

Harmonizace českých metadat se směrnicí INSPIRE

Petr Říha

T-MAPY spol. s r.o., Špitálská 150,
50003, Hradec Králové
Kontakt.petr.riha@tmapy.cz

Abstrakt. Současná česká metadata většinou neodpovídají potřebám INSPIRE. Byla vytvořena podle českých zvyklostí a pro potřeby českých uživatelů. Příspěvek analyzuje, které typy problémů je potřebné řešit pro plnou slučitelnost s INSPIRE, a navrhuje způsob, jak toho dosáhnout efektivně a za odpovídající ceny.

Autoři široce používaného metadatového editoru mapují situaci v České republice a navrhují řešení. Jak Česká, tak i Slovenská republika již formálně implementovaly INSPIRE do svého zákonodárství, ale geografická data jsou jen málokde nabízena a prodávána současně s metadaty ve formátu INSPIRE. Naštěstí se v poslední době objevilo několik dobrých metadatových nástrojů, takže producenti dat, kteří o to budou mít zájem, budou moci nabízet data již ve formátu slučitelném s INSPIRE. V tomto příspěvku jsou prezentovány jak nejčastější nedostatky dat, tak i nástroje připravené k okamžitému použití. Firmy jako T-MAPY nebo Intergraph mají aktualizované metadatové editory, další nástroje jsou dostupné ze zahraničí, některé z nich dokonce i v češtině. ERDAS, Cenia, T-MAPY a další nabízejí metadatové katalogy s různou koncepcí řešení včetně OpenSource. V tomto příspěvku jsou také prezentovány výsledky analýzy zkoumající, co bude nutno doplnit před povolením harvestingu na metadatový portál. Jsou zmíněny problémy, slepé uličky a skryté nástrahy, kterými metadata v posledních letech procházejí. Okrajově jsou uvedeny použité technologie a protokoly.

Klíčová slova: metadata, INSPIRE, harmonizace.

Abstract. Need for harmonisation between INSPIRE and Czech metadata. Harmonisation between INSPIRE metadata requests and Czech metadata answers is necessary. The question is, how it should be done and which tools should be used to be effective enough, to avoid thousands items of poor quality metadata, and to do the job quickly and at an appropriate price. Authors of widely used metadata editor map the situation in the Czech Republic and propose appropriate solutions.

Both Czech and Slovak Republics have formally implemented INSPIRE to their national laws but in practice, businesses are rarely prepared to offer their data to others this way. Fortunately several good metadata editors appeared, organizations will soon be able to offer and even sell their data through portals. In the paper new opportunities in this area are presented, that can be immediately used. Both T-MAPY and Intergraph have renewed metadata editors, another editors are accessible from abroad, some of them are in Czech language. ERDAS, Cenia and T-MAPY offer metadata portals using various solutions including OpenSource. The paper also presents a thorough analysis of the necessary steps to be done before the data are harvested by portals. The author mentions problems, traps and setbacks, which metadata editors must comply with. Finally he mentions technology and protocols used.

Keywords: metadata, INSPIRE, harmonisation

1 Úvod

Metadata ke geoprostorovým datům vytváří v České republice několik desítek institucí, většinou institucí ve vlastnictví státu nebo územních samospráv. Tato metadata jsou jen omezeně kompatibilní a až na výjimky je nevyužívá nikdo jiný, než jeho tvůrci. Bude-li dosavadní vývoj v oblasti metadat pokračovat jako dosud, pak ani INSPIRE nedává naději, že by se na tom něco mohlo změnit. V následujícím textu si ukážeme hlavní nedostatky a některé jejich příčiny nebo možná řešení.

2 Data a metadata

V České republice jsou tisíce veřejnoprávních subjektů, významná geodata produkují stovky z nich, avšak informace o svých datech poskytuje jen málokdo. Situace je to překvapivá, když si uvědomíme

trvalý nedostatek financí ve veřejné správě (možnost prodeje dat) a ohromné úsilí o vybudování a udržení kreditu organizací.

V následujícím textu mapujeme některé producenty metadat a způsoby, jak metadata vytvářejí. Výčet se v žádném případě nesnaží o úplnost, naopak má pouze ilustrovat jak velmi rozdílný je přístup k metadatům v ČR.

Metadatový portál CENIA. Portál na adrese mis.cenia.cz [1] umožňuje základní vyhledávání fulltextem a podle organizací. Vyhledávání lze omezit na data MŽP (rezortní data). Portál působí velmi jednoduše, neposkytuje o sobě téměř žádné informace, nenašli jsme dokonce ani nápovědu. Situace se může výrazně změnit vzhledem k již uzavřenému výběrovému řízení, jsou však pochybnosti, zda bude zadavatel dostatečně silný, aby prosadil vlastní vyhlášené požadavky a kritéria.

Přehled metadat portál neposkytuje, odhadem je zde necelých 400 záznamů vytvořených ručně z klávesnice v metadatovém profilu firmy HS-RS [4]. Za poslední rok bylo přidáno nebo upraveno 180 záznamů. Metadatový záznam je členěn na části Identifikace, Kvalita, Distribuce a Metadata.

Mapové služby, Krajský úřad Plzeňského kraje. Seznam prostorových dat a metadat na adrese kr-plzensky.cz [9] je strukturován tématicky s vyznačením přístupnosti zdroje. V tomto případě se jedná o skutečný katalog naplňující smysl tohoto slova [8]. Ke každému zdroji je uveden přímý odkaz na metadata.. Vyhledávání je také možné, a to prostřednictvím metainformačního systému MetIS (T-MAPY).

Systém MetIS Plzeňského kraje poskytuje přes tisíc záznamů vytvořených v metadatovém profilu firmy T-MAPY [11] zčásti ručně a zčásti importem. Metadata je možno vyhledávat, je možno zobrazit i jejich přehled, data řadit a filtrovat. K importu metadat se využívá software ArcGIS (ESRI) a zabudovaná funkčnost MetIS umožňující převádět data z amerického formátu FGDC. Metadatový záznam je rovněž členěn na části Identifikace, Kvalita, Distribuce a Metadata, kromě toho však poskytuje i důležité informace v části Atributy. Nápověda je dostupná nejen k ovládání aplikace, ale také jako kontextová nápověda k položkám datového modelu ISO.

MICKA Liberecký kraj. Tzv. metainformační katalog Micka na adrese maps.kraj-lbc.cz [6] poskytuje pouze rozšířené vyhledávání. Informace o programu jsou dvouřádkové, nápověda má délku sotva tři strany textu. Pochybnosti nastávají ohledně úrovně technické podpory, protože zde jsou uvedeny pouze dvě freemailové adresy.

Přehled metadat tento „katalog“ neposkytuje, odhadem je zde 70 záznamů vytvořených ručně z klávesnice v metadatovém profilu firmy HS-RS. Množství novějších datových záznamů se nepodařilo zjistit.

Geoportál ČÚZK. Geoportál na adrese geoportal.cuzk.cz [2] je rozdělen na metadata o produktech a o tzv. výdejních jednotkách (viz dále). Základní vyhledávání umožňuje hledat v názvech nebo fulltextově. Portál o sobě neposkytuje informace ani nápovědu.

Obsahem Geoportálu ČÚZK je na nejvyšší úrovni pouze 42 metadatových záznamů. Metadatový záznam je v podrobném členění rozdělen do osmi částí. Pokud jsou data aktualizována postupně, lze datum aktualizace jednotlivých výdejních jednotek dále zjišťovat „interaktivním dotazováním nad přehledovými mapami s klady listů jednotlivých produktů“. Kompatibilitu ani množství novějších dat se nepodařilo zjistit. Primárním cílem Geoportálu ČÚZK je pravděpodobně podpora internetového obchodu.

Geoportál ŽP Středočeského kraje. Nový portál na adrese mapy.kr-stredocesky.cz [10] dosud neobsahuje žádná metadata. Software a způsob ovládání jsou totožné s portálem CENIA.

MetIS Dvůr Králové. Ukázka typického metadatového serveru městského úřadu [7]. Poskytuje přes 600 záznamů vytvořených v metadatovém profilu firmy T-MAPY [11] ručně a importem. Pro metadata platí totéž co pro mapové služby KÚ Plzeňského kraje.

Tento nedokonalý výčet ukazuje pouze zástupce zdrojů, které jsou veřejně dostupné. Bohužel mnoho organizací, a to i z veřejnoprávní sféry, tvoří metadata pouze pro svoji potřebu a nemá tedy zájem je veřejně prezentovat.

Co tedy můžeme z předložených ukázek a provedených testů vyvodit:

- V drtivé většině jsou metadata jen v českém jazyce. Jejich využití pro uživatele mimo české jazykové prostředí tedy je a bude velmi obtížné, přestože ani v současné době není žádný problém použít anglickou mutaci software nebo například vlastního klienta CS-W.

- Metadata jsou téměř vždy vyplněna jen částečně, takže datový obsah neodpovídá ani požadavkům norem ISO, ani INSPIRE. Aktuálnost metadat nebylo možno prozkoumat, ale vzhledem k neexistující veřejné kontrole zde budou patrně velké rezervy.
- S výjimkou netypicky stručných metadat ČÚZK nejsou rozlišována vlastní a zakoupená data. Vytvoření statistických reportů z takových zdrojů, jaké jsou zde uvedeny, bude tedy velmi obtížné.
- Dalším problémem je jen částečná kompatibilita dat, která je způsobena především chybějícím českým metadatovým profilem. Jak se dalo předpokládat, a bylo to prakticky otestováno, i z podrobně vyplněného záznamu se převodem do katalogu velká část metadat ztratí.
- V INSPIRE není podpora pro hierarchii a vazby mezi metadatovými záznamy, ani pro zobrazení umístění záznamu v datovém modelu organizace. Přesněji řečeno, hierarchie a vazby lze implementovat různým způsobem, takže o nějaké kompatibilitě se zde vůbec nedá hovořit.
- Metadata zkopírovaná do jiného katalogu (harvesting) bohužel často ztrácí vazbu (odkaz) na původní záznam, kde jsou uloženy dodatečné informace, a tím se jejich vypovídací schopnost a užitečnost dále omezuje.

3 Software

V tomto příspěvku jsme software omezili na typické zástupce metadatových editorů. Při veškeré snaze o objektivnost je možné, že zde zmíněné vlastnosti jsou ovlivněny osobním postojem autora.

ArcCatalog (ESRI). Komerční software ArcCatalog je dodáván jako součást produktové řady ArcGIS Desktop. Editor těží z postavení ESRI jako vedoucí firmy v oblasti GIS.

Silné stránky. Dobře dokumentovaný datový model, možnost využívat metadata, která jsou součástí dat (platí pouze pro FGDC editor). Podpůrné jednoúčelové nástroje.

Slabé stránky. Značně nepřehledné ovládání, problematická podpora češtiny. Oblasti, do kterých se ukládají metadata typu FGDC a ISO, jsou bohužel zcela oddělené.

INSPIRE Metadata Editor (JRC). Webová aplikace [5], kterou pro Evropskou komisi provozuje Joint Research Centre.

Silné stránky. Umožňuje validovat metadata, vytvořit metadatový záznam a uložit jej na počítači uživatele jako XML soubor.

Slabé stránky. Je to jen ukázka aplikace (proof of concept), navíc pouze v angličtině. Pro tento server je typická pomalá odezva a časté výpadky.

MetIS (T-MAPY). Webová aplikace, kterou používá více než 30 organizací veřejné správy v ČR.

Silné stránky. Efektivní plnění dat, nadstavbové funkce, dobrá technická podpora (zázemia velké SW firmy).

Slabé stránky. Prozatím dostupné jen v češtině a angličtině. Pevný datový model (kompatibilní s ISO i INSPIRE).

Micka (HS-RS). Webová aplikace, kterou používá několik organizací veřejné správy v ČR.

Silné stránky. Pružný datový model (lze zvolit strukturu vytvářených metadat), jazykové mutace.

Slabé stránky. Méně efektivní ovládání, pomalejší vývoj, nedostatečná nápověda.

XML Copy Editor. Zástupce obecného XML editoru s možností validace dat. XML Copy Editor byl uvolněn pod licencí GNU General Public License. Je dostupný na serveru sourceforge.net [3] mimo jiné i ve slovenštině.

Silné stránky. Ničím neomezená možnost editace metadat, funkce Lock Tags, cena.

Slabé stránky. Nutná znalost požadované struktury metadat, nedostatečná možnost využívat číselníky, problematická podpora. Nutnost použít klienta pro zobrazení a katalog k prohledávání metadat.

Závěr

Na konkurenčním trhu je již dostatek metadatových editorů, takže producenti dat, kteří o to mají zájem, mohou již nyní nabízet data s metadaty ve formátu slučitelném s INSPIRE. Zatím se to však neděje a metadat přibývá. Potřeba metadatové záznamy strukturovat (datový model) je prokazatelně větší, než potřeba nabízet je prostřednictvím katalogů. Domníváme se, že producenti geodat by měli zvážit následující priority:

- vytvářet metadata s použitím prověřených nástrojů se zaručenou technickou podporou,
- důrazně odlišovat vlastní a zakoupená data a
- dodržovat platné evropské standardy.

Opatření, která by bylo k naplnění směrnice INSPIRE potřebné v ČR provést na centrální úrovni, nejsou tématem tohoto příspěvku.

Reference

1. CENIA. <http://mis.cenia.cz/portal/client/?lang=cze>
2. ČÚZK. <http://geoportal.cuzk.cz/default.aspx?a=cswwclient&mode=siple>
3. Gerald Schmidt. <http://xml-copy-editor.sourceforge.net>
4. HELP SERVICE – REMOTE SENSING spol. s.r.o. <http://www.bnhelp.cz/>
5. INSPIRE geoportal. <http://www.inspire-geoportal.eu/index.cfm>
6. Liberecký kraj. <https://maps.kraj-lbc.cz/metadatas/>
7. Městský úřad Dvůr Králové. <https://mapy.mudk.cz/tgs/metis/>
8. Petr Říha. *Slovník počítačové informatiky*. Montanex 2002 Ostrava. ISBN 80-7225-083-3.
9. Plzeňský kraj, Mapy. <http://www.kr-plzensky.cz/article.asp?sec=556>
10. Středočeský kraj. <http://mapy.kr-stredocesky.cz/mapserv/catalogue/?SID=&lang=cze>
11. T-MAPY spol. s r.o. <http://www.tmapy.cz>