

# Analýza zverejňovania datasetov ústredných orgánov štátnej správy na Portáli otvorených dát data.gov.sk k 30.06.2018

## Obsah:

<b>Analýza celkového počtu datasetov a porovnanie s minulým rokom</b>	<b>2</b>
<b>Formáty zverejňovaných datasetov</b>	<b>9</b>
<b>Licencie</b>	<b>12</b>
<b>Pravidelnosť aktualizácie:</b>	<b>13</b>
<b>Reálne aktualizácie</b>	<b>14</b>
<b>Dátum zverejnenia datasetu</b>	<b>15</b>
<b>Dátum zverejnenia dátového zdroja</b>	<b>16</b>
<b>Veľkosť datasetov</b>	<b>17</b>
<b>Záver a odporúčania</b>	<b>18</b>

## Analyza celkového počtu datasetov a porovnanie s minulým rokom

Na Portáli otvorených údajov data.gov.sk je k 30.06.2018 celkovo zverejnených 7208 datasetov, pričom však množstvo datasetov je zarátaných viackrát, nakoľko sú zverejnené vo viacerých formátoch (napr. aj xml aj csv). **Po odstránení týchto duplicit je celkovo zverejnených 2181 unikátnych datasetov.** Na výsledný počet datasetov mala okrem duplicit vo forme formátov vplyv aj metodika zarátavania datasetov, ktoré sú publikované opakovane, resp. denne, napríklad:

- **Úrad pre verejné obstarávanie** publikuje denne Vestník verejného obstarávania - štruktúrou opakujúce sa datasety boli zarátané iba 1x, čo znížilo počet unikátnych datasetov ÚVO: celkovo zverejnených bolo 1450, pričom zarátaných unikátnych je iba 16
- **Ministerstvo vnútra SR** publikovalo celkovo 2188 datasetov, z čoho však 1804 sú len zmenové dávky Registra adries

Organizácia	Absolútny počet datasetov	Počet unikátnych datasetov
Agentúra pre rozvoj vidieka, Nitra	4	4
Agroinštitút Nitra, štátny podnik	6	6
Generálna prokuratúra SR	8	8
Hydromeliorácie, štátny podnik	10	2
Lesopoľnohospodársky majetok Ulič, štátny podnik	1	1
Mesto Prešov	238	58
Mesto Trnava	5	5
Ministerstvo dopravy a výstavby SR	43	32
Ministerstvo financií SR	20	20
Ministerstvo hospodárstva SR	22	17
Ministerstvo kultúry Slovenskej republiky	37	35
Ministerstvo obrany SR	24	19
Ministerstvo pôdohospodárstva a rozvoja vidieka SR	189	111

Ministerstvo práce, sociálnych vecí a rodiny SR	5	5
Ministerstvo spravodlivosti SR	27	17
Ministerstvo školstva, vedy, výskumu a športu Slovenskej republiky	37	34
Ministerstvo vnútra SR	2188	253
Ministerstvo zahraničných vecí a európskych záležitostí SR	29	25
Ministerstvo zdravotníctva SR	59	24
Ministerstvo životného prostredia SR	46	41
Národná agentúra pre sieťové a elektronické služby - NASES	192	21
Národné lesnícke centrum	26	12
Národné poľnohospodárske a potravinárske centrum Lužianky	18	9
Národný bezpečnostný úrad	4	4
Nitriansky samosprávny kraj	5	4
Plemenárske služby Slovenskej republiky, štátny podnik	9	6
Pôdohospodárska platobná agentúra	2	2
Protimonopolný úrad SR	45	7
Regionálny úrad verejného zdravotníctva Banská Bystrica	12	11
Regionálny úrad verejného zdravotníctva Bardejov	33	9
Regionálny úrad verejného zdravotníctva Bratislava	11	5
Regionálny úrad verejného zdravotníctva Čadca	7	3
Regionálny úrad verejného zdravotníctva Galanta	6	6
Regionálny úrad verejného zdravotníctva Komárno	6	6
Regionálny úrad verejného zdravotníctva Levice	14	7
Regionálny úrad verejného zdravotníctva Liptovský	14	5

Mikuláš		
Regionálny úrad verejného zdravotníctva Lučenec	4	3
Regionálny úrad verejného zdravotníctva Michalovce	67	7
Regionálny úrad verejného zdravotníctva Nitra	15	12
Regionálny úrad verejného zdravotníctva Nové Zámky	6	6
Regionálny úrad verejného zdravotníctva Prievidza	9	2
Regionálny úrad verejného zdravotníctva so sídlom vo Veľkom Krtíši	10	4
Regionálny úrad verejného zdravotníctva Stará Ľubovňa	7	5
Regionálny úrad verejného zdravotníctva Svidník	22	8
Regionálny úrad verejného zdravotníctva Topoľčany	35	5
Regionálny úrad verejného zdravotníctva Trebišov	36	5
Regionálny úrad verejného zdravotníctva Trenčín	22	7
Regionálny úrad verejného zdravotníctva Vranov nad Topľou	19	5
Regionálny úrad verejného zdravotníctva Žiar nad Hronom	22	5
Regionálny úrad verejného zdravotníctva, Žilina	9	7
SK-NIC	1	1
Slovenská agentúra životného prostredia	2	2
Správa štátnych hmotných rezerv Slovenskej republiky	12	5
Štatistický úrad SR	1737	1101
Štátny Ústav pre kontrolu liečiv	5	3
Úrad geodézie, kartografie a katastra SR	2	2

Úrad jadrového dozoru Slovenskej republiky	7	6
Úrad podpredsedu vlády Slovenskej republiky pre investície a informatizáciu	53	14
Úrad pre normalizáciu, meterológiu a skúšobníctvo Slovenskej republiky	17	3
Úrad pre verejné obstarávanie	1450	16
Úrad priemyselného vlastníctva Slovenskej republiky	112	19
Úrad verejného zdravotníctva SR Bratislava	26	24
Úrad vlády SR	23	13
Ústredný kontrolný a skúšobný ústav poľnohospodársky v Bratislave	75	56
Všeobecná zdravotná poisťovňa	1	1
<b>spolu</b>	<b>7208</b>	<b>2181</b>

Do budúca je potrebné určiť metodiku zarátavania počtu datasetov, aby sa výsledné čísla takto extrémne nelíšili a aby sa nastavili pravidlá, čo je to individuálny dataset, ktorý má byť zarátaný a vykazovaný.

V (tým pádom nedokonalom) porovnaní voči minulému roku, keď bolo zverejnených 1034 datasetov je to nárast o 1147, t.j. o viac ako 110%. (nedokonalé, nakoľko v minulom roku bola odlišná metodika zarátavania počtu individuálnych datasetov).

Čo sa týka absolútneho počtu zverejnených datasetov, tak už tradične neohrozene na prvom mieste stojí Štatistický úrad SR s 1101 unikátnymi datasetmi, čo predstavuje viac ako 50% všetkých unikátnych datasetov. Výrazný nárast počtu datasetov si pripísali aj Ministerstvo vnútra SR a Ministerstvo pôdohospodárstva a rozvoja vidieka SR. **Veľmi proaktívne a v prostredí samospráv unikátne, sa k zverejňovaniu datasetov naďalej stavia mesto Prešov, ktoré sa v absolútnom poradí dostalo na celkovo štvrtú priečku**, pričom oproti predchádzajúcej analýze z roku 2017 celkový počet ich zverejnených datasetov vzrástol z 18 na 58.

Organizácia	Počet unikátnych datasetov
-------------	----------------------------

Štatistický úrad SR	1101
Ministerstvo vnútra SR	253
Ministerstvo pôdohospodárstva a rozvoja vidieka SR	111
Mesto Prešov	58
Ústredný kontrolný a skúšobný ústav poľnohospodársky v Bratislave	56
Ministerstvo životného prostredia SR	41
Ministerstvo kultúry Slovenskej republiky	35
Ministerstvo školstva, vedy, výskumu a športu Slovenskej republiky	34
Ministerstvo dopravy a výstavby SR	32
Ministerstvo zahraničných vecí a európskych záležitostí SR	25
Ministerstvo zdravotníctva SR	24
Úrad verejného zdravotníctva SR Bratislava	24
Národná agentúra pre sieťové a elektronické služby - NASES	21
Ministerstvo financií SR	20
Ministerstvo obrany SR	19
Úrad priemyselného vlastníctva Slovenskej republiky	19
Ministerstvo hospodárstva SR	17
Ministerstvo spravodlivosti SR	17
Úrad pre verejné obstarávanie	16
Úrad podpredsedu vlády Slovenskej republiky pre investície a informatizáciu	14
Úrad vlády SR	13
Národné lesnícke centrum	12
Regionálny úrad verejného zdravotníctva Nitra	12
Regionálny úrad verejného zdravotníctva Banská Bystrica	11

Národné poľnohospodárske a potravinárske centrum Lužianky	9
Regionálny úrad verejného zdravotníctva Bardejov	9
Generálna prokuratúra SR	8
Regionálny úrad verejného zdravotníctva Svidník	8
Protimonopolný úrad SR	7
Regionálny úrad verejného zdravotníctva Levice	7
Regionálny úrad verejného zdravotníctva Michalovce	7
Regionálny úrad verejného zdravotníctva Trenčín	7
Regionálny úrad verejného zdravotníctva, Žilina	7
Agroinštitút Nitra, štátny podnik	6
Plemenárske služby Slovenskej republiky, štátny podnik	6
Regionálny úrad verejného zdravotníctva Galanta	6
Regionálny úrad verejného zdravotníctva Komárno	6
Regionálny úrad verejného zdravotníctva Nové Zámky	6
Úrad jadrového dozoru Slovenskej republiky	6
Mesto Trnava	5
Ministerstvo práce, sociálnych vecí a rodiny SR	5
Regionálny úrad verejného zdravotníctva Bratislava	5
Regionálny úrad verejného zdravotníctva Liptovský Mikuláš	5
Regionálny úrad verejného zdravotníctva Stará Ľubovňa	5
Regionálny úrad verejného zdravotníctva Topoľčany	5
Regionálny úrad verejného zdravotníctva Trebišov	5
Regionálny úrad verejného zdravotníctva Vranov nad Topľou	5
Regionálny úrad verejného zdravotníctva Žiar nad Hronom	5
Správa štátnych hmotných rezerv Slovenskej republiky	5
Agentúra pre rozvoj vidieka, Nitra	4

Národný bezpečnostný úrad	4
Nitriansky samosprávny kraj	4
Regionálny úrad verejného zdravotníctva so sídlom vo Veľkom Krtíši	4
Regionálny úrad verejného zdravotníctva Čadca	3
Regionálny úrad verejného zdravotníctva Lučenec	3
Štátny Ústav pre kontrolu liečiv	3
Úrad pre normalizáciu, meteorológiu a skúšobníctvo Slovenskej republiky	3
Hydromeliorácie, štátny podnik	2
Pôdohospodárska platobná agentúra	2
Regionálny úrad verejného zdravotníctva Prievidza	2
Slovenská agentúra životného prostredia	2
Úrad geodézie, kartografie a katastra SR	2
Lesopôľnohospodársky majetok Ulič, štátny podnik	1
SK-NIC	1
Všeobecná zdravotná poisťovňa	1
<b>spolu</b>	<b>2181</b>

Odlíšny pohľad na počet zverejnených datasetov ponúka samotný Portál otvorených dát, ktorý informuje o celkovo 1483 datasetoch, pričom aj počty zverejnených datasetov sa pri jednotlivých organizáciách líšia od našej analýzy:





## Formáty zverejňovaných datasetov

Ďalšie zaujímavé porovnanie je z pohľadu toho, v akých formátoch sú datasety zverejnené. Množstvo organizácií zverejňuje datasety vo viacerých formátoch, čo na jednej strane uľahčuje ich otvorenie a používanie pre širšiu skupinu užívateľov, na druhú stranu spôsobuje komplikácie pri určení, či ide o unikátny dataset, alebo ten istý, len v inom formáte.

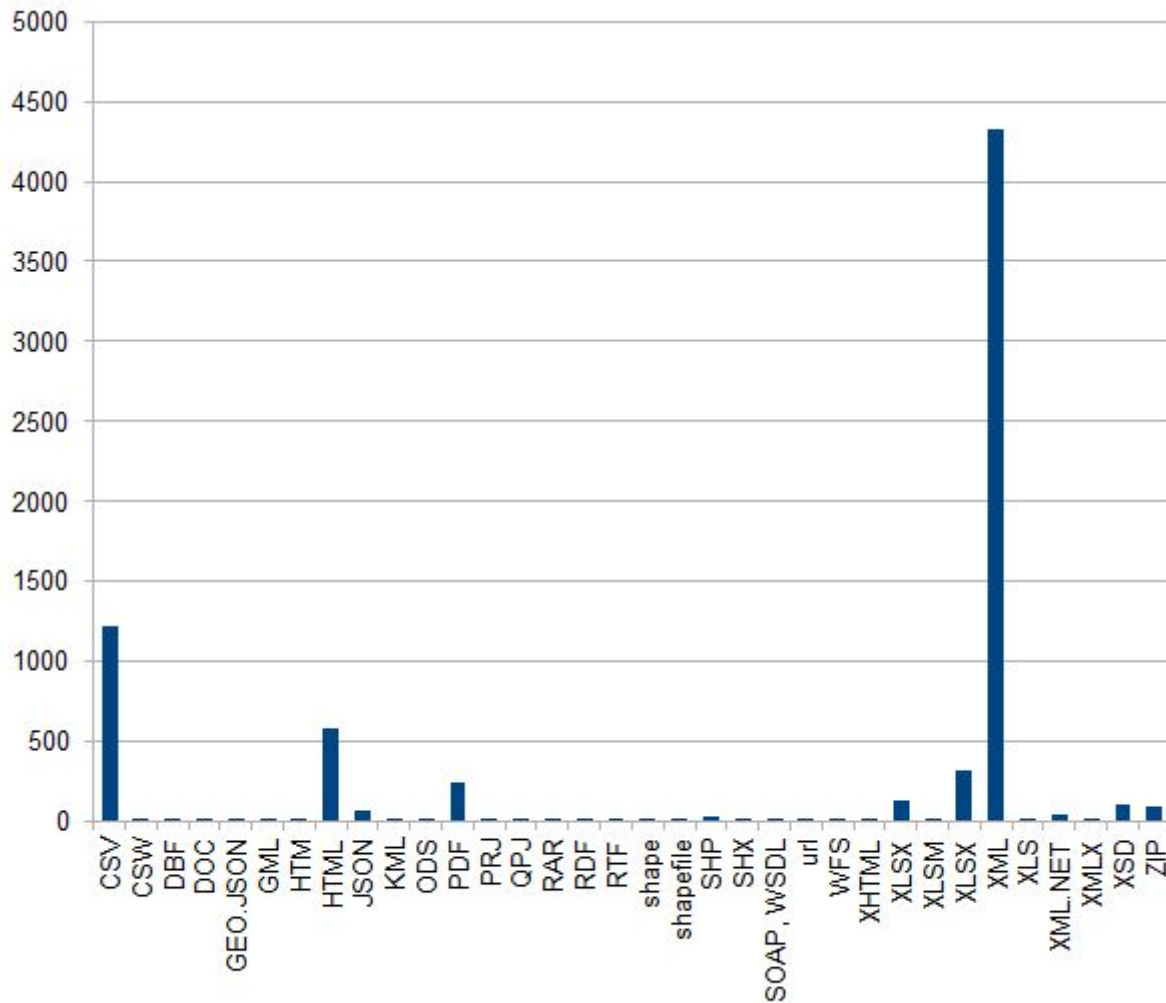
Organizácie najčastejšie zverejňujú svoje datasety v XML, CSV a HTML, ktoré sú aj užívateľsky najlepšie použiteľné aj pre dátových “neodborníkov”.

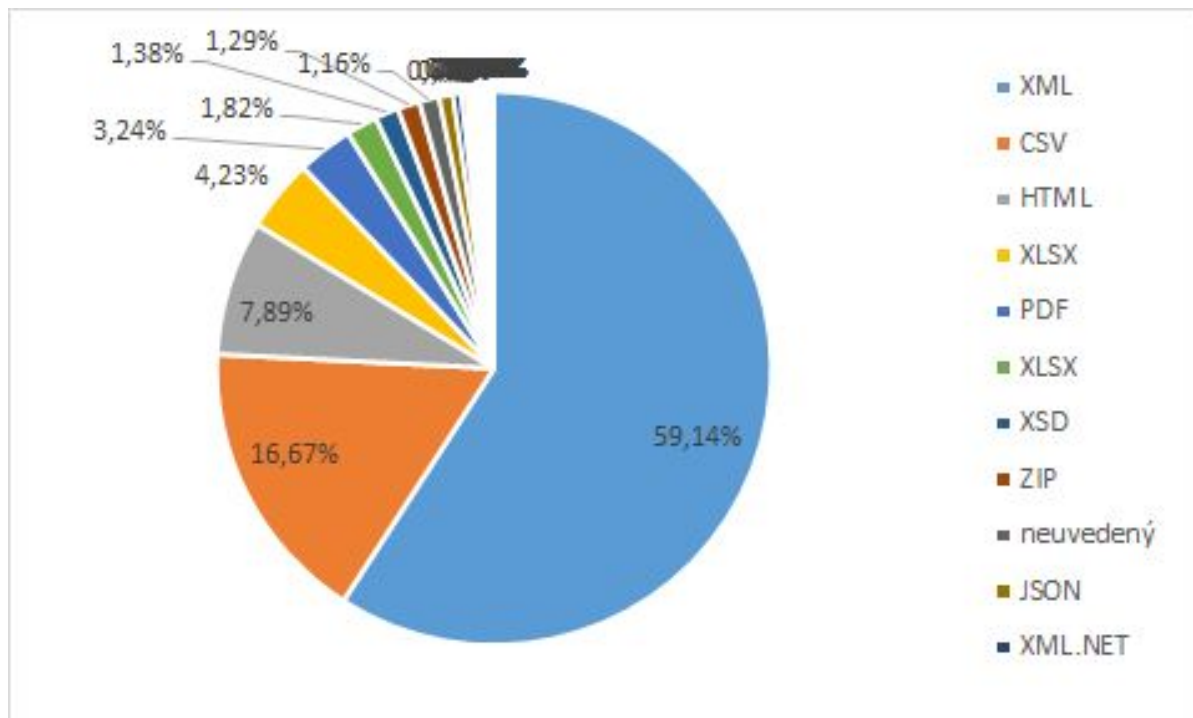
Množstvo datasetov je však stále uvádzaných v “open data unfriendly” formátoch, ako napríklad PDF. V týchto prípadoch sa bijú 2 snahy - snaha o zverejňovanie čo najväčšieho počtu datasetov voči snahe o čo ich najvyššiu kvalitu (“hviezdičková” konvencia <https://5stardata.info/cs/> resp. <http://opendatahandbook.org/glossary/en/terms/five-stars-of-open-data/> ).

XML	4320	59,14%
CSV	1218	16,67%
HTML	576	7,89%
XLSX	309	4,23%
PDF	237	3,24%
XLSX	133	1,82%
XSD	101	1,38%
ZIP	94	1,29%

neuvedený	85	1,16%
JSON	59	0,81%
XML.NET	35	0,48%
SHP	24	0,33%
GEO.JSON	20	0,27%
KML	20	0,27%
url	14	0,19%
RTF	13	0,18%
ODS	6	0,08%
RDF	6	0,08%
GML	5	0,07%
shapefile	4	0,05%
WFS	4	0,05%
CSW	3	0,04%
SOAP, WSDL	3	0,04%
DOC	2	0,03%
HTM	2	0,03%
RAR	2	0,03%
XHTML	2	0,03%
DBF	1	0,01%
PRJ	1	0,01%
QPJ	1	0,01%
shape	1	0,01%
SHX	1	0,01%
XLSM	1	0,01%
XLS	1	0,01%

XMLX	1	0,01%
<b>spolu</b>	<b>7305</b>	



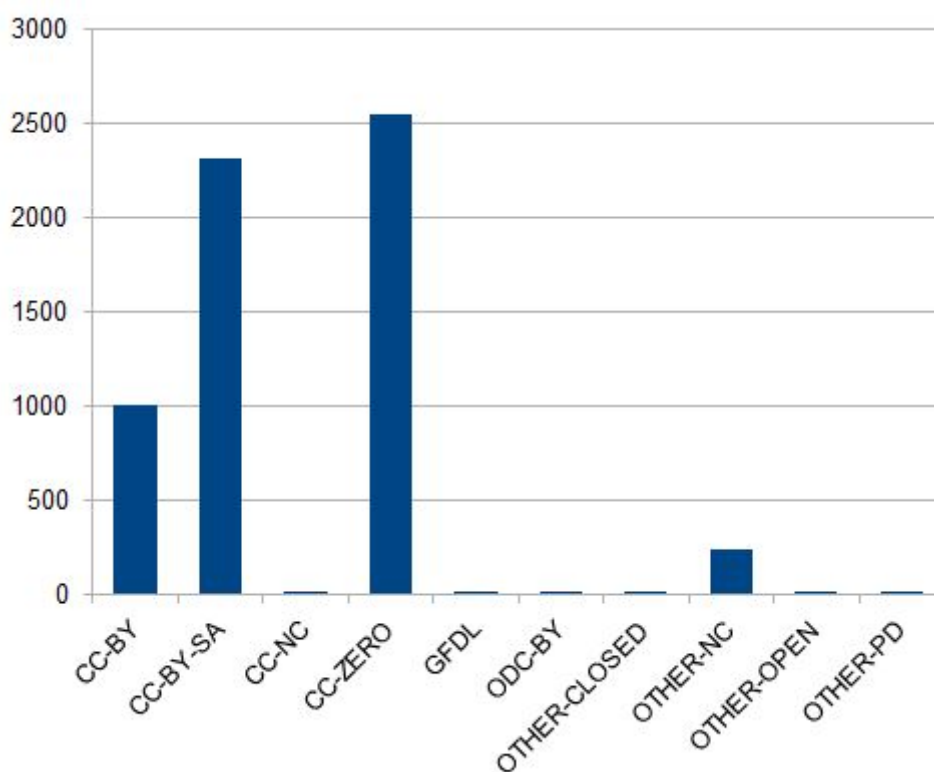


## Licencie

Na to, aby datasety spĺňali požiadavky na “open data” - otvorené údaje, musia okrem vhodného formátu mať uvedenú aj licenciu, čo sa s týmito dátami môže robiť, t.j. ako sa s nimi môže ďalej nakladať. V prostredí Open data by tieto licencie mali byť čo najširšie (bez obmedzenia; resp. aspoň s uvedením zdroja). Na Slovensku nie je stanovená licencia na zverejňovanie údajov verejnej správy a tak si organizácie môžu zvoliť inú, už existujúcu, alebo si vytvoriť vlastnú. Najčastejšie sa používajú medzinárodné licencie Creative Commons (<https://sk.creativecommons.org/>) a to CC-ZERO a CC-BY-SA. Je zaujímavé, že Úrad pre verejné obstarávanie, ktorý zverejňuje denne Vestník verejného obstarávania, k týmto vestníkom nepridal žiadnu licenciu...

CC-ZERO	2546	35,32%
CC-BY-SA	2318	32,16%
(prázdne pole)	1053	14,61%
CC-BY	1001	13,89%
OTHER-NC	233	3,23%
OTHER-PD	16	0,22%
CC-NC	13	0,18%

GFDL	8	0,11%
OTHER-OPEN	8	0,11%
nešpecifikovaná	8	0,11%
ODC-BY	2	0,03%
OTHER-CLOSED	2	0,03%
spolu	7208	

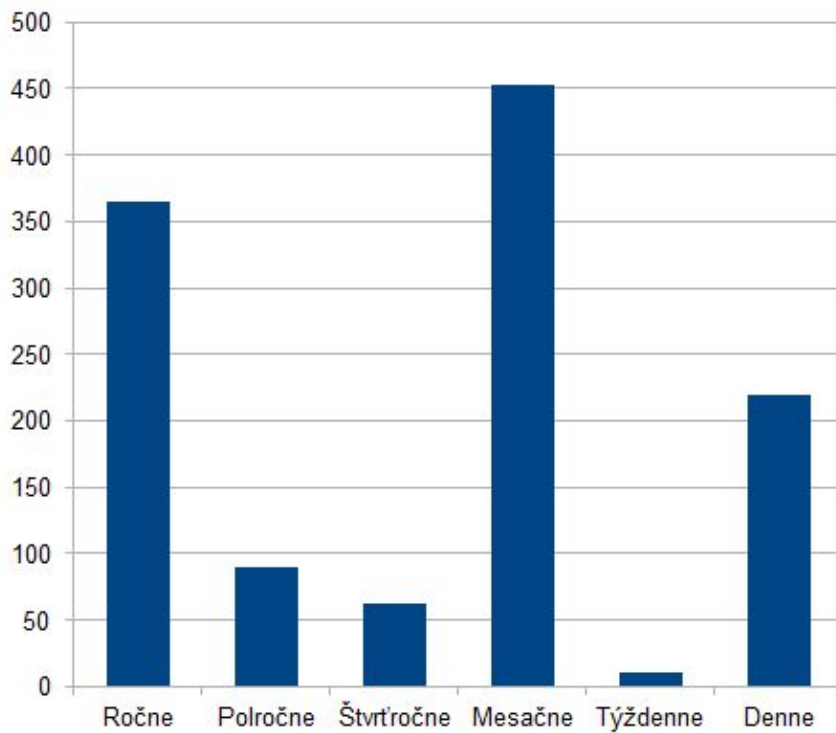


### Pravidelnosť aktualizácie:

Pre použiteľnosť zverejnených datasetov je ďalšou kľúčovou informáciou, či a ako často bude dataset aktualizovaný. Tvorcovia aplikácií využívajúcich zverejnené datasety sa tak môžu pripraviť a svoje aplikácie tvoriť tak, aby používali vždy najaktuálnejšie informácie.

Zo 7208 všetkých datasetov malo pravidelnosť aktualizácie uvedených iba 1200, ostatné buď nie sú aktualizované (jednorázovo zverejnené datasety), alebo túto aktualizáciu neuvádzajú.

<b>Ročne</b>	365
<b>Polročne</b>	90
<b>Štvrťročne</b>	62
<b>Mesačne</b>	453
<b>Týždenne</b>	10
<b>Denne</b>	220
nepravidelne	358
neuvedené	5650
<b>spolu</b>	<b>7208</b>



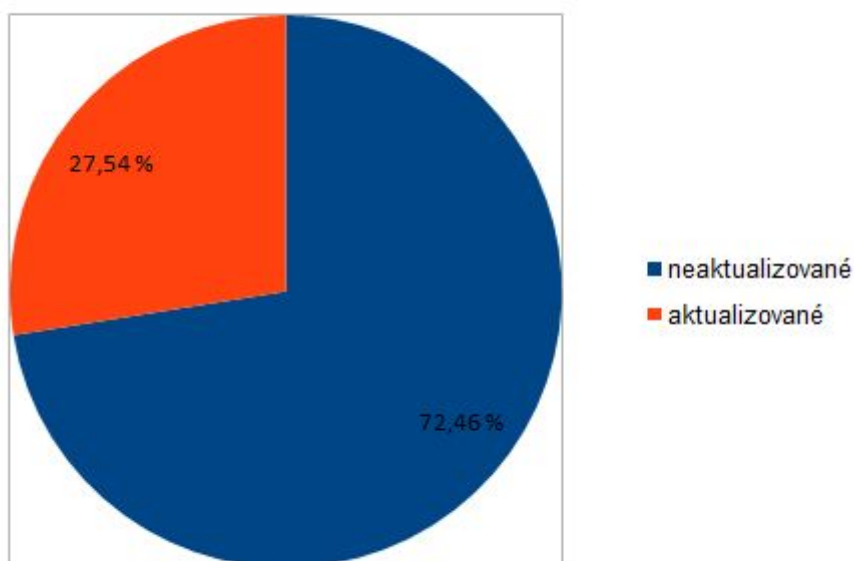
## Reálne aktualizácie

Údaj o pravidelnosti aktualizácie má zverejnených 1200 datasetov, koľko ich však bolo reálne, aspoň raz aktualizovaných?

Na výsledky však má opäť vplyv metodika zarátania počtu datasetov (niektoré organizácie aktualizujú ten istý dataset, niektoré vytvárajú stále nový, resp. publikujú nové zmenové dávky. Do počtu aktualizovaných datasetov môžu byť dokonca zarátané aj opravy chybných údajov v datasetoch).

Celkovo však má aspoň 1 aktualizáciu uvedených 1985 datasetov.

<b>neaktualizované</b>	5223	72,46 %
<b>aktualizované</b>	1985	27,54 %
<b>spolu:</b>	<b>7208</b>	

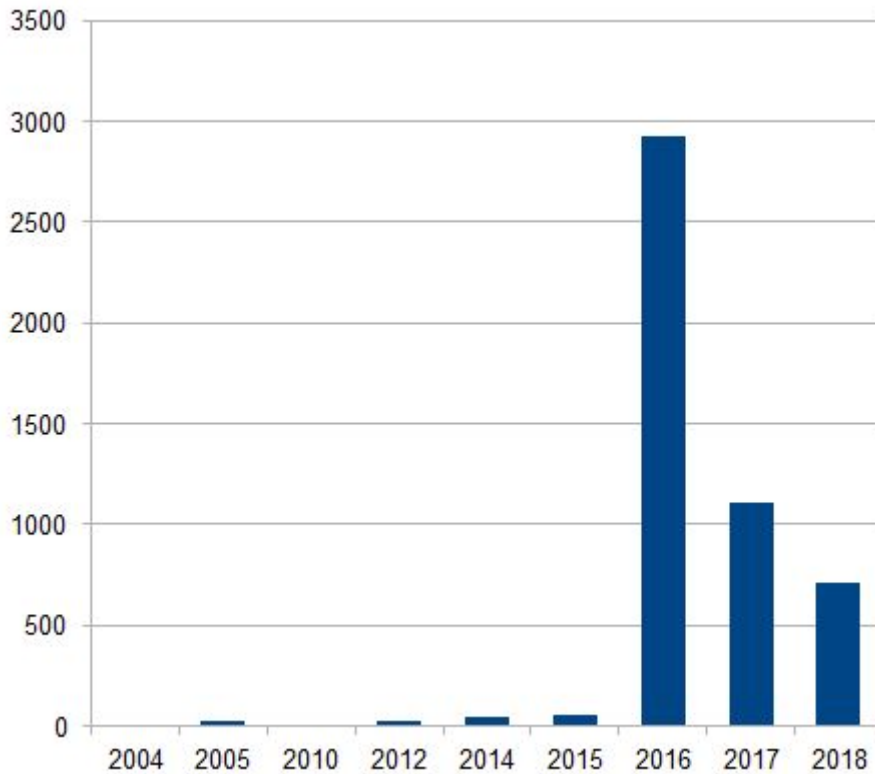


## Dátum zverejnenia datasetu

Pri sledovaní vývoja počtu datasetov je zaujímavé posúdiť, kedy boli zverejnené datasety, aby bolo možné analyzovať, či trend v počte zverejňovaných datasetov má rastúcu, alebo stagnujúcu tendenciu. Ako z uvedenej tabuľky vyplýva, najviac datasetov má uvedené, že boli zverejnené v roku 2016.

rok neuvedený	2303
2004	3
2005	26
2010	5
2012	28
2014	44
2015	55
2016	2931
2017	1106
2018	707
<b>spolu</b>	<b>7208</b>



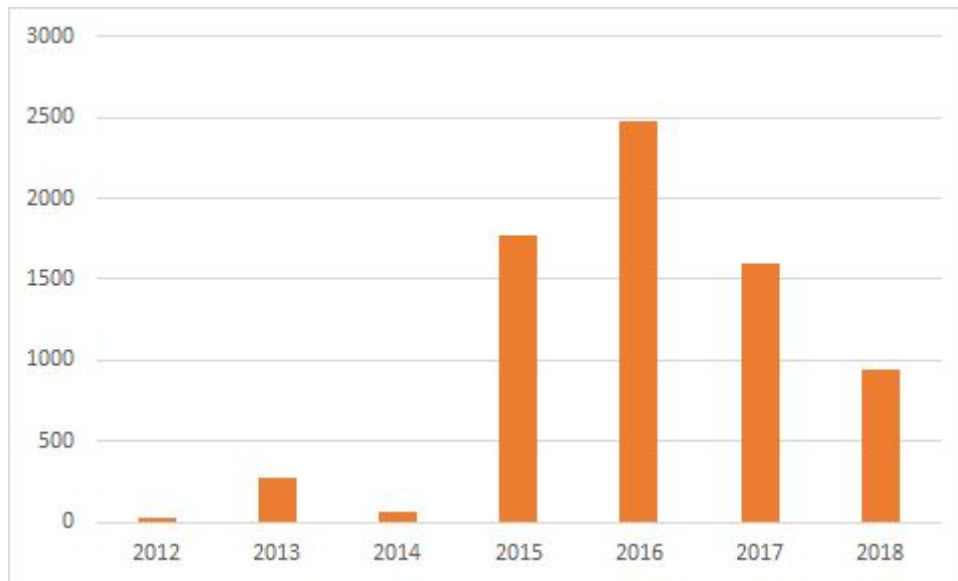


## Dátum zverejnenia dátového zdroja

Okrem dátumu zverejnenia je zaujímavé posúdiť, kedy bol zverejnený samotný dátový zdroj (t.j. nie kedy bol zverejnený dataset na data.gov.sk). Tento údaj v podstate poukazuje na "vek" dátového zdroja.

0-zlý formát	62
2012	23
2013	274
2014	63
2015	1775
2016	2476
2017	1597
2018	938

<b>spolu</b>	<b>7208</b>
--------------	-------------

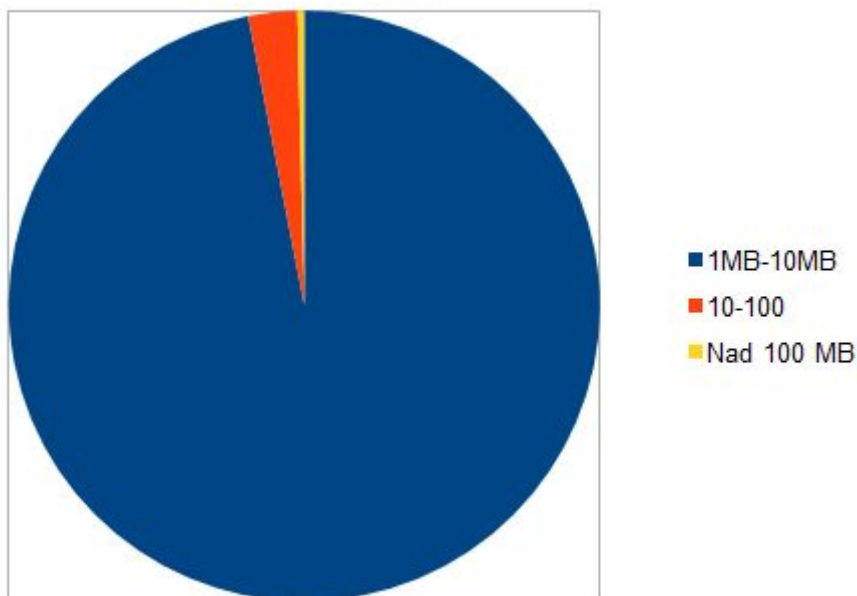


V roku 2017 sa rekordný rok 2016 nepodarilo prekonať, čo je možné pripočítať aj predpokladu, že v rokoch 2015-2016, keď došlo k obrovskému nárastu počtu zverejňovaných datasetov, zverejnili organizácie všetky "jednoduché" a rýchlo zverejniteľné datasety, ktoré mali k dispozícii. Následné vytváranie nových datasetov už nemusí byť také ľahké ani rýchle.

## Veľkosť datasetov

Niektoré organizácie sa obávajú, že ich obrovské datasety (rádovo v stovkách MB až GB) pri veľkom záujme o ich sťahovanie a používanie spôsobia zahltenie ich serverov, resp. úložísk dát. Portál otvorených dát ponúka pre tieto organizácie, aby si svoje datasety nahrali priamo na Portál otvorených dát a tak celú "traffic" presmerovali na portál na to určený. Niektoré organizácie to vyriešili tak, že svoje datasety zverejňujú po častiach, čo opäť vplyva na počet zverejnených datasetov. Z datasetov, ktoré majú uvedenú veľkosť, je najväčším datasetom Register adries - Register vchodov (orientačných čísiel) Ministerstva vnútra SR s 899 MB, pričom oficiálny maximálny limit na data.gov.sk predstavuje 1GB.

<b>1MB-10MB</b>	4319
<b>10-100</b>	119
<b>Nad 100 MB</b>	17
<b>neuvedené</b>	2753
<b>spolu</b>	<b>7208</b>



## Závery a odporúčania

1. Celkový počet datasetov výrazne rastie, chýbajú však informácie o ich využívanosti, čiže či ich zverejňovanie má význam a sú ďalej používané, t.j. posúdenie možného potenciálu konkrétneho datasetu. Na data.gov.sk sú zaregistrované iba 4 aplikácie, ktoré využívajú zverejnené open data: <https://data.gov.sk/related> . Ako spätnú väzbu pre zverejňujúce organizácie by bolo vhodné pridať počítadlo prístupov/stiahnutí ku všetkým datasetom na data.gov.sk, keďže vytvárané aplikácie zjavne nie sú zaregistrované, tak aby sa životaschopnosť zverejňovania datasetov preukázala aj počtom ich siahnutí. V poslednom období sa organizuje množstvo hackatonov, na ktorých sú vytvárané zaujímavé a kreatívne aplikácie využívajúce open data - tieto však často nie sú "dotiahnuté" a následne aj chýba motivácia autorov na ich registráciu na data.gov.sk
2. Zverejnené popisy/metadáta/atribúty datasetov často chýbajú, alebo sú chybne uvedené, NASES ako správca data.gov.sk by mal vyzvať organizácie, ktoré datasety zverejnili, aby doplnili, resp. opravili informácie. Identifikácia zjavných chýb a chýbajúcich dát je v celkovom exporte z data.gov.sk: [...LINK...](#)
3. S cieľom umožniť automatizované a presnejšie kvalitatívne posúdenie obsahu datasetov zaviesť v popise datasetu aj kategóriu "oblasť" resp. "klaster", pričom je možné využiť napr. 13 primárnych klastrov (<https://index.okfn.org/methodology/> ), ktoré sú využívané pri zostavovaní Global Open Data Index (<https://index.okfn.org/> ), aby bolo možné aj medzinárodné porovnanie.

4. Vytvoriť metodiku zaratavania unikátnych datasetov, čo je dôležité aj z pohľadu napĺňania ukazovateľov strategických dokumentov - Operačný program Integrovaná infraštruktúra, Národná koncepcia informatizácie verejnej správy a ďalších. Využiť by bolo možné napr. Open Data Charter Measurement Guide <https://drive.google.com/file/d/1yNOPMP1rO68l4Swg16zqD8aTWDgFml-/view>
5. Vytvoriť odporúčania a popis pre organizácie, aké sú preferované licencie k datasetom verejnej správy a legislatívny popis toho, aký dopad má ich použitie (napr. množstvo organizácií sa bojí použiť otvorené licencie z obavy, že ich datasety budú následne modifikované, čo by v prípade neskorších sporov predstavovalo reputačné riziko pre nich ako pôvodných autorov datasetov). Aktuálne tvorený Zákon o údajoch v gescii Úradu podpredsedu vlády SR pre investície a informatizáciu by mohol túto otázku vyriešiť.