	13,986	15,775	22,412	29,350
jedle	119	117	107	115	116	136	157
borovice	1,879	1,805	1,558	1,368	1,363	1,127	1,288
modřín	532	523	462	424	457	522	505
listnaté	2,102	2,004	1,778	1,693	1,652	1,476	1,273
z toho:							
topol osika	485	448	410	391	353	305	264
buk	949	897	763	747	721	654	567
lípa	74	73	66	54	52	42	37
topolosika	88	91	76	60	64	48	42
Nahodilá těžba	4,248	4,527	8,153	9,399	11,743	23,013	30,945
v tom:							
živelní	2,277	2,455	4,388	2,636	4,345	8,378	5,879
exhalační	22	19	28	29	20	19	20
hmyzová	1,052	1,133	2,309	4,420	5,853	13,059	22,780
ostatní	897	920	1,428	2,314	1,525	1,557	2,266

Pozn.: Těžba dřeva zahrnuje hmotu hrubí i část nehroubí (většina nehroubí – těžební zbytky – nezapočteno), která byla přijata jako hotový sortiment
Zdroj: ČSÚ

Těžba dřeva - dřeviny

Dřeviny	2013	2014	2015	2016	2017	2018	2019
	tis. m ³ b. k.						
smrk	10,667	10,984	12,230	13,986	15,775	22,412	29,350
jedle	119	117	107	115	116	136	157
borovice	1,879	1,805	1,558	1,368	1,363	1,127	1,288
modřín	532	523	462	424	457	522	505
ostatní jehličnaté	32	43	28	31	24	16	13
dub	485	448	410	391	353	305	264
buk	949	897	763	747	721	654	567
lípa	74	73	66	54	52	42	37
topol osika	88	91	76	60	64	48	42
ostatní listnaté	506	495	463	441	462	427	363

Nahodilá těžba

Nahodilá těžba	2013	2014	2015	2016	2017	2018	2019
	tis. m ³ b. k.						
živelní	2,277	2,455	4,388	2,636	4,345	8,378	5,879
exhalační	22	19	28	29	20	19	20
hmyzová	1,052	1,133	2,309	4,420	5,853	13,059	22,780
ostatní	897	920	1,428	2,314	1,525	1,557	2,266

Metadata



Evropská unie
Evropský sociální fond
Operační program Zamestnanost



Číselník = výčet všech hodnot dimenze

Těžba dřeva - dřeviny

Dřeviny	2013	2014	2015	2016	2017	2018	2019
	tis. m ³ b. k.						
smrk	10,667	10,984	12,230	13,986	15,775	22,412	29,350
jedle	119	117	107	115	116	136	157
borovice	1,879	1,805	1,558	1,368	1,363	1,127	1,288
modřín	532	523	462	424	457	522	505
ostatní jehličnaté	32	43	410	31	24	16	13
dub	485	448	410	391	353	305	264
buk	949	897	763	747	721	654	567
lípa	74	73	66	54	52	42	37
topol osika	88	91	76	60	64	48	42
ostatní listnaté	506	495	463	441	462	427	363

Nahodilá těžba

Nahodilá těžba	2013	2014	2015	2016	2017	2018	2019
	tis. m ³ b. k.						
živelní	2,277	2,455	4,368	2,636	4,345	8,378	5,879
exhalační	22	19	28	29	20	19	20
hmyzová	1,052	1,133	2,309	4,420	5,853	13,059	22,780
ostatní	897	920	1,428	2,314	1,525	1,557	2,266

Číselníky jsou společné pro všechny DS

Dřeviny

smrk jehličnaté

jedle jehličnaté

borovice jehličnaté

modřín jehličnaté

dub listnaté

buk listnaté

lípa listnaté

topol osika listnaté

javor listnaté

jasan listnaté

DSC1

DS1

DS2

Nahodilá těžba

živelní

exhalační

hmyzová

ostatní

DSC2

Rok

1998

1999

2000

2002

2003

2004

2005

2006

2007

2008

2009

2010

2011

2012

2013

2014

2015

2016

2017

2018

2019



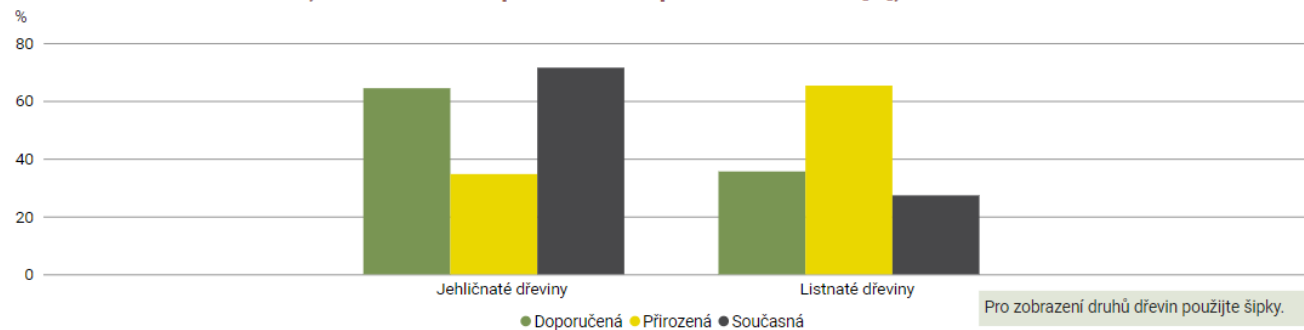
Hierarchie

DOMŮ > VÝPIS INDIKÁTORŮ > DRUHOVÁ SKLADBA LESŮ (2018)

Druhová skladba lesů (2018)

Toto je popis indikátoru.

Druhová skladba lesů v ČR, rekonstruovaná přirozená a doporučená skladba [%], 2018

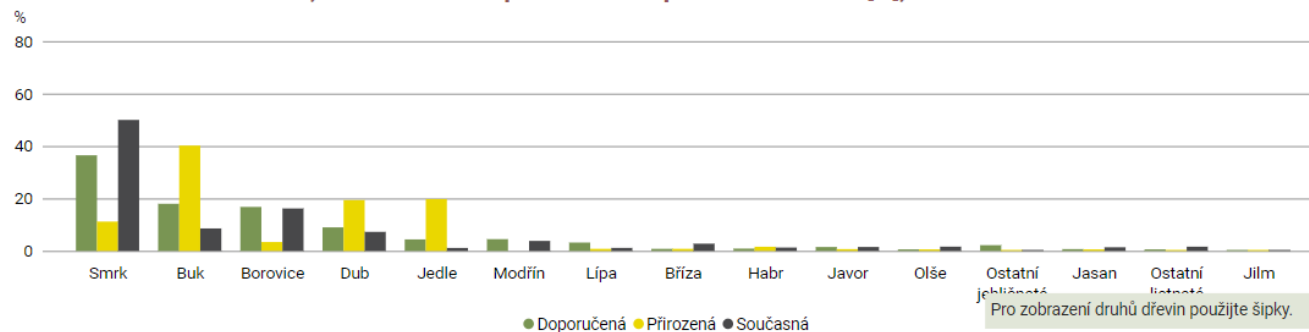


DOMŮ > VÝPIS INDIKÁTORŮ > DRUHOVÁ SKLADBA LESŮ (2018)

Druhová skladba lesů (2018)

Toto je popis indikátoru.

Druhová skladba lesů v ČR, rekonstruovaná přirozená a doporučená skladba [%], 2018



Číselníky

Nahodilá těžba	1970	1980	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000
celkem	3,226	7,060	9,822	6,939	5,400	8,100	9,282	7,855	6,855	5,527	3,840	3,736	3,288
živelní	-	-	8,701	5,309	2,803	3,935	4,355	2,766	4,058	3,982	2,592	2,743	2,388
exhalační	-	-	289	290	286	323	276	303	192	212	171	126	78
hmyzová	-	-	178	257	837	2,097	2,211	2,376	1,145	546	408	327	320
ostatní	-	-	654	1,083	1,474	1,745	2,440	2,410	1,460	787	669	540	502

Zdroj: ČSÚ

Nahodilá těžba	1970	1980	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000
celkem	3,226	7,060	9,822	6,939	5,400	8,100	9,282	7,855	6,855	5,527	3,840	3,736	3,288
živelní			8,701	5,309	2,803	3,935	4,355	2,766	4,058	3,982	2,592	2,743	2,388
exhalační			289	290	286	323	276	303	192	212	171	126	78
hmyzová			178	257	837	2,097	2,211	2,376	1,145	546	408	327	320
ostatní nahodilá těžba			654	1,083	1,474	1,745	2,440	2,410	1,460	787	669	540	502
nedefinováno	3,226	7,060	0	0	0	0	0	0	0	0	0	0	0
Celkem	3,226	7,060	9,822	6,939	5,400	8,100	9,282	7,855	6,855	5,527	3,840	3,736	3,288

Výpočet v BI

Číselníky

Celkem



Nedefinováno



Ostatní xyz





Faktové tabulky

- Formát dat (celé číslo, desetinné číslo, text, datum, ...)
- Hodnoty jako **číslo**, ne **poznámky, texty, zvláštní znaky, vzorce**
- Chybějící hodnota = **prázdný záznam**
- Zaokrouhlení
- Vhodná agregace
- Poznámky





Datové formáty

Typické strojově zpracovatelné datové formáty:

- csv + json
- json
- xml



Formát csv

```
1 C-ROKY,C-GEO-KRAJE,HODNOTA,C-JEDNOTKY,C-GEO-TYP-UZEMI,C-AUTOR
2 2018,Hlavní město Praha,0,ha,Národní parky,AOPK
3 2018,Středočeský kraj,0,ha,Národní parky,AOPK
4 2018,Jihočeský kraj,34073,ha,Národní parky,AOPK
5 2018,Plzeňský kraj,34511,ha,Národní parky,AOPK
6 2018,Karlovarský kraj,0,ha,Národní parky,AOPK
7 2018,Ústecký kraj,7926,ha,Národní parky,AOPK
8 2018,Liberecký kraj,11649,ha,Národní parky,AOPK
9 2018,Královéhradecký kraj,24671,ha,Národní parky,AOPK
10 2018,Pardubický kraj,0,ha,Národní parky,AOPK
11 2018,Kraj Vysočina,0,ha,Národní parky,AOPK
12 2018,Jihomoravský kraj,6274,ha,Národní parky,AOPK
13 2018,Olomoucký kraj,0,ha,Národní parky,AOPK
14 2018,Zlínský kraj,0,ha,Národní parky,AOPK
```



Strojově čitelný formát

Rozloha chráněných území dle krajů			
	2018	2018	2018
Národní parky		Chráněné krajinné oblasti	Národní pří
ha		ha	ha
Hlavní město Praha	0	518	145
Středočeský kraj	0	109,785	615
Jihočeský kraj	34,073	164,034	1,335
Plzeňský kraj	34,511	96,876	221
Karlovarský kraj	0	5	
Ústecký kraj	7,926	1	
Liberecký kraj	11,649	9	
Královéhradecký kraj	24,671	7	
Pardubický kraj	0	3	

```

1 C-ROKY,C-GEO-KRAJE,HODNOTA,C-JEDNOTKY,C-GEO-TYP-UZEMI,C-AUTOR
2 2018,Hlavní město Praha,0,ha,Národní parky,AOPK
3 2018,Středočeský kraj,0,ha,Národní parky,AOPK
4 2018,Jihočeský kraj,34073,ha,Národní parky,AOPK
5 2018,Plzeňský kraj,34511,ha,Národní parky,AOPK
6 2018,Karlovarský kraj,0,ha,Národní parky,AOPK
7 2018,Ústecký kraj,7926,ha,Národní parky,AOPK
    
```

C-ROKY	C-GEO-KRAJE	HODNOTA	C-JEDNOTKY	C-GEO-TYP-UZEMI	C-AUTOR
2018	Hlavní město Praha	0	ha	Národní parky	AOPK
2018	Středočeský kraj	0	ha	Národní parky	AOPK
2018	Jihočeský kraj	34073	ha	Národní parky	AOPK
2018	Plzeňský kraj	34511	ha	Národní parky	AOPK
2018	Karlovarský kraj	0	ha	Národní parky	AOPK
2018	Ústecký kraj	7926	ha	Národní parky	AOPK
2018	Liberecký kraj	11649	ha	Národní parky	AOPK
2018	Královéhradecký kraj	24671	ha	Národní parky	AOPK





Csv formát

Základní pravidla pro distribuce datových sad ve formátu csv:

- ✓ utf-8 kódování
- ✓ oddělovač hodnot, nebo řetězců = čárka
- ✓ textové řetězce (stringy) v uvozovkách
- ✓ desetinná tečka v číselných hodnotách
- ✓ jednotky zvlášť ve sloupci
- ✓ prázdná hodnota jako prázdná (ne „null“, „n/a“ apod.)
- ✓ pojmenování hlaviček konzistentní napříč soubory
- ☒ prázdné řádky a sloupce

Xls šablona datové sady

Odběry podzemních a povrchových vod

	Povrchové vody	Podzemní vody	Povrchové vody	Podzemní vody	Povrchové vody	Podzemní vody	Povrchové vody	Podzemní vody	Povrchové vody	Podzemní vody
	2016	2016	2017	2017	2018	2018	2019	2019	2020	2020
Odvětví dle CZ-NACE	m3	m3	m3	m3	m3	m3	m3	m3	m3	m3
01-03 Zemědělství, lesnictví a rybářství	33 351 000	14 124 000	31 895 000	14 701 000	32 439 000	15 282 000	27 352 000	15 259 000	23 350 000	15 572 000
05-33 Průmysl (včetně dobývání nerostných surovin)	213 801 000	36 328 000	219 977 000	37 308 000	224 103 000	36 553 000	218 942 000	34 980 000	198 784 000	32 887 000
35 Výroba a rozvod elektřiny, plynu, tepla a klimatizovaného vzduchu	700 480 000	1 957 000	677 758 000	2 039 000	626 884 000	2 093 000	562 160 000	1 868 000	464 385 000	1 852 000
36 Shromažďování, úprava a rozvod vody	313 097 000	292 529 000	320 566 000	296 063 000	325 961 000	299 085 000	326 171 000	290 095 000	312 292 000	288 914 000
37 Činnosti související s odpadními vodami	13 000	274 000	13 000	262 000	11 000	273 000	0	277 000	0	277 000
38-96 Ostatní (včetně stavebnictví)	11 313 000	17 589 000	10 798 000	19 009 000	11 327 000	17 077 000	12 336 000	16 839 000	12 226 000	15 392 000

Popis datové sady:

Zpráva pro zpracovatele:

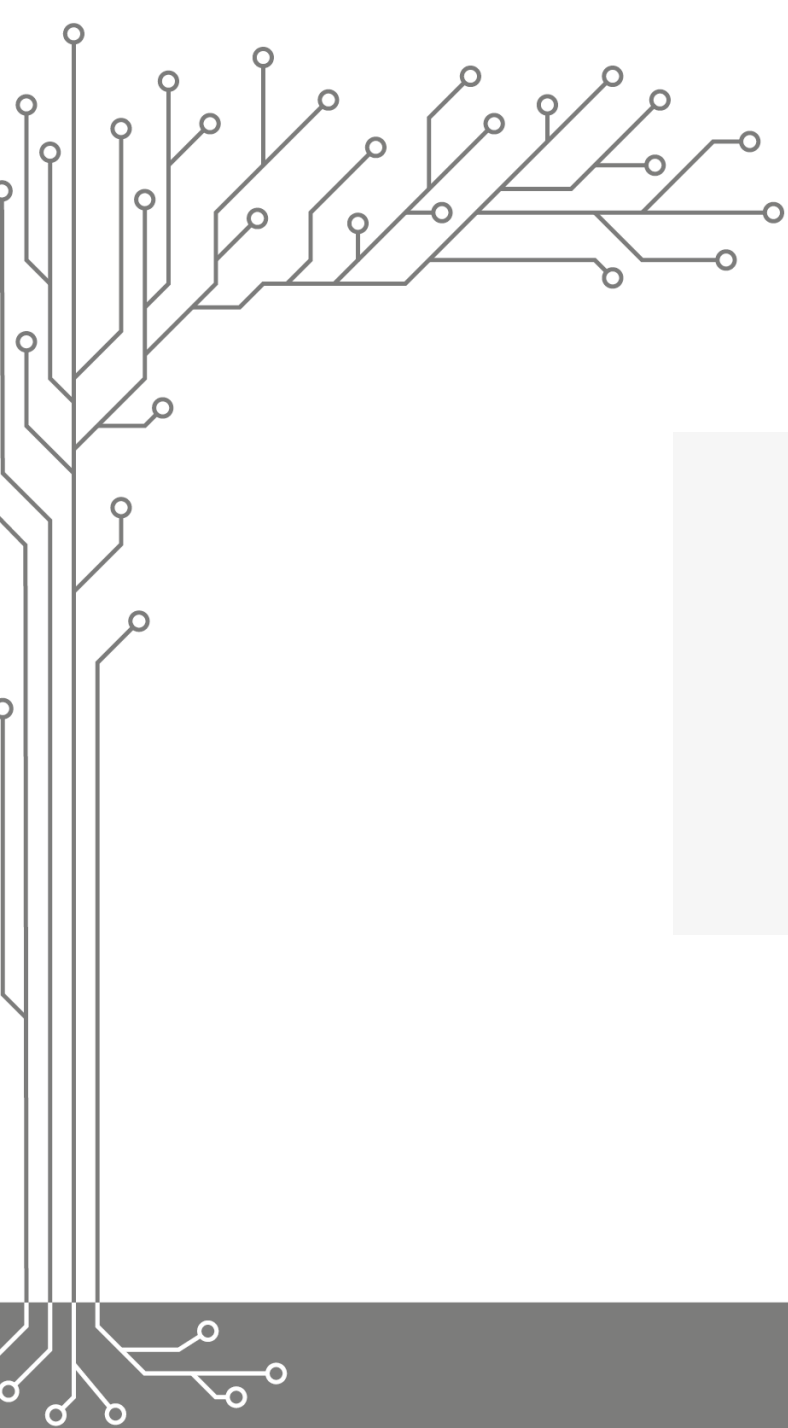
Kód datové sady: csu-voda-odbery

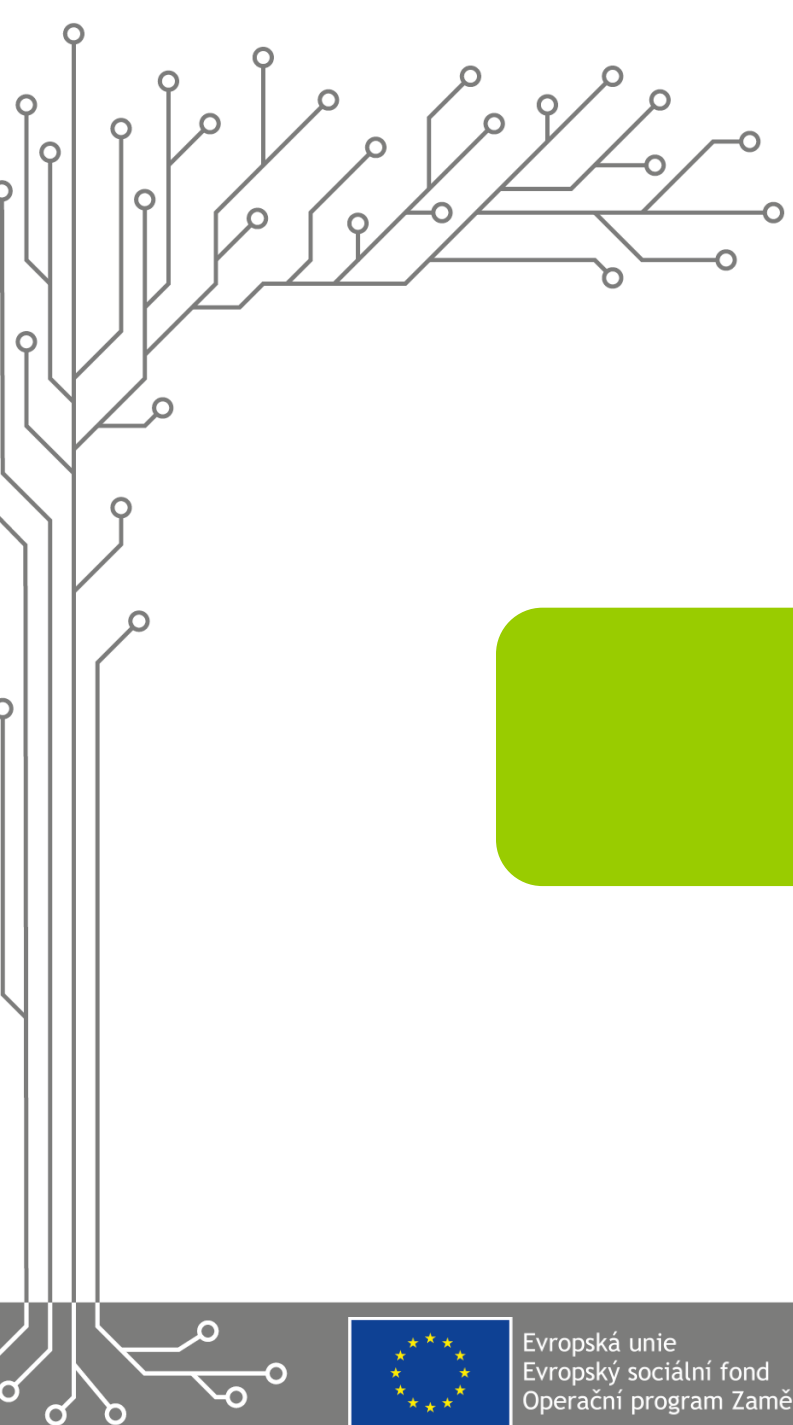
Zdroj: ČSÚ

Standardizace

- Otevřená data
 - Metadata
 - Strojově čitelný formát



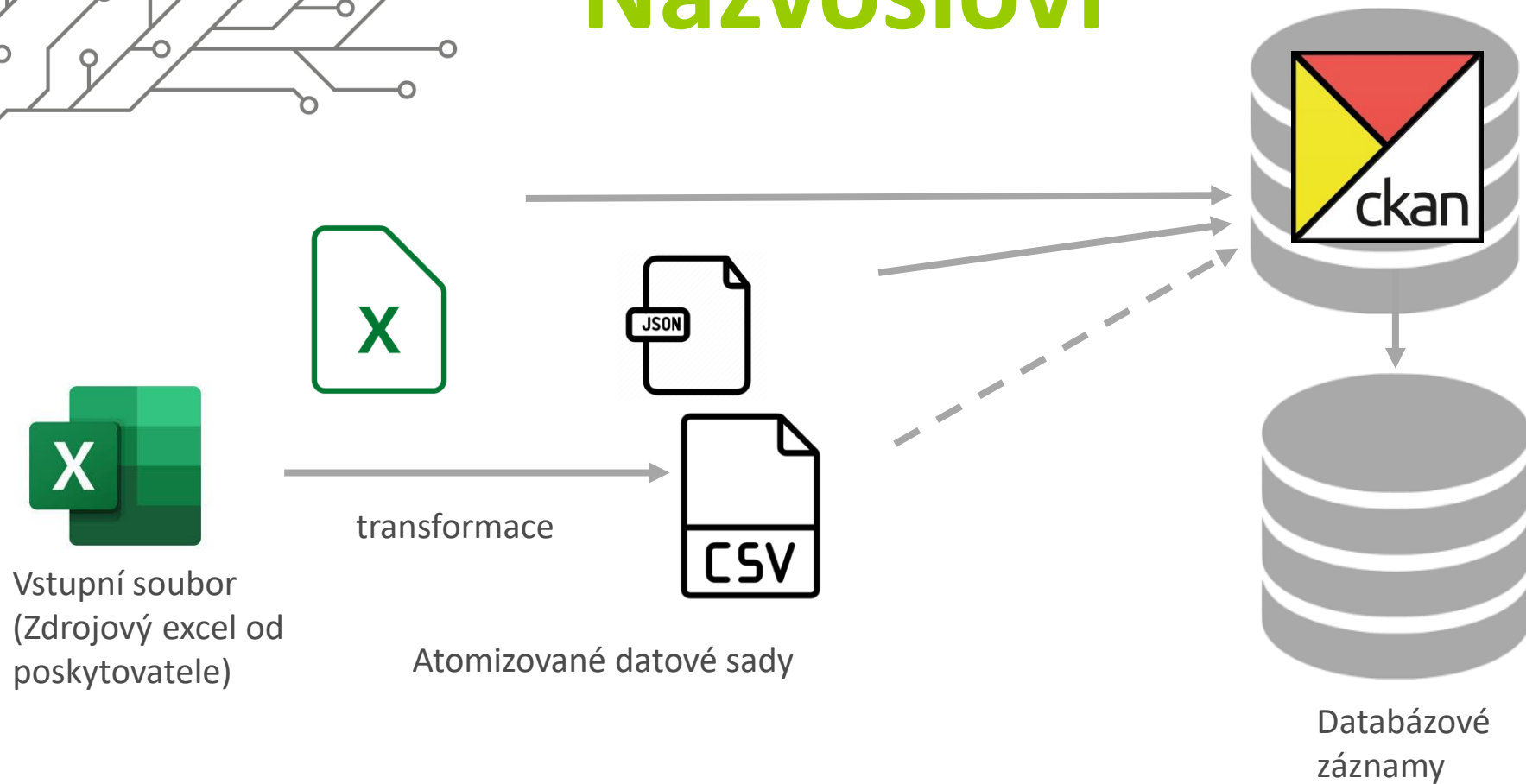




Backup



Názvosloví



poskytovatel_soubor_DS_rok_datum_platnosti

csu-lesy-tezba-celkem-2019-20201007

csu-lesy-nahodila-tezba-2019-20201007



Evropská unie
Evropský sociální fond
Operační program Zaměstnanost



Schema datové sady

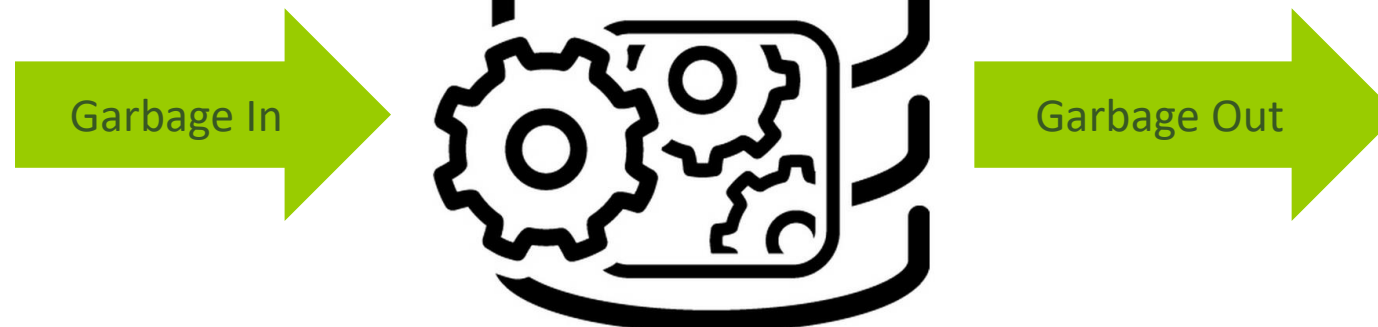
Schema musí být definováno pro každou datovou sadu před jejím prvním příjmem

Schema obsahuje všechny parametry datové sady, které jsou potřebné pro její strojové zpracování

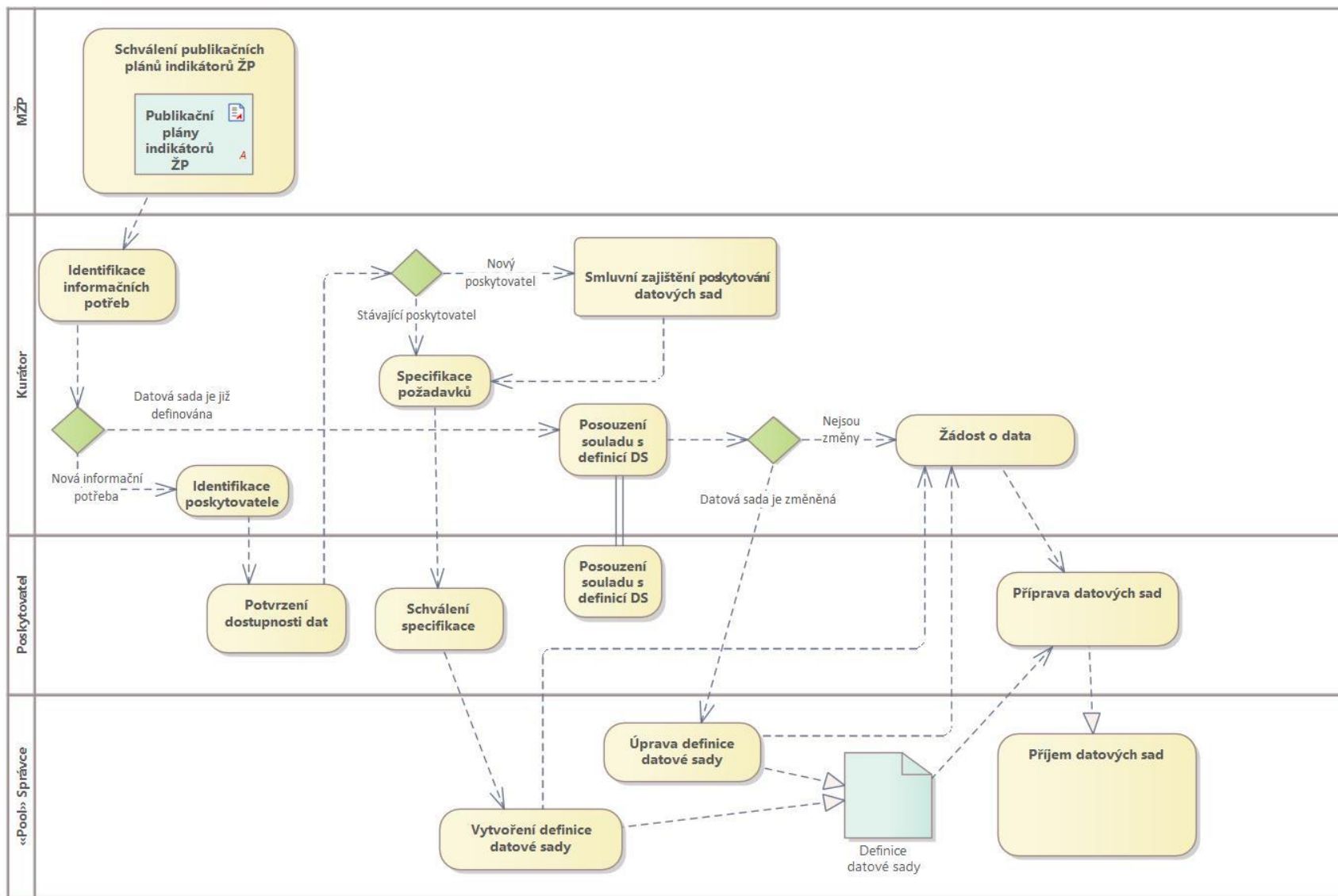
Ze **Schema** datové sady se odvozuje např.:

- Transformační skript
- Validací skript
- Importní skript

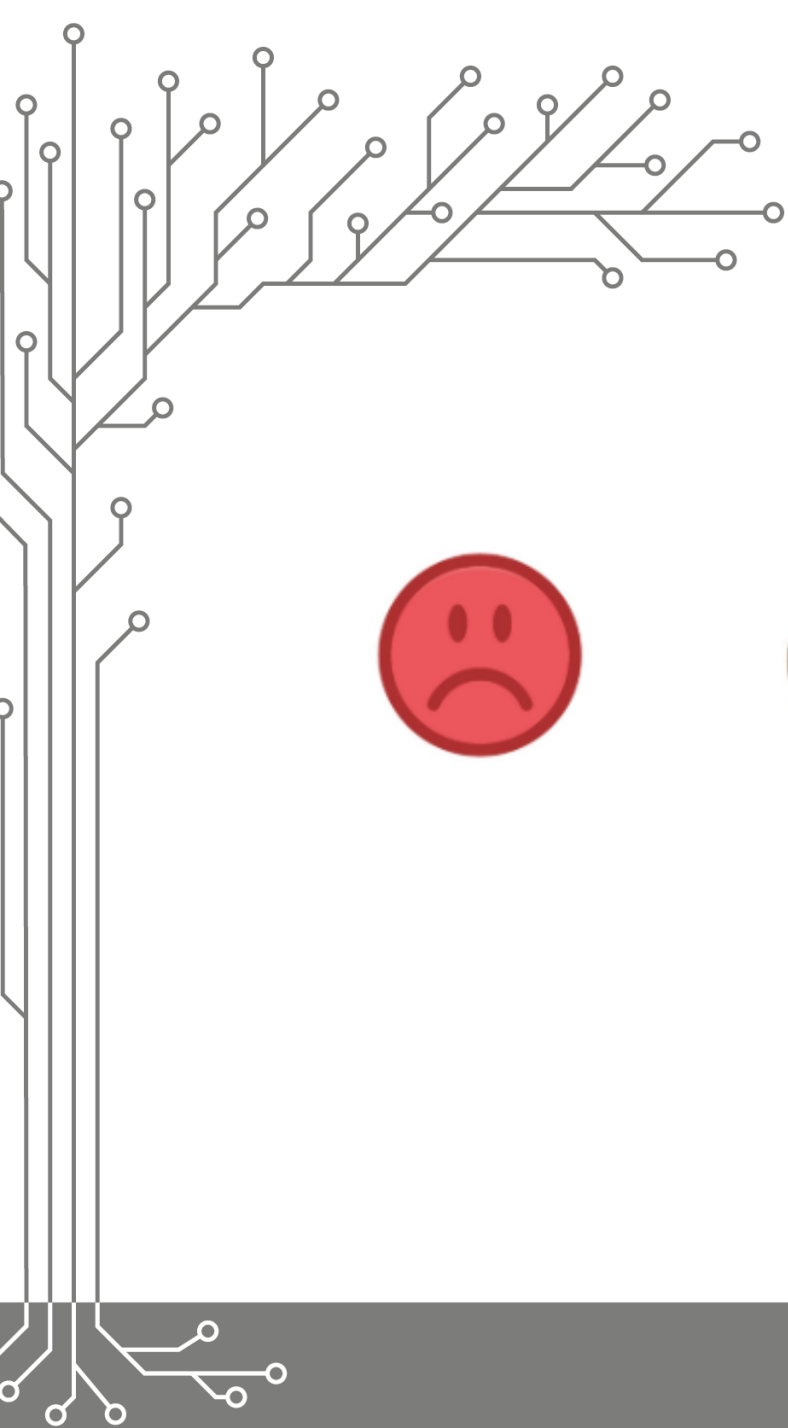


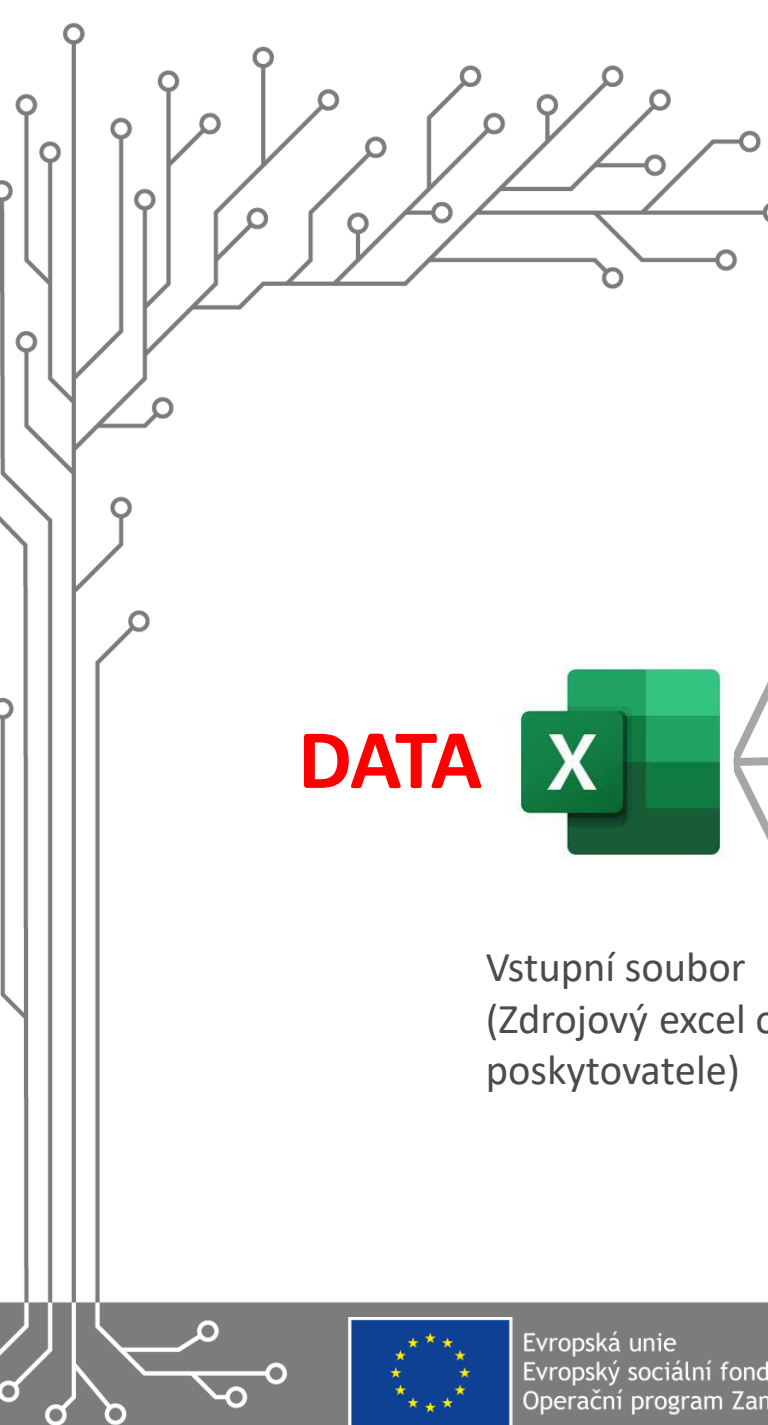


Čistota dat – standardizace, čištění, validace



Proces akvizice dat – TO BE





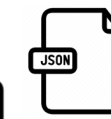
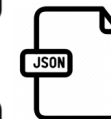
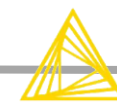
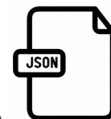
METADATA



DATA



Vstupní soubor
(Zdrojový excel od
poskytovatele)



Listy=
Atomizované (homogenní)
datové sady

Strojově zpracovatelné
datové sady



Evropská unie
Evropský sociální fond
Operační program Zaměstnanost





Důležité aspekty zpracování dat

- Otevřená rozhraní
- Číselníky a jednoznačnost dimenzí
 - (datum – vzniku?, aktualizace?, platnosti?, loadu?)
 - Referenční číselníky
 - Vazby
- Standardizace formátu na strojově zpracovatelné
- Čistota dat – standardizace, čištění, validace
- Zdroj pravdy



Metadata – data o datech

- **Popisují**, vysvětlují a lokalizují **data** a jejich zdroj
- **Zjednodušují** získávání, používání a **správu dat**

Descriptive metadata

Identifikace dat a důvod zpracování (**časové a geografické pokrytí**)

Structural metadata

Popis struktur a datových modelů (**schéma datové sady**)

Administrative metadata

Umožňují spravovat data (**podmínky užití dat, ...**)

Klíčový význam u dat z nesourodých zdrojů

*Pro zabezpečení **interoperability mezi systémy** a propojování dat, by metadata měla být mapována na **standardní model** (např. DCAT, Dublin Core).*

