

VYUŽITÍ TECHNOLOGIE XLINK K SURFOVÁNÍ MEZI GEODATY

Autor: Ing. Michal MED

Abstrakt: Organizace jako W3C, OGC, ISO a další definují řadu standardů. S nástupem INSPIRE by tyto standardy měly být využívány ve větší míře, což přináší řadu výhod. Se znalostí použitého standardu lze očekávat, jak budou data nebo služba vypadat a co od nich očekávat. S INSPIRE je spojena zásada sbírání a publikování dat pouze jednou. Objem dat je celkově menší, ale vzniká potřeba jednotlivé geoprostorové prvky navzájem pospojovat. Jedním ze způsobů, jak navzájem propojit prostorová data využívá standardů standardů GML 3.2.1, WFS 2.0.0 a XLink a umožňuje surfování mezi prostorovými daty podobně jako mezi hyperlinkovými odkazy v internetu, plně v souladu s principy a technologiemi INSPIRE.

Jazyk: Cze

1 | Technologie XLink

Organizace W3C¹ definuje XML Linking Language (XLink) jako jazyk umožňující vytváření vnitřních i vnějších odkazů – linků – v XML dokumentech a přiřazování metadat k těmto odkazům. W3C spravuje standard stejného jména, který technologii XLink komplexně popisuje.

V souborech GML je XLink využívána k referencování mezi jednotlivými objekty a přesně tak je plánováno XLink využít při tvorbě a správě dat dle specifikací INSPIRE. Prozatím jsme se potýkali s problémem samotné tvorby dat, nyní jsou data připravena a je třeba zajistit provázanost mezi nimi. To je spojené s mezinárodně uznávanými standardy o otevřených datech, které bychom chtěli dodržet.



Úroveň otevřenosti dat je dána podle následujících pravidel pro Linked Open Data²

- ★ Data jsou publikována na internetu pod otevřenou licenci,
- ★★ Data jsou publikována ve strukturované formě,
- ★★★ Data jsou opatřena metadaty,
- ★★★★ Data jsou katalogově vyhledatelná,
- ★★★★★ Data jsou navzájem provázána – Linked Data.

Data INSPIRE již z principu splňují stupeň ★★³, tedy strukturovaná data publikovaná na internetu a opatřena metadaty. K získání čtvrté hvězdy je potřeba tezaurus, což je v českých podmínkách dlouhodobý problém, který se snad pomalu začíná řešit. K získání páté hvězdy je potřeba publikovat provázaná data, k čemuž chci využít právě technologie XLink.

INSPIRE data spravovaná Sekcí centrální databáze jsou data prostorových témat Parcely, Adresy, Územní správní jednotky a připravované téma Budovy. Každé z témat obsahuje řadu prostorových prvků, které by měly obsahovat odkazy na prostorové prvky jiných témat (např.: budova by měla mít vazbu na adresní místo a parcelu atd.). Kompletní seznam prostorových prvků je v kapitole 2. Navrhované způsoby řešení linkování mezi nimi je v kapitole 3.

¹World Wide Web Consortium – <http://www.w3.org/>

²<http://www.w3.org/DesignIssues/LinkedData.html>

³TR

Technologie XLink vytváří odkazy mezi prvky XML prostřednictvím spojujícího elementu (linking element), což je element stvrzující existenci linku. Linking element odkazuje na zdroj nebo jeho část, což může být soubor, obrázek, dokument, program nebo výsledek dotazu. Pro adresování je použit IRI (Internationalized Resource Identifier). Umožňuje odkazování nejen zdroje, ale i jeho části (e.g. pokud je zdrojem XML dokument, může XLink odkazovat na jeho jednotlivé elementy). Jedním z nejběžnějších použití technologie XLink je vytváření hyperlinků.

Při použití XLinků musí být dodržována jistá pravidla. Standard je dělí na Markup Conformance (značkovou shodu) a Application Conformance (aplikační shodu). Zjednodušeně řečeno značková shoda ošetřuje, aby byl linking element správně zapsán po technické stránce a aplikační shoda zajišťuje, aby byl XLink v dokumentu použit pouze tam, kde je to vyžadováno/povoleno.

Při praktickém používání XLink je používána řada atributů. Reference prostřednictvím technologie XLink pak může vypadat například takto:

```
<my:crossReference
  xmlns:my="http://example.com/"
  xmlns:xlink="http://www.w3.org/1999/xlink"
  xlink:type="simple"
  xlink:href="students.xml"
  xlink:role="http://www.example.com/linkprops/studentlist"
  xlink:title="Student List"
  xlink:show="new"
  xlink:actuate="onRequest">
Current List of Students
</my:crossReference>
```

Při použití pro linkování prostorových dat INSPIRE je zatím nejpravděpodobnější, že XLink bude odkazovat na dotaz webové služby stahování dat. Bude tedy používat parametry href s dotazem, type obsahující simple a title, který bude obsahovat v textové podobě název nebo identifikátor odkazovaného prvku.

2 | Linkovatelné prostorové prvky

Data sérií datových sad INSPIRE jsou strukturována a zapsána v souboru ve formátu GML ve verzi 3.2.1 v souladu se směrnicí INSPIRE. Tyto soubory jsou validovány proti schématům XSD, které definují obsah a strukturu GML. Prostorové prvky jsou členěny do tzv. FeatureTypes. Jednotlivé prvky pak obsahují řadu atributů, mezi nimi i atributy odkazující na prvky jiného FeatureType, který může být umístěn ve stejném souboru, v jiném souboru, nebo dokonce v souboru jiného tématu prostorových dat. Tyto atributy jsou zde vypsány podle toho, do kterého featureType patří:

Addresses

- **FeatureType Address**

- **building** – odkazuje na feature typu AbstractConstruction (nebo jeho dědice) z tématu Budovy (**BU**),
- **parcel** – odkazuje na feature typu CadastralParcel z tématu Parcely (**CP**),
- **component** – odkazuje na feature abstraktního typu AddressComponent z tématu Adresy (**AD**), jehož potomky jsou features typu AddressAreaName, AdminUnitName, ThoroughfareName nebo PostalDescriptor,

- **FeatureType AddressComponent** a jeho potomci

- **adminUnit** – odkazuje na feature typu AdministrativeUnit z tématu Územní správní jednotky (**AU**),
- **transportLink** – odkazuje na feature typu TransportLink z tématu Dopravní síť (**TN**),
- **namedPlace** – odkazuje na feature typu NamedPlace z tématu Zeměpisná jména (**GN**),

Administrative Units

- **FeatureType AdministrativeUnit**

- **lowerLevelUnit** – odkazuje na feature typu AdministrativeUnit z tématu Územní správní jednotky (**AU**),
- **upperLevelUnit** – odkazuje na feature typu AdministrativeUnit z tématu Územní správní jednotky (**AU**),
- **NUTS** – odkazuje na feature typu NUTSRegion z tématu Územní správní jednotky (**AU**), který je ve verzi 3.1 specifikace zrušen,
- **boundary** – odkazuje na feature typu AdministrativeBoundary z tématu Územní správní jednotky (**AU**),

- **FeatureType AdministrativeBoundary**

- **admUnit** – odkazuje na feature typu AdministrativeUnit z tématu Územní správní jednotky (**AU**),

- **FeatureType NUTSRegion**

- ve verzi 3.1 specifikace zrušen,

Cadastral Parcels

- **FeatureType CadastralParcel**

- **administrativeUnit** – odkazuje na feature typu AdministrativeUnit z tématu Územní správní jednotky (**AU**) na nejnižší úrovni členění, tedy na úrovni obce (4thOrder),
- **zoning** – odkazuje na feature typu CadastralZoning z tématu Parcely (**CP**),

- **FeatureType CadastralBoundary**

- **parcel** – odkazuje na feature typu CadastralParcel z tématu Parcely (**CP**),

- **FeatureType CadastralZoning**

- linky jsou volitelné v případě, že je v daném státě více úrovní členění katastrálních území, která jsou vztahena do INSPIRE, což není případ České republiky,

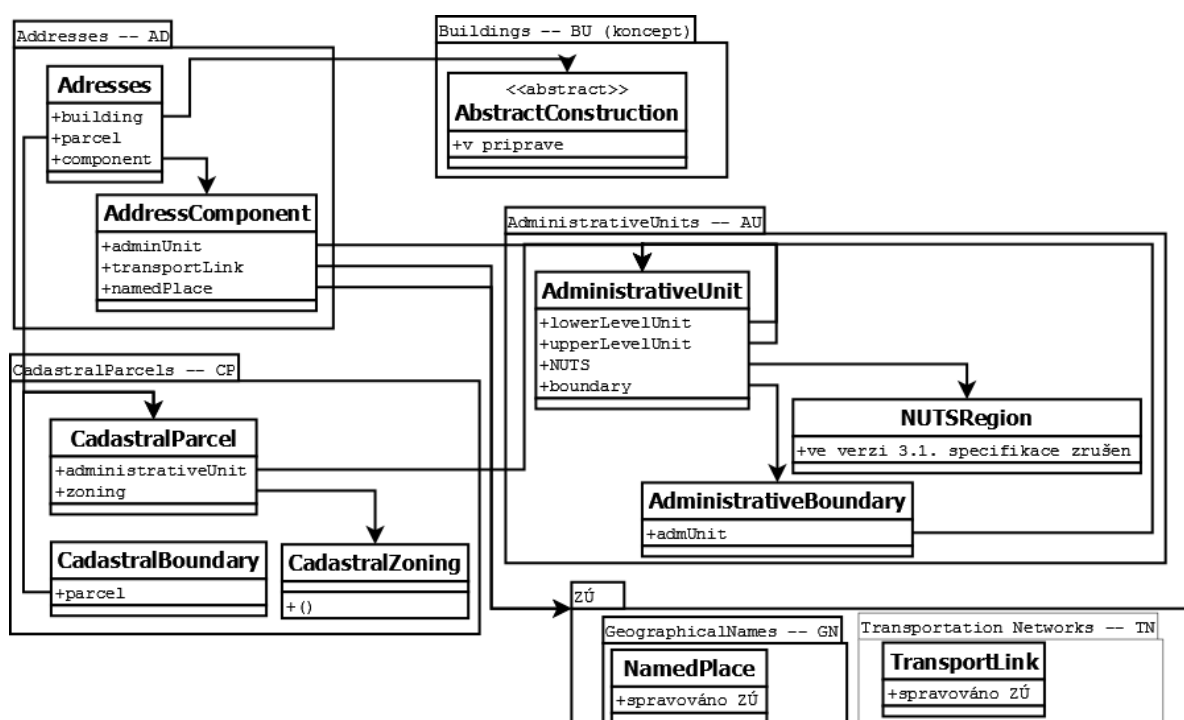


Figure 2.1: Diagram vztahů mezi vybranými atributy a featureTypy v tématech AD, AU a CP

Aby byly vztahy mezi jednotlivými vztahy jasnější, jsou znázorněny v obrázku 2.1.

3 | Návrhy možností implementace

Implementace technologie `XLink` je zřejmá. V hlavičce dokumentu je potřeba zahrnout namespace s XSD schématem `xlink` a u prvků, které mají obsahovat link je potřeba přidat atribut `xlink:href="url_odkazu"`. Tato kapitola se zabývá návrhy, jak má vypadat právě `url_odkazu`. Současná implementace služeb INSPIRE umožňuje několik variant vyhledávání konkrétního feature prostřednictvím dotazu `getFeature` služby WFS:

- s využitím parametru `FILTER`,
- s využitím parametru `FeatureId`,
- s využitím rewritů dotazů,
- odkazem na datový soubor, který feature obsahuje.

Parametr `FILTER` V dotazech `getFeature` je implementována možnost filtrování výsledků na základě různých parametrů. Parametrem pro filtrování výsledků podle identifikátoru je `resourceId`. Příkladem volání takového dotazu pro téma `Parcely` je dotaz:

```
http://services.cuzk.cz/wfs/inspire-cp-wfs.asp?service=WFS&version=2.0.0&request=GetFeature&
typenames=CadastralParcel&srsName=urn:ogc:def:crs:EPSG::5514&FILTER=<Filterxmlns="http://www.
opengis.net/fes/2.0"xmlns:gml="http://www.opengis.net/gml/3.2"xmlns:xsi="http://www.w3.org/2001/
XMLSchema-instance"><ResourceIdrid="1623333801"/></Filter>
```

Tento dotaz vrací z tématu `CadastralParcels` feature s identifikátorem `1623333801`, kterým je parcela `222/1` z katastrálního území `752657`. V této verzi linkování by feature `zoning` ve výsledku tohoto dotazu obsahoval `xlink` s dotazem:

```
http://services.cuzk.cz/wfs/inspire-cp-wfs.asp?service=WFS&version=2.0.0&request=GetFeature&
typenames=CadastralZoning&srsName=urn:ogc:def:crs:EPSG::5514&FILTER=<Filterxmlns="http://www.
opengis.net/fes/2.0"xmlns:gml="http://www.opengis.net/gml/3.2"xmlns:xsi="http://www.w3.org/2001/
XMLSchema-instance"><ResourceIdrid="752657"/></Filter>
```

Tento dotaz vrací feature typu `CadastralZoning` s katastrálním územím číslo `752657`, kterým je `Křížov u Sovince`.

Parametr `FeatureId` Některé služby podporují dotaz `FeatureId`. Stahovací služby INSPIRE provozované Sekcí centrální databáze zatím ne, v příští verzi ale bude tento parametr naimplementován. Funkce je v podstatě stejná jako předchozí způsob, s tím rozdílem, že `FeatureId` má čitelnější a kratší zápis. Stejné dotazy, tedy na katastrální parcelu `222/1` a katastrální území `Křížov u Sovince`, vypadají následovně:

```
http://services.cuzk.cz/wfs/inspire-cp-wfs.asp?service=WFS&version=2.0.0&request=GetFeature&
typenames=CadastralParcel&srsName=urn:ogc:def:crs:EPSG::5514&featureId=1623333801,
```

<http://services.cuzk.cz/wfs/inspire-cp-wfs.asp?service=WFS&version=2.0.0&request=GetFeature&typenames=CadastralZoning&srsName=urn:ogc:def:crs:EPSG::5514&featureId=752657>.

Výsledky, které vrací WFS dotazy s podporou parametru FeatureId jsou stejné jako při použití filtru.

Rewrity Jendou z posledních úprav stahovacích služeb ze strany Sekce centrální databáze byla úprava dotazů prostřednictvím rewrítů. Jedná se o zjednodušenou formu zápisu dotazu užívajícího FILTER. Nejedná se doslova o jiný způsob volání. Dotaz zavolaný touto formou zápisu udělá pouze to, že zavolá odpovídající dotaz za použití filtru. Výhodou je tedy pouze jednoduchost a systematičnost zápisu. Nevýhodou je, že rewrity byly provedeny pouze pro některé souřadnicové referenční systémy (S-JTSK a ETRS-89). Předchozí dotazy vypadají takto:

<http://services.cuzk.cz/wfs/cp/epsg-5514/CP.1623333801>, <http://services.cuzk.cz/wfs/cp/epsg-5514/CZ.752657>.

Prefix CP určuje, že se jedná o feature typu CadastralParcel, CZ znamená CadastralZoning. Tento způsob je nejpřehlednější, omezuje ho jen menší podpora souřadnicových systémů. Přidělat další systémy by nemělo být problematické.

Poslední z uvedených způsobů je využitelný pouze někdy, především v případě, že se jedná o xlink ve velkém souboru a odkazovaný prvek je také jeho součástí. Příkladem může být linkování adresních komponentů v předpřipravených souborech po obcích pro téma Adresy, nebo vytvoření hierarchie územních správních jednotek v předpřipraveném souboru tématu Územní správní jednotky. Pro širší využití je ale tento způsob nepraktický a proto není dále rozveden.

Vyhledávání s použitím parametru BBOX je pro tento případ nevhodné, neznáme totiž přesné geografické umístění vyhledávaného feature (na rozdíl od identifikátoru).

4 | Praktické dopady

V předchozích kapitolách byly popsány technické parametry, jak dosáhnout provázanosti mezi datovými sadami. V praxi je ale důležité ptát se i na to, jaké= výhody řešení přináší. To je i důvod, proč jsou všechny ukázky směřovány ke Směrnici INSPIRE – data obsahově spadající do témat INSPIRE mají jednoznačně dané vazby, jak bylo ostatně popsáno v předchozích kapitolách. Podle výkladu směrnice INSPIRE však stačí, aby tyto vazby byly reprezentovány pouhým identifikátorem daného prvku v uzavřené datové sadě. Pokud však bude vazba provedená prostřednictvím odkazu na volání konkrétního prvku službou Web Feature Service, výsledkem budou data ve standardizovaném formátu obsahující harmonizovaný prostorový objekt.

Nutným dalším krokem by měl být software, nebo spíše plugin do existujícího GIS software (například QGIS¹), který umožní vyhledat dotazy WFS v attributech `xlink` prostorových dat a umožní jejich zavolání a tím stažení souvisejících prostorových prvků.

V kontextu dat harmonizovaných v souladu se Směrnicí INSPIRE by se jednalo například o vyhledání adresních míst k vybraným budovám, nebo k přiřazení územního správního celku adresním místům. Potenciál využití této technologie však dalece přesahuje hranice INSPIRE a umožňuje využití nad prakticky libovolnými sadami prostorových dat.

Nutno ještě dodat, že v kontextu propojování dat není `xlink` zdaleka jedinou dostupnou technologií, v současné době probíhá několik projektů pokoušejících se o propojování různých datových sad pomocí technologie Resource Description Framework (RDF)², která byla původně navržena pro správu metadat, nutno ovšem podotknout, že valná většina z nich se nezabývá prostorovými daty.

¹ Software QGIS je v současnosti jediným software dostupným pod otevřenou licencí, který podporuje stahování dat prostřednictvím služby WFS 2.0.0, která je pro tuto funkcionalitu nezbytná.

² Především se jedná o sémantický web a tzv. Linked Data.