

## OTEVÍRÁNÍ GEODAT - PŘÍPAD IPR PRAHA

JÁCHYM ČEPICKÝ

<sup>1</sup> Otevřená GeoInfrastruktura, Česká republika  
*jachym.cepicky@opengeolabs.cz*

### Abstrakt

Na základě nově schválených Pravidel pro poskytování dat a výstupů z datových souborů z datového skladu Geografického informačního systému hl. m. Prahy zajišťovaným IPR Praha, která nabudou účinnosti od 1. 5. 2015, bude možné poskytování geodat formou otevřených dat a otevřených služeb. IPR Praha ve spolupráci se sdružením Otevřená GeoInfrastruktura a Fondem Otakara Motejla sestavili dotazník pro potenciální uživatele těchto dat, abychom jim při otevírání dat vyšli co nejvíce vstříc a podpořili tak jejich další široké využití. Dotazník se obecně ptá na pozadí respondentů, zda-li přicházejí z oboru GIS či jiného, s jakými typy programů nejčastěji pracují (desktop vs. server nebo databáze či web), na jakých platformách se pohybují (např. otevřený software, proprietární), v jaké formě data nejraději požadují (surová či předzpracovaná) a jakou formu distribuce požadují (webové služby OGC či předgenerované soubory). Výsledkem bude, jak věříme, jací uživatelé pro data přicházejí a jaké je jejich technické zázemí a odborné pozadí. Také na základě dat z dotazníku vzniká analýza otevírání (geo)dat nejen pro IPR Praha, která popisuje možné způsoby otevírání geodat, jejich vhodné formáty pro různé případy použití a "best practices" při otevírání geodat obecně.

### Abstract

Based on new rules for distribution of data files from data warehouse of Geographical information system of the city of Prague, it will be possible to distribute geodata as open data, using open services. Institute of Planning and Development of Prague, together with Open GeoInfrastructure Association and Otakar Motejl's fund put together a questionnaire for potential users of those data, in order to be able to be able to fulfill their needs and support of data usage as much as possible. Questionnaire is asking on the background of the users, on their technological background, whether they come from GIS field or from other area, whether they work rather with desktop, database or server programs. It also tries to find out, which form of the data is preferred (raw versus postprocessed, ...), which types of services are required and so on. Also based on the results of the questionnaire, new analysis of opening of (geo)data is to be prepared, which should be usable not only for the city of Prague, but also for other bigger public (geo)data providers.

**Klíčová slova:** otevřená data; dotazník; evaluace; otevřené formáty; IPR Praha

**Keywords:** open data, questionnaire; evaluation; open formats; Prague Institute of Planning and Development

### ÚVOD

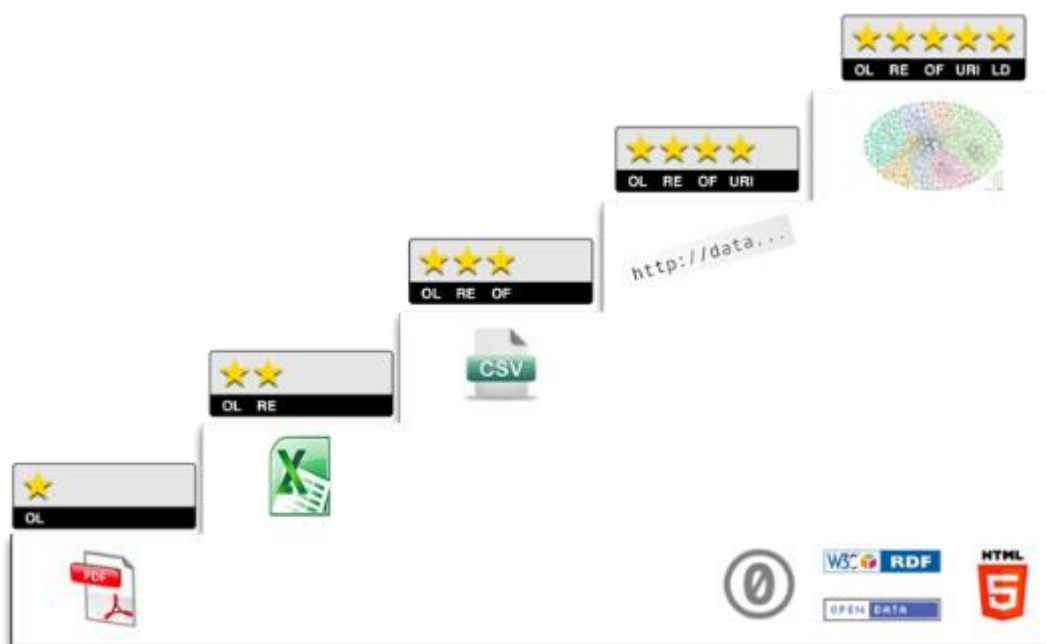
Je všeobecně uznávaným faktem, že sebelepší řešení není udržitelné, pokud není naplněno kvalitními a aktuálními daty. V poslední době dochází v této souvislosti v čím dál větší míře ke sdílení a šíření dat prostřednictvím sítě Internet. Pojmy jako "linked data" nebo "sémantický web" nejsou dávno jen teoretickým konceptem, ale aktuální skutečností. Bez nadsázky lze říci, že data jsou jedna z komodit, které podporují jak technický pokrok, tak rozvoj občanské společnosti. Tento fakt začíná být po vzoru našich nejbližších sousedů, především Německa, ale i Rakouska či Velké Británie reflektován i veřejnou správou v České republice v podobě postupného otevírání dat.

Aby data mohla být označena za otevřená, musí splňovat několik legislativních a technických podmínek (viz níže). Tyto podmínky nejsou v čase konstantní, vyvíjí se spolu s technickými prostředky, standardy ale i legislativou či postoji společnosti.

V současnosti těmito podmínkami myslíme zejména:

- Data jsou dostupná na síti Internet a dohledatelná běžnými nástroji.
- Data jsou čitelná, tedy v textovém či binárním formátu.
- Data musí být úplná, tj. jsou zveřejněná v takovém rozsahu, aby nechyběly některé fragmenty umožňující jejich využití. Data by zároveň měla být zveřejněna v maximálním možném rozsahu.
- Legislativní a technické překážky pro využití dat jsou minimální.

Otevírání dat přináší kromě nákladů na vlastní otevření a provoz také nemalé přínosy jak pro poskytovatele, tak pro uživatele. Tim Berners-Lee [1] sestavil tzv. pětihvězdičkový systém hodnocení otevřenosti dat, jehož smyslem je umožnit jednoduchou orientaci ve stupni otevřenosti datových sad. Čím výše se datová sada dostane, tím větší je možnost jejího využití veřejností.



Obr. 1: Pětihvězdičkový systém otevřených dat podle Tim Berners-Lee [1]

### 5ti hvězdičkový systém

#### ★ Dostupná data

Data jsou zveřejněna na Internetu, nezáleží na formátu. Data, která jsou zveřejněna na síti Internet v jakémkoli formátu, ale pod otevřenou licencí pro jejich využití. Existují tedy určitá technická omezení pro jejich využití, nicméně legislativní omezení jsou odstraněna. Příkladem může být mapové dílo ve formátu PDF.

#### ★★ Strukturovaná data

Data jsou zveřejněna ve strojově čitelném formátu. Data jsou zveřejněna ve strojově zpracovatelném formátu, který není otevřený. Příkladem mohou být data ve formátu Microsoft Excel nebo data uložená v Esri GeoDatabase.

**★★★ Data v otevřeném formátu**

Data jsou zveřejněna v otevřeném formátu. Strojový formát, ve kterém jsou data zveřejněna, je otevřený, tj. veřejně publikovaný s licencí umožňující jeho využití. Příkladem jsou například formáty Geospatial Markup Language (GML), OGC GeoPackage nebo Esri Shapefile

**★★★★ Data s identifikátorem**

Data jsou dohledatelná, mají unikátní identifikátor. Data jsou opatřena identifikátorem Uniform Resource Identifier (URI), unikátním v rámci celé sítě Internet. Jednotlivé prvky datové sady by měly být identifikované pomocí URI ve tvaru použitelném pro protokol HTTP tzn. v podobě adresy URL. Tím se zajistí, že uživatelé mohou daný prvek kdykoli najít.

**★★★★★ Provázaná data**

Data jsou navzájem prolinkovaná, lze mezi nimi navigovat, "surfovat" [ref71]. Data jsou nejen identifikovatelná pomocí URI, ale obsahují odkazy na další datové sady. Stejně jako datové sady jsou pomocí odkazů provázány i jednotlivé prvky z datových sad. Data jsou dále opatřena popisnými informacemi (metadaty) tak, aby v nich bylo možno jednoduše vyhledávat.

**OTEVÍRÁNÍ DAT**

Proces otevírání dat v sobě zahrnuje zejména následující kroky [2]:

- Zvolení souborového formátu (GeoTIFF, Esri Shapefile, GeoDatabáze, atd.)
- Zvolení způsobu distribuce (např. Server FTP, webové služby, CD k vyzvednutí osobně na vrátnici úřadu a podobně)
- Zvolení licence databáze
- Způsob aktualizace
- A další

Jeden z rozhodujících faktorů je cílová skupina uživatelů otevřených dat. Jedná se o odborné pracovníky v daném oboru? Pracovníky GIS? Datové analytiky bez specifického zaměření? Programátory bez hlubší znalosti problematiky? Jaký používají software, ze kterého budou k datům přistupovat? Tyto a další otázky si je potřeba položit – nalézt odpověď nemusí být zcela jednoznačné.

**SPECIFIKA GEODAT**

Na rozdíl od běžných datových sad, jsou geodata svým způsobem specifická oblast a to zejména z těchto důvodů:

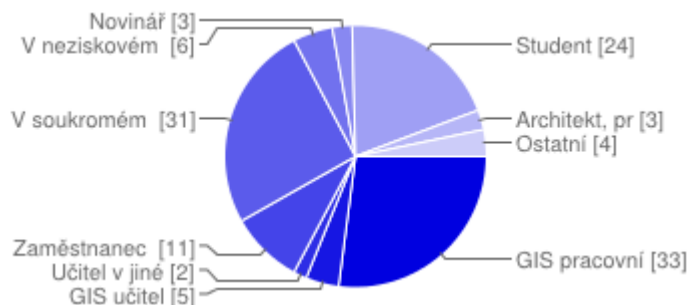
- Existuje nezávislá uznávaná standardizační organizace, jejích normy jsou akceptovány jak výrobci, tak uživateli software (Open Geospatial Consortium OGC).
- Existují všeobecně uznávané otevřené formáty datových sad
- Existují standardizované a využívané způsoby distribuce geodat pomocí webových služeb (opět standardy OGC)
- Díky INSPIRE jsou tyto normy dále využívány

**PRŮZKUM IPR**

Jedním ze způsobů, jak zjistit preference potenciálních uživatelů otevřených dat je dotazníkové šetření. Ve spolupráci s IPR Praha, Fondem Otakara Motejla a spolkem Otevřená GeoInfrastruktura jsme se pokusili takový dotazník sestavit. Jeho dílčí výsledky bychom rádi představili v tomto příspěvku. Dotazník je možné spustit z adresy <http://www.geoportalpraha.cz/cs/clanek/260/pomozte-otevrit-prazska-geograficka-data> k

prosinci 2014 dotazník vyplnilo 120 respondentů. Níže uvedeme a částečně okomentujeme výsledky některých odpovědí [3].

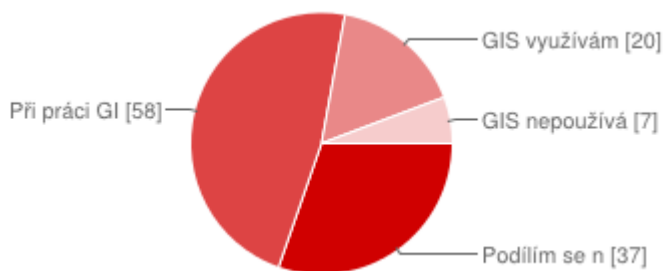
### Uživatelé



Obr. 2: Odpovědi na otázku, "Pracujete jako...?"

Převažující skupinou respondentů jsou „Pracovníci GIS“ (27%), následovaná „Pracovníky v soukromém sektoru“ (25%) a „Studenty“ (20%). Ostatní skupiny byly spíše minoritní.

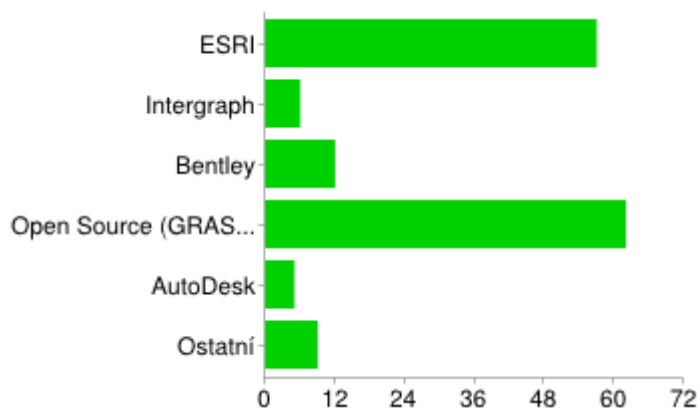
### Práce s GeoDaty



Obr. 3: Jakým způsobem pracujete s geodatami, jaký je váš vztah ke GIS?

Na otázku „Jakým způsobem pracujete s geodatami?“ odpověděla skoro polovina respondentů (48%), že při své práci GIS aktivně využívají (jako zdroj dat, používají analýzy, tvoří výstupy a podobně). Pouhých 6% respondentů odpovědělo, že GIS ve své praxi prakticky nevyužívá.

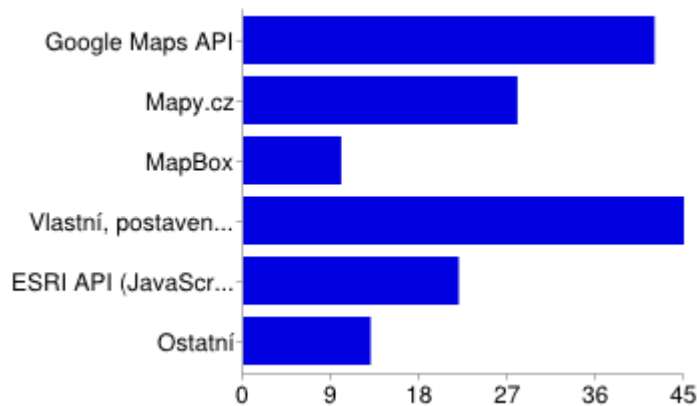
### GIS na desktopu



Obr. 4: Desktopová platforma

70% dotázaných odpovědělo, že ve své práci využívá desktopové programy pro tvorbu GIS. Nepřekvapí, že významného zastoupení na desktopech našich uživatelů dosáhla platforma firmy ESRI (47%). Významnějšího podílu však dosáhla skupina „open source“ programů (QGIS, GRASS GIS, Jump a podobně) s 51%.

### Web GIS

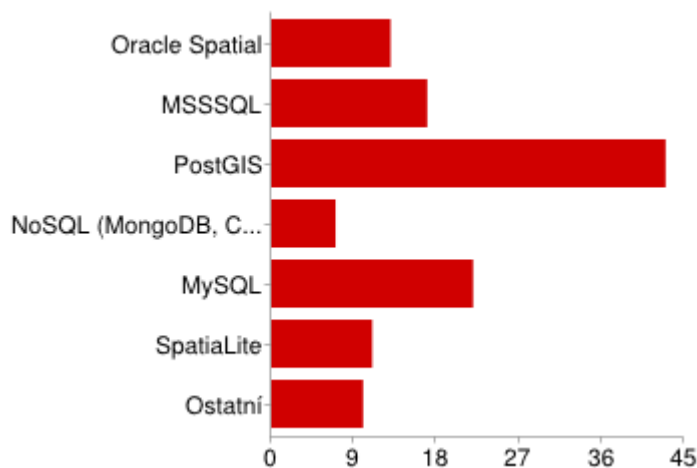


Obr. 5: Webová platforma

66% dotázaných využívá některou z webových platform pro tvorbu mapových aplikací. Nejčastější platforma je „vlastní“, postavené na některé z open source knihoven (OpenLayers, Leaflet) (37%), a Google Maps API (34%).

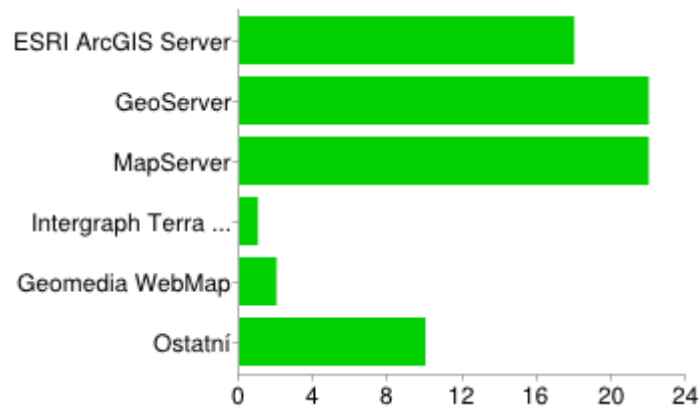
### Prostorové databáze

66% účastníků průzkumu uvedlo, že prostorové databáze nějakým způsobem využívá.

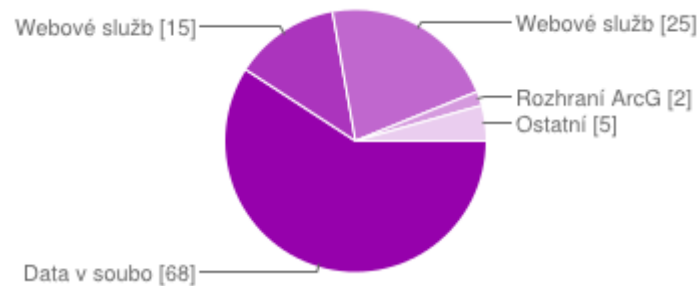


Obr. 6: Prostorové databáze

Nejpoužívanější prostorovou databází je s odstupem PostGIS (35%). NoSQL databáze zatím nejsou v našich končinách příliš využívány.

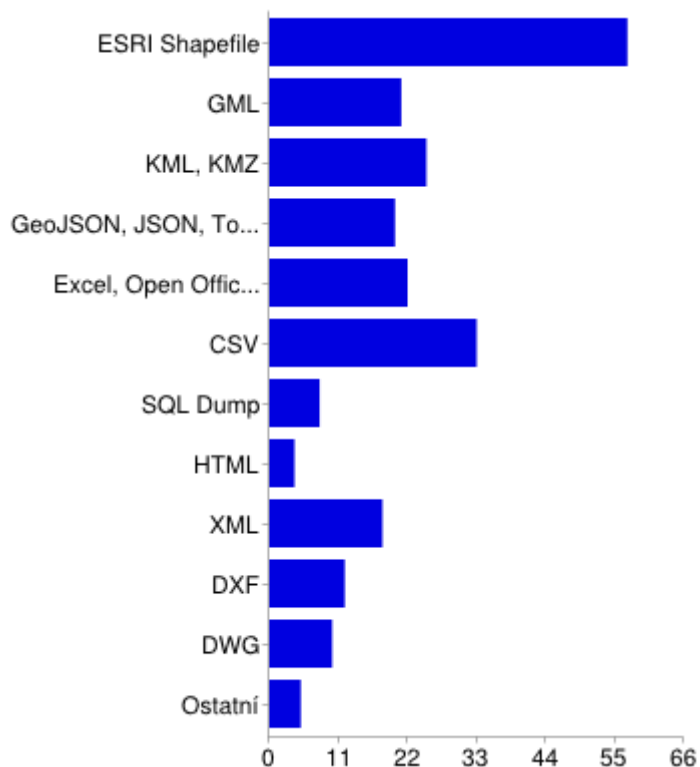
**Mapové servery****Obr. 7:** Mapové servery

41% respondentů uvedlo, že využívá mapové servery. Nejpoužívanější software jsou GeoServer a MapServer (oba shodně 18%), následovaný ArcGIS Serverem (15%).

**Datové zdroje****Obr. 8:** Způsob distribuce otevřený geodat

Na otázku, jakým způsobem uživatelé preferují dostávat otevřená geodata, většina odpověděla (56%), že prostřednictvím souborů. Druhým nejčastějším „přáním“ způsobu distribuce jsou webové služby OGC (OGC OWS) 20%.

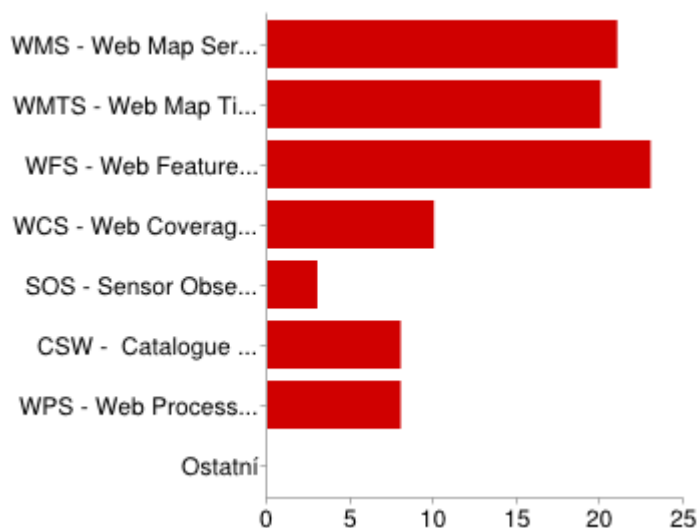
## Souborové výměnné formáty



Obr. 9: Preferovaný souborový formát pro otevřená geodata

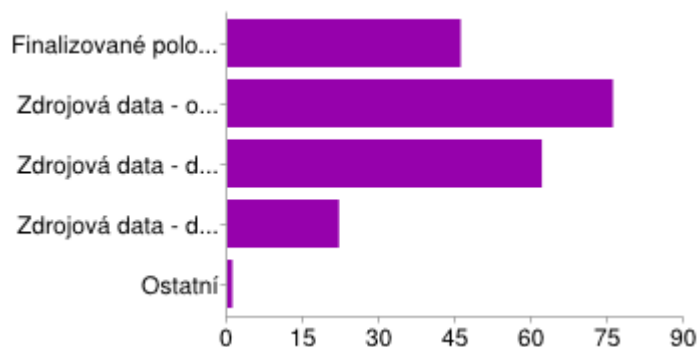
Nejvíce preferovaným výměnným souborovým formátem je ESRI Shapefile (47%). Mezi dalšími formáty převažují ty otevřené (GML, KML, JSON, ...). V seznamu chybí formát OGC GeoPackage, ale vzhledem k jeho malému povědomí v naší odborné veřejnosti, by pravděpodobně nebyl příliš preferován.

## Webové služby OGC



Obr. 10: Požadované služby OGC

Podíváme-li se na preferované služby OGC, s překvapením zjistíme, že vedle „INSPIRE prohlížečích služeb“ (WMS, WMTS), je silně zastoupena i stahovací služba vektorových dat WFS (19%).

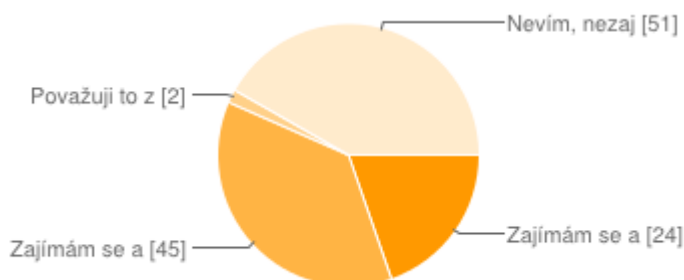


Obr. 11: Forma zpracování geodat

Podíváme-li se na požadavek na před-zpracování dat veřejnou správou, zjistíme, že převažuje požadavek na zdrojová data (přes 60%), nad požadavkem na předpřipravené analýzy.

### Provázaná (linked) data

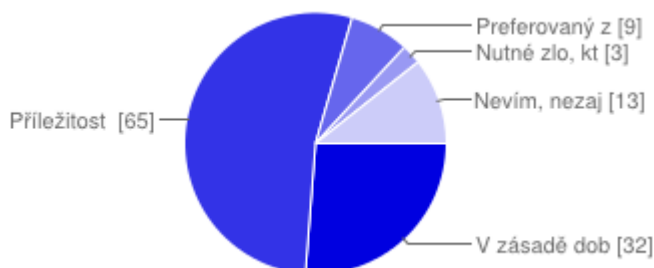
V dotazníku jsme se také ptali na pohled našich respondentů na tzv. provázaná data (linked data). Provázaná data jsou z hlediska výše zmíněného 5ti hvězdičkového klasifikačního systému na „nejvyšším vývojovém stupni“ otevřených dat. Z pohledu uživatele dat mají data nejvyšší hodnotu, protože z jednoho datového prvku se lze díky jejich vzájemnému provázání dostat na jakýkoliv další. Z pohledu poskytovatele dat se ale jedná o nejvíce náročnou úpravu dat. Jednotlivé datové prvky je potřeba opatřit identifikátory a datové sady se na sebe následně musí navzájem navázat. Jak se data mění, musí poskytovatel dat na tuto skutečnost reagovat a své datové sady udržovat stále aktuální a to tak, aby provázání nepřestalo platit jak v rámci vlastních datových sad, tak v případě prolinkování vlastní dat s datovými sadami třetích stran.



Obr. 12: Provázaná data

20% respondentů by preferovali publikování stavajících datových sad ve formě otevřených provázaných dat. 37% se domnívá, že se jedná o přirozený ale budoucí vývoj v oblasti publikace dat. Ale plných 42% dotázaných se o provázaná data vůbec nezajímá.

### INSPIRE



Obr. 13: Jak vnímají INSPIRE

V odpovědi na obecnou otázku „Co si myslíte o INSPIRE“ se ukázalo, že INSPIRE je vnímán v naší odborné veřejnosti spíše pozitivně. Většina respondentů si myslí, že je buď „Preferovaný způsob výměny dat“ (9%)



nebo alespoň vnímají INSPIRE jako „Příležitost udělat pořádek v poskytování geodat“ (53%). Jako „nutné zlo“ jej vnímají pouze 2% respondentů.

## ZÁVĚR A DISKUSE

V tomto článku jsme prezentovali některé odpovědi na otázky položené v dotazníku sestaveném ve spolupráci s IPR Praha, který směřoval na potenciální uživatele otevřených geodat. Cílem bylo prozkoumat, z jakého prostředí tito uživatelé pocházejí, jaké mají odborné a technické zázemí, v jakém formátu data preferují a jakým způsobem by preferovali jejich distribuci.

Dotazník zatím nebyl zcela uzavřen a zpracován, nicméně již nyní můžeme vidět, že

- *open source* je vnímán jako *validní alternativa* a to jako v desktopovém, tak na serverovém prostředí. Nejpřesvědčivěji je to vidět na poli prostorových databází,
- uživatelé *kladou důraz na otevřené formáty, otevřené standardy* a preferují je před proprietárními
- uživatelé *preferují původní, nikým nezpracovaná data*, před nějakým způsobem zpracovanými datovými sadami.

Po uzavření dotazníku budou jeho výsledky zveřejněny a věříme, že nejen IPR Praha přihlédnou k jeho výsledkům při své iniciativě otevírání geodat.

Otevírání dat je proces, kdy (mimo jiné) instituce ztrácí kontrolu nad „svými“ daty a zároveň tak přichází o důležitou zpětnou vazbu. Chce-li instituce zjistit, co se dále s jejich daty děje, musí tuto informaci poměrně náročně získávat zpět. Tato činnost je vždy aktivní – instituce musí pro získání této zpětné vazby „něco udělat“. Obvyklou formou je pořádání přednášek či (lépe) aktivních workshopů a tzv. hackathonů. Další možností je právě dotazníkové šetření, které bude bez pochyby po nějaké době zopakovat. Výsledky dotazníkového šetření jsou spíše než závazným dogmatem určitým návodem, který dává instituci cennou zpětnou vazbu a poskytuje určitý náhled na problematiku. Cenné jsou samozřejmě i volné komentáře, jejichž zpracování je ovšem již náročnější, nicméně bude součástí vyhodnocení dotazníku.

## LITERATURA

[1] Tim Berners-Lee: Linked Data, <http://www.w3.org/DesignIssues/LinkedData.html> [cit. 2014-09-06]

[2] Augustýn, R., Čepický, J., Landa, M. a kol: Otevírání geografických dat - Případová studie, <http://opengeolabs.cz/otevrena-geodata/>, [cit. 2014-12-19]

[3] Kolektiv autorů: Dotazník Pražská Open Geodata, <https://docs.google.com/forms/d/1oPlmmgfcml4R3YqOyAvW6jZ1GeWSVjGJaobU7EEHrlw/viewform>, [cit. 2014-12-19]