

C5V1 - Definice scénářů využití konceptuálních modelů

Vytvořeno v rámci projektu

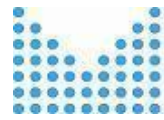
Rozvoj datových politik v oblasti zlepšování kvality
a interoperability dat veřejné správy
CZ.03.4.74/0.0/0.0/15_025/0013983

KA 6: Návrhy a realizace prostředí pro ontologické konceptuální datové modelování

Výstup: C5V1 Definice scénářů využití konceptuálních modelů

Verze výstupu: 01

Indikátor: 8 05 00 Počet napsaných a zveřejněných analytických a strategických dokumentů
(vč. evaluačních)



1. Definice výstupu

Typ výstupu: Dokument s popisem scénářů

Budou definovány a analyzovány prioritní scénáře využití ontologických konceptuálních datových modelů legislativy a agend pro zlepšení sémantické datové interoperability informačních systémů, které použijeme v dalších krocích pro validaci výstupů cíle.

2. Úvod

Tento dokument popisuje scénáře práce s konceptuálními modely ve veřejném sektoru - zejména tedy při publikaci a údržbě otevřených dat, a obecněji též v agendách veřejné správy.

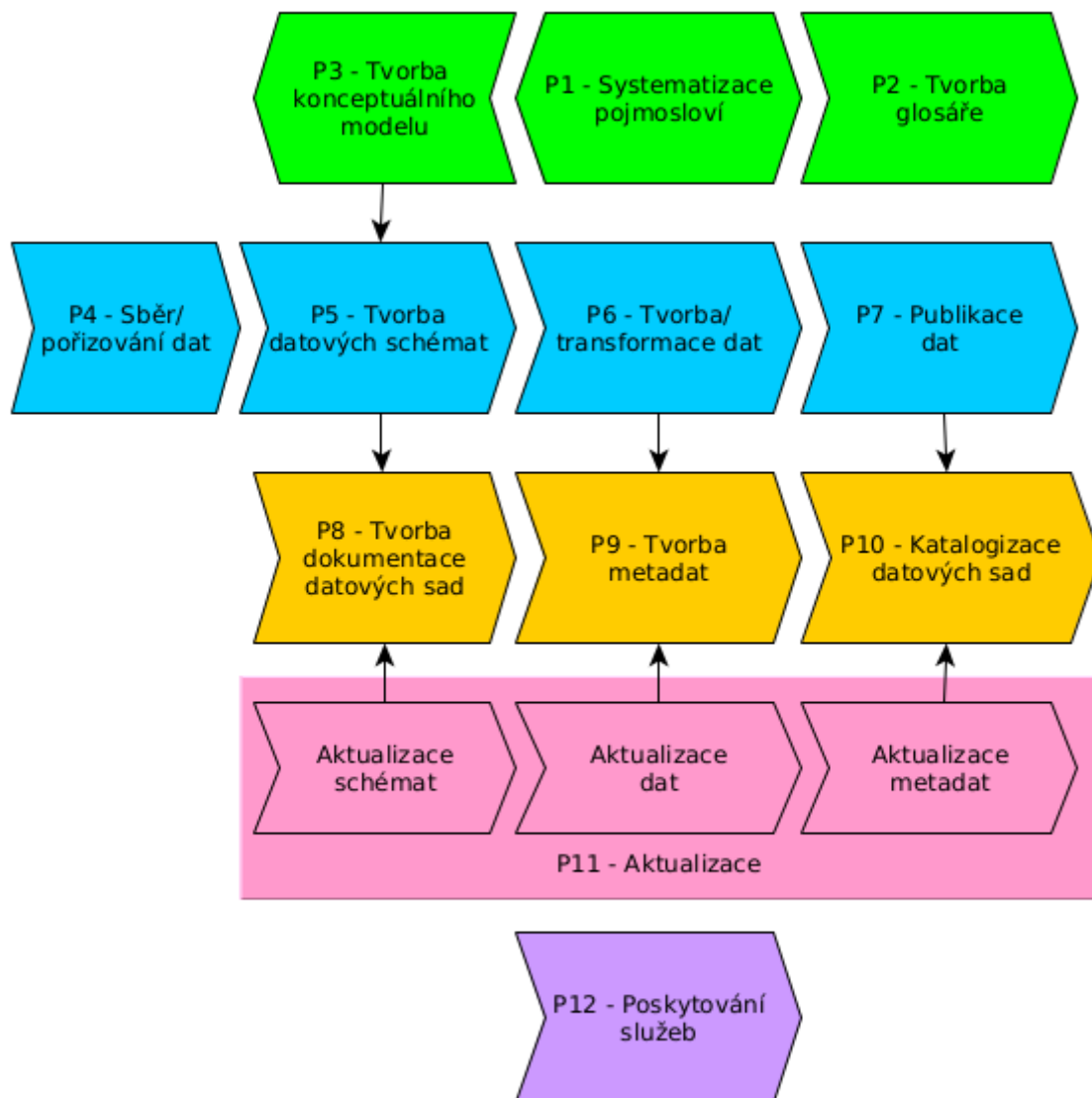
V [kapitole 3](#) identifikujeme jednak klíčové procesy (aktivity), podle kterých se při práci s daty a datovými zdroji již nyní postupuje, jednak ty, které s sebou nově přináší agenda tvorby a údržby konceptuálních modelů.

Jednotlivé identifikované prioritní scénáře, představené v [kapitole 4](#), pak popisujeme ve vztahu k těmto klíčovým aktivitám. Navržené scénáře mohou spadat do oblastí, ve kterých se budou s určitou variabilitou konkrétní scénáře obměňovat. Typy scénářů v oblastech se budou ve svých charakteristikách opakovat a většina problémů, které bude vhodné řešit za použití konceptuálních modelů, bude odpovídat některému ze zde popsaných prioritních scénářů. I přes to zde nejsou popsány (a ani nebylo snahou popsat) všechny scénáře, k nimž je použití konceptuálního modelu výhodné.

3. Procesy tvorby a využití konceptuálních modelů ve veřejné správě

Na úvod je vhodné popsat, jaké procesy doprovází tvorbu a využití konceptuálních modelů. Potřeby subjektů veřejné správy, které stojí za těmito procesy, do značné míry vysvětlují, jaké scénáře se pro využití konceptuálního modelování nabízejí. Obrázek 1 popisuje procesy na čtyřech úrovních. První úroveň (na obrázku zelená) popisuje proces tvorby konceptuálního modelu, druhá úroveň (modrá) směřuje k publikaci dat, třetí úroveň (žlutá) se zabývá dokumentací a čtvrtá (růžová) procesem aktualizace aktiv vytvořených na ostatních úrovních. Proces P12 - Poskytování služeb s konceptuálním modelováním nesouvisí věcně, ale jedná se o služby, které mohou sloužit k zpřístupnění datových výstupů kterékoliv dílčí aktivity vhodnou formou.

Hierarchie procesů ukazuje souvislou posloupnost od potřeby znalosti nové domény až po publikovaná, popsaná a aktualizovaná data dané domény propojená s ostatními daty veřejné správy, ovšem vstupy do jednotlivých procesů mohou pocházet i z jiných zdrojů. Každý scénář využívá a prioritizuje jiné procesy, procesy se v různých scénářích opakují a pomocí procesu jsou scénáře propojeny, nebo na sebe přímo navazují.



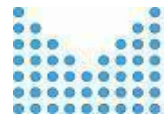
Obrázek 1: Procesy tvorby a využití konceptuálního modelování ve veřejné správě.

3.1. Procesy tvorby konceptuálního modelu

První úroveň obsahuje tři procesy. Pracovní postup začíná u systematizace pojmosloví. Ta v první fázi probíhá v hlavách, na papírech a v tabulkových procesorech expertů na danou doménu. Typicky se může jednat například o sjednocení názvosloví mezi různými skupinami účastníků meziresortních projektů nebo pro tvorbu nového legislativního dokumentu. Cílem je identifikovat klíčové pojmy dané domény a jejich jednoznačné a jedinečné definice.

Následuje proces formalizace pojmů s přesně specifikovaným významem do tzv. glosářů - uspořádaných hierarchií. Hierarchická struktura představuje vztahy definující širší význam pojmů (např. specializace *obec* → *město*) a slouží tak k lepší orientaci v pojmech samotných.

Vzhledem k množství vztahů, které širší význam může představovat, může být pojem ve vztahu k více



pojům a prostá hierarchická struktura pro toto zobrazení není vhodná. Závěrečný proces na této úrovni spočívá v tvorbě vztahů mezi pojmy a specifikaci jejich významu (například pojem “barva očí” je vlastností pojmu “člověk” a představuje konkrétní barvu, kterou mají oči daného člověka - zároveň je možné definovat i obor hodnot, kterých vlastnost bude nabývat). Součástí procesu může být i vizualizace typů pojmů a konkrétních vztahů, například formou diagramů.

První úroveň procesů konceptuálního modelování vytváří formalizované slovníky, které popisují podrobně a strojově čitelně procesy ve státní a veřejné správě, legislativní a jiné dokumenty včetně zákonů, vyhlášek a nařízení a vzájemně je propojují. Jedná se o procesy, které je nutné nově zavést, aby bylo možné využít výhod konceptuálních modelů v jednotlivých scénářích. Procesy na dalších úrovních využívají výstupů těchto tří procesů.

3.2. Procesy publikace dat

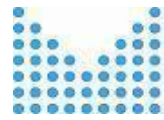
Na různých úrovních veřejné nebo státní správy vznikají data, která bývají pořizována na základě legislativních či jiných dokumentů a na základě procesů veřejné nebo státní správy. Jedná se například o data Registru území identifikace, adres a nemovitostí, která jsou pořizována a publikována na základě zákona č. 111/2009 Sb., o základních registrech, nebo o data digitální technické mapy poskytovaná podle zákona č. 47/2020 Sb., se kterými se pracuje ve scénářích č. 2 a č. 4. Tato data by měla být publikována v podobě, která umožňuje využívat pojmů v legislativě či procesech zavedených. To přináší řadu výhod, jmenujme například popis datových sad na základě definic v legislativě nebo propojení různých datových sad napříč veřejnou správou. K publikaci dat směřují čtyři procesy.

První z nich, samotný sběr dat, se konceptuálním modelováním až tolik nezabývá, může však využít konceptuálních modelů pro zjednodušení samotného sběru dat např. tak, že formuláře sloužící ke sběru dat budou přímo automaticky vytvářeny z konceptuálního modelu a tím poskytnou tvůrcům dat např. legislativní definice pro jednotlivé atributy. V řadě scénářů jsou data již posbírána a týkají se pouze jejich publikace. V případě, že jsou data sbírána za účelem jejich publikace v souladu s procesem veřejné správy nebo podle legislativy, konceptuální model umožňuje zefektivnit sběr dat - například jednou měřící metodou může být naměřeno více atributů.

Oproti tomu tvorba datových schémat může využít konceptuálního modelování v plné míře. Datová schémata jsou technické artefakty, které definují obsah a strukturu poskytovaných dat, například v závislosti na formátu. Datová sada může být publikována ve více formátech, ty ovšem musí obsahovat data se stejným obsahem. I tak by se ovšem mohlo stát, že atributy budou mít jiné názvy a data jinou strukturu (může to být dáno například formátem). Využití konceptuálního modelování v první řadě zajišťuje, že je pro různá datová schémata stejné datové sady jasné, které objekty a atributy datového schématu představují konkrétní prvky reálného světa. Datová schémata mohou být v případě precizně provedených modelů pro různé datové formáty dokonce automaticky generována.

Další proces transformuje data ze zdrojové podoby (což může být databáze, proprietární nebo otevřený formát, ale i data na papíře nebo video) do podoby dané datovým schématem, tedy odpovídající konceptuálnímu modelu. Data mohou být tvořena až během tohoto procesu, například vkládáním do formuláře, který je generován z konceptuálního modelu.

Posledním procesem je samotná publikace, která už jde mimo konceptuální modelování. Publikace



zajišťuje dostupnost dat na internetu, ať už vystavením dat na konkrétní web, nebo prostřednictvím webových služeb.

3.3. Procesy dokumentace a katalogizace

Kromě toho, že jsou data publikovaná, je nezbytné, aby byla i vyhledatelná v souladu s [FAIR principy](#). K tomu mají primárně sloužit metadata popisující standardizovaným způsobem datové sady. Metadata jsou sdružena v katalozích a umožňují vyhledávání konkrétních datových sad podle informací v nich uvedených.

Metadata bývají publikována ve strojově čitelném formátu a zpravidla odpovídají nějakému standardu, např. ISO normě, [CKAN](#) nebo [DCAT-AP \(v aktuální verzi 1.1\)](#). Jejich struktura bývá daná a jejich obsah může být z větší části generován z konceptuálního modelu. Metadata kromě informací o datové sadě obsahují i informace o územním rozsahu, aktuálnosti dat, jejich úplnosti nebo o samotném poskytovateli. Tyto informace musí být do metadata doplněny jiným způsobem.

Kromě metadat, která jsou strojově čitelná, může být z konceptuálních modelů (a to plně automaticky) vytvořena i dokumentace pro lidi, například ve formátu HTML. Kvalita vygenerované dokumentace bude úměrná kvalitě a podrobnosti popisu datových sad v konceptuálním modelu, případně s nimi spojených zdrojů (dokumentů, diagramů atd.).

Katalogizace není pak ničím jiným než publikací metadat v katalogu, aby byla veřejně vyhledatelná. Zatímco na data se mohou vztahovat licenční podmínky či poplatky, pro metadata je to nevhodné - placenou datovou sadu si nekoupí nikdo, kdo o ní nic neví.

3.4. Procesy aktualizace

Publikací životní cyklus nekončí. Procesy na nejvyšší úrovni se starají o to, aby při změně výstupů na některém z nižších procesů byly změny zapracovány a procesy publikace a katalogizace nadále generovaly aktuální data a metadata. Změny se prakticky mohou stát na třech úrovních - změní se datové schéma, změní se data samotná, nebo se změní informace v metadatach.

Změna schémat vychází zpravidla ze změny legislativy nebo jejich procesů a promítá se do procesu tvorby datových schémat skrze konceptuální modely. Změna schémat má ve většině případů za následek změnu dat (je potřeba obohatit data o nové informace) a metadat (minimálně aktualizovat informace o aktuálnosti dat a datovém schématu).

Změna dat je způsobena například aktualizací nebo novým sběrem dat, nebo rozšířením časového nebo územního rozsahu a má za následek aktualizaci metadat.

Aktualizace metadat může být sama o sobě způsobena například změnou poskytovatele dat, nebo jeho kontaktních údajů.

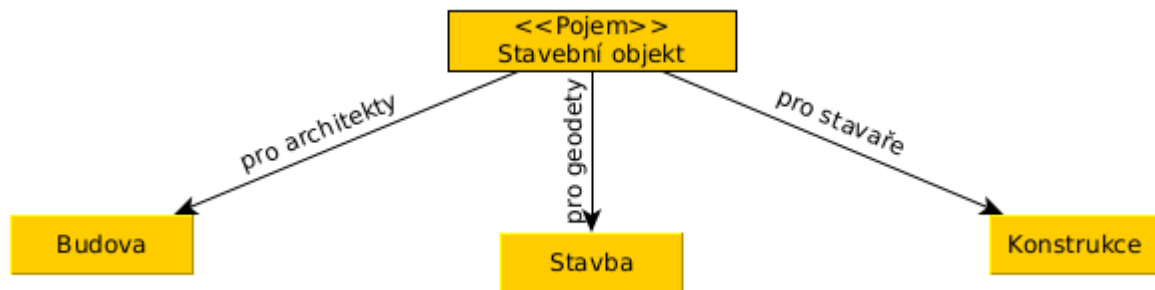
4. Prioritní scénáře

Výše popsané procesy popisují jádro tvorby konceptuálních modelů a jejich využití při publikaci dat ve

veřejné správě. Jednotlivé subjekty budou procesy využívat pro různé účely v souvislosti s tím, jaká je jejich role ve veřejné správě a jaké konkrétní úkoly a problémy řeší. V následující sekci je definováno několik oblastí pro prioritní scénáře, u kterých předpokládáme, že pokryjí většinu scénářů, které budou subjekty veřejné správy v rámci své agendy řešit. Pro každou oblast je vždy uvedeno několik příkladů konkrétních scénářů.

4.1. Uspořádání pojmosloví dané domény

Scénáře v této oblasti se týkají problematiky žádné nebo nedostatečné systematizace názvosloví. Cílem scénářů je zpravidla pouze zajistit, aby pracovníci správně a v plném rozsahu chápali významy pojmů, které používají. Alternativním scénářem je vzájemné porozumění pracovníků s různým zaměřením, kteří pro různé významy (tzv. pojmy) používají různá slova (slovo, které pro pojem používají).



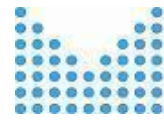
Obrázek 2: Pojem "stavební objekt" je používán v Zákoně č 111/2009 Sb. o základních registrech. Objekty tohoto typu jsou pracovníky na různých pozicích nazývány různě: geodeti jim mohou říkat "stavby", architekti "budovy" a stavaři "konstrukce". Konceptuální model zajistí, že si členové všech tří skupin rozumí i při používání různých slov pro stejný pojem.

Scénáře v této oblasti využívají procesů P1 - P2 na první úrovni z obrázku 1. Vstupem bývá zpravidla legislativní nebo jiný dokument, nebo seznam nesourodých pojmů na papíře, na HTML stránce nebo v tabulkovém procesoru. Precizním popisem a zařazením pojmů včetně zavedení přesných definic, zdrojů a upřednostňovaných, ale i alternativních názvů vzniká glosář. Ten je v řadě případů dostačujícím nástrojem pro porozumění, ovšem řešení některých problémů lze následně automatizovat vytvořením modelů nad pojmy glosáře. Kombinaci glosáře a modelu říkáme významový slovník, množinu vzájemně propojených slovníků pak nazýváme Sémantický slovník pojmů (SSP).

Uspořádání názvosloví šetří čas a umožňuje pracovat s legislativními pojmy v přesném významu a to i vzhledem k dalším slovníkům. Následující příklady konkrétních scénářů ukazují, jaké problémy uspořádání pojmosloví pomáhá řešit.

Scénář č. 1 - Uspořádání pojmosloví pro tvorbu Metropolitního plánu Prahy

Institut plánování a rozvoje hlavního města Prahy (IPR) pracuje na návrhu Metropolitního plánu Prahy. Dokument, který je v podstatě územním plánem hlavního města, se skládá z textové části, mapové části a tzv. krycích listů, které popisují jednotlivé lokality. Textová část definuje pravidla, kterými se řídí zbylé části, ale kterými se řídí i orgány činné v územním řízení. Metropolitní plán musí dodržovat další závazné legislativní dokumenty, které se zabývají územním plánováním a přepoužívá pojmy, které se v nich



vyskytují, například pro vysvětlení nových pojmů. Jedná se hlavně o Pražské stavební předpisy, které jsou aktualizovány každé dva roky a Stavební zákon (č. 183/2006 Sb.).

Uspořádání pojmosloví řeší dva dílčí problémy:

1. Identifikace pojmů z jiné legislativy použitých v Metropolitního plánu,
2. Porozumění mezi architekty, legislativci, GIS specialisty a územními plánovači,

Metropolitní plán využívá pojmy, které jsou definovány nebo používány na jiných úrovních legislativy. Pro tvorbu a používání Metropolitního plánu je nezbytné význam těchto pojmů znát i v kontextu propojenosti s ostatními pojmy. Jako velký problém se ukazuje, že v legislativě řada pojmů není definována, protože se předpokládá jejich "generický" význam. V praxi pojmy každý používá jinak a dochází ke konfliktům v jejich použití. Příkladem může být použití pojmu ve významu, který je v přímém rozporu s použitím stejného pojmu v jiném legislativním dokumentu.

Přestože Metropolitní plán Prahy má sloužit pro územní plánování, v širším hledisku se týká i legislativců, GIS specialistů nebo architektů, jejichž povinnosti jsou Metropolitním plánem ovlivněny, nebo kteří se přímo podílí na jeho tvorbě. Členové těchto skupin používají zpravidla vlastní názvosloví, pro hladký průběh procesů územního plánování je ovšem třeba, aby si členové těchto skupin navzájem rozuměli.

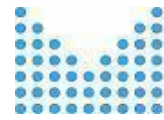
Scénář č. 2 - Uspořádání pojmosloví pro systematizaci digitálních technických map krajů

Od začátku 21. století se na úrovni krajů mluví o zavedení standardu o Digitální technické mapě (DTM). Roku 2018 podepsaly dotčené instituce memorandum o spolupráci při zajištění vzniku Digitální technické mapy České republiky, její součástí mají být digitální technické mapy krajů, obcí a digitální mapa veřejné správy. Obce a kraje již daty, které mají být náplní DTM, disponují a publikují je. Chybějící standardizace je však překážkou výstavby infrastruktur a nadměrnou zátěží pro byrokracii a výrazně zvyšuje náklady. Vznikl zákon č. 47/2020 Sb., který implementuje Digitální technickou mapu a zavádí pojmy definující její obsah, které se mohou lišit od pojmů, které jsou používány správci a pořizovateli stávajících datových sad.

Digitální technická mapa upravuje strukturu, ve které jsou taková data publikovaná. Obce i kraje dotčená data pořizují a publikují, ovšem struktura (a asi i obsah) se liší. Přichází problém určit, která data odpovídají objektům definovaných v DTM. Kromě toho je vhodné propojit model DTM k evropským a mezinárodním standardům, jako je například klasifikace tříd CCI pro Building Intelligence Model (BIM).

Řešení problému vyžaduje vytvoření glosáře objektů Digitální technické mapy a s ní spojených nařízení (zákon, vyhláška) s přesnou specifikací toho, co konkrétní objekty znamenají, aby bylo možné vytvořit funkční mapování mezi stávajícími datovými schémata a DTM. Podobným způsobem je potřeba vytvořit glosáře (případně i modely) evropských a mezinárodních standardů, nebo využít již existující.

V případě existence slovníků v podobě, která neodpovídá potřebám konceptuálního modelování může být jednodušší použít některý existující nástroj pro převod např. tabelárních dat do podoby SKOS slovníku. Pro tvorbu slovníků založenou na legislativních či jiných dokumentech je možné vytvořit specializované nástroje založené na principu textové analýzy textových dokumentů.



Dobře popsané pojmy by mělo být snazší propojit pomocí mapování. To bude často mezi pojmy ze dvou slovníků v poměru m ku n. Zároveň mapování bude probíhat mezi více slovníky zároveň. To umožňuje odvozovat propojení mezi pojmy slovníků, které mezi sebou žádná mapování nemají.

Sémantické slovníky pojmů samy o sobě vytvářejí znalostní bázi nutnou pro porozumění objektům DTM a jejich mapování na data sbíraná a publikovaná v obcích a krajích. Přesunem na druhou úroveň procesů využití konceptuálních modelů bude možné slovníky využít k tvorbě schémat datových sad DTM a transformaci stávajících dat publikovaných kraji a obcemi do této struktury.

Scénář č. 3 - Uspořádání pojmosloví v oblasti eGovernmentu

Podobný problém jako v předchozích scénářích může řešit v podstatě libovolná veřejná i soukromá instituce. Směrem k digitalizaci veřejné správy je významným projektem Národní Architektonický Plán, který určuje metodiku pro služby eGovernmentu v České Republice. Pro srozumitelnost architektonického plánu vznikl [slovník eGovernmentu](#). Tento slovník je podle výše uvedeného významu glosářem a slouží ke sjednocení názvosloví v oblasti digitalizace veřejné správy.

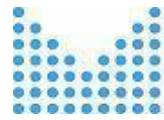
4.2. Podpora tvorby a publikace otevřených dat

Druhá oblast prioritních scénářů se zabývá tvorbou a publikací otevřených datových sad. Jedná se o jeden z primárních cílů projektu KODI a konceptuální modelování v něm má nezastupitelnou roli. Za součást publikace je z pohledu otevřených dat považována i jejich dokumentace a katalogizace. Proces tvorby a publikace otevřených dat je tvořen procesy P3 - P11, ale konkrétní scénáře v závislosti na problémech, které řeší mohou využívat pouze některé z nich.

Zákon č. 106/1999 v aktuálním znění definuje otevřená data jako „*informace zveřejňované způsobem umožňujícím dálkový přístup v otevřeném a strojově čitelném formátu, jejichž způsob ani účel následného využití není omezen a které jsou evidovány v národním katalogu otevřených dat*“. Z této definice jsou otevřenými daty pouze ta data, která jsou publikována (P7), popsána pomocí metadat (P9) a zveřejněna v Národním katalogu otevřených dat (P10). Konceptuální modelování (jak je popsáno v kapitole [Procesy tvorby a využití konceptuálních modelů ve veřejné správě](#)) do samotných těchto procesů nemusí tolik zasahovat, scénáře v této oblasti se zabývají převážně procesy tvorby datových schémat (P5) a tvorby dokumentace (P8) a metadat (P9).

Účastníky těchto scénářů jsou zpravidla organizace, které chtějí data, která mají nebo se chystají pořizovat, publikovat jako otevřená data. Případně se může jednat o organizaci, která chce publikaci otevřených dat umožnit jiným stranám.

Vzhledem k tomu, že výstupem ze scénářů v této oblasti jsou publikovaná otevřená data popsaná datovými schématy napojenými na konceptuální datové modely, nebo prostředky, které k tomuto cíli vedou, považujeme zde za vhodné uvést příklady využití těchto dat. Otevřená data odpovídající schématům vytvořeným podle konceptuálních modelů jsou propojená mezi sebou na základě svého významu, a též na legislativu, která je předepisuje. Vztahy existující mezi třídami (pojmy) konceptuálního modelu existují i v instancích těchto pojmů - konkrétních datech těchto datových sad. Takto publikovaná data jsou auditovatelná vzhledem k legislativním dokumentům, do jaké míry plní legislativní požadavky atd.



Scénář č. 4 - Zjednodušení tvorby sdíleného datového schématu pro Digitální technickou mapu (DTM)

V návaznosti na scénář č. 2, jehož výstupem je mimo jiné Sémantický slovník pojmů Digitální technické mapy je nyní úkolem vytvořit datové schéma a efektivně jej sdílet mezi poskytovatele. Účastníkem tohoto scénáře pravděpodobně nebudou sami poskytovatelé, ale centrální autorita, která poskytovatelům dat poskytne schéma, ve kterém mají data publikovat. V případě Digitální technické mapy je touto autoritou Český úřad zeměměřický a katastrální. Publikace dat DTM je dána legislativou, schéma by tedy mělo být jednotné pro všechny dílčí poskytovatele.

Na základě konceptuálního modelu Digitální technické mapy kraje, resp. obce, vytvoří ČÚZK datové schéma (nebo schémata). Datové schéma je technický artefakt, který určuje obsah a strukturu dat v daném konkrétním formátu - datovým schématem pro soubory XML je soubor XML Schema Definition (XSD), pro JSON je to JSON schéma atd. Objekty a atributy v datových schématech významem odpovídají objektům a vlastnostem Sémantického slovníku pojmů. V ideálním případě existuje strojově čitelné mapování mezi schématem a slovníkem, například pomocí kontextu v JSON-LD schématech, nebo pomocí [RDF serializace dat](#). Tím budou vzniklá data nejen otevřená, ale i propojená ([Linked Open Data](#)).

Datová schémata vybraných formátů mohou být generovaná automatizovaně z SSP. Poté je důležité zajistit, aby schémata byla dostupná jak pro poskytovatele dat (aby publikovaná data schématům odpovídala) tak i pro jejich konzumenty, aby mohli ověřit, že publikovaná data schématům odpovídají a jaký je význam jednotlivých objektů a atributů.

Scénář č. 5 - Zefektivnění tvorby nové OFN pro publikaci dat o turistických cílech

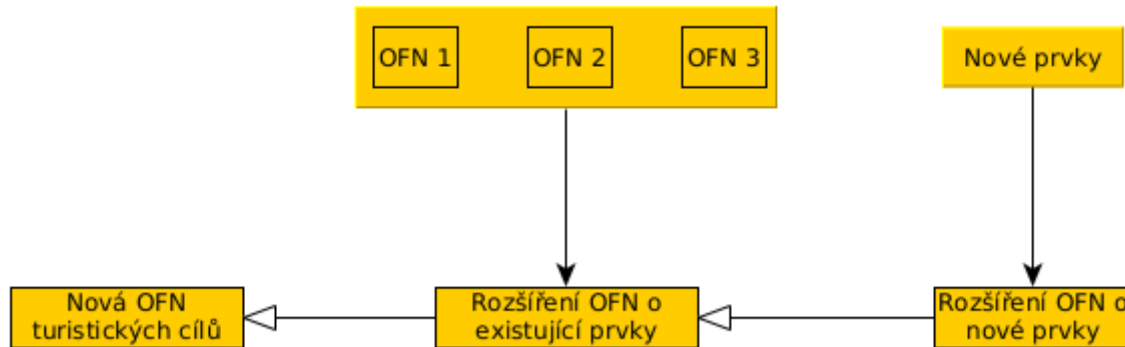
Pro účely publikace dat, jejichž typy se často vyskytující opakují, vznikají standardy v podobě otevřených formálních norem (OFN). Osoby publikující otevřená data mohou pomocí OFN zajistit kompatibilitu publikovaných dat s daty publikovanými jinými osobami a zpřístupnit tak otevřená data veřejnosti pomocí jednotných služeb a aplikací. Poskytování dat způsobem, kde mají data stejnou strukturu a datová schémata napojená na SSP umožňuje projektům na vyšších úrovních, například kraje, státu, nebo celé Evropy pracovat s daty všech dílčích poskytovatelů stejným způsobem. Informace o turistických cílech mohou být jednotným způsobem zpracovány např. na Wikipedii.

Samotné OFN jsou vytvářeny v souladu s principy konceptuálního modelování, které efektivně umožňuje jejich přepoužívání a rozšiřování. OFN v sobě kombinuje konceptuální model a z něj vygenerované datové schéma a dokumentaci.

Toho je využíváno i při tvorbě modelů nových otevřených datových sad, tedy například zmiňovaných turistických cílů. Je pravděpodobné, že struktura dat napříč různými poskytovateli dat turistických cílů v ČR se bude lišit. Pro snadné sjednocení využije poskytovatel k popisu "své" datové sady cílů kombinaci existujících OFN a dalších slovníků již vytvořených pomocí konceptuálního modelování, které může používat a/nebo rozšiřovat, pokud chce poskytovat data, která v OFN popsána nejsou.

Pro zpracování dat je důležité vycházet z jednotného standardu - Otevřené formální normy (OFN) - který definuje strukturu a obsah datových sad bez závislosti na samotném datovém schématu (ovšem různá

datová schémata mohou být jeho součástí). Taková norma musí být nejprve vytvořena. Základem OFN je podmnožina konceptuálního modelu. Na základě pojmů SSP je vytvořeno datové schéma a je k němu vygenerována dokumentace. Otevřená formální norma poté bude publikována na portálu OFN, který funguje jako katalog OFN. Při publikaci dat si poskytovatel nejprve projde OFN v katalogu, jestli již neexistuje takové, které pokrývá jeho potřeby. To je potom možné přepoužít, nebo rozšířit.



Obrázek 3: Proces tvorby a rozšiřování Otevřených formálních norem o nové nebo existující prvky

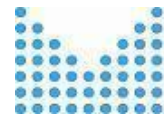
V procesu publikace turistických cílů je nejprve vytvořena základní OFN, která pokrývá všechny vlastnosti turistických cílů, u kterých autor předpokládá, že se mohou objevit v datech. Poté je OFN publikována k používání a vznikají podle ní první datové sady. Vzhledem k tomu, že atributy a objekty se v OFN mohou vyskytovat v libovolném počtu, publikované datové sady nemusí obsahovat všechny objekty a atributy uvedené v OFN a datová sada přesto odpovídá OFN. Pokud pak někdo z poskytovatelů chce publikovat data, která poskytují oproti OFN informace navíc, musí stávající OFN rozšířit. Pokud se informace dá najít v jiné OFN, například v Otevřené formální normě pro Umístění, může být tato OFN nebo její část využita pro rozšíření stávajícího datového schématu pro Turistické cíle. Může se však stát, že se nový typ informace v žádné OFN nenachází. V takovém případě je potřeba stávající OFN rozšířit a publikovat v katalogu OFN.

Scénář č. 6 - Napojení užívaných mezinárodních standardů na českou legislativu

Řada slovníků popisující typy objektů a jejich vlastností již ve standardizované podobě existuje. Jedná se o například o slovníky [ISA2 Core Vocabularies](#) Evropské komise. Není cílem SSP obsah těchto slovníků modelovat znovu, ale integrovat je do stávajícího systému. Pojmy v SSP modelují především českou legislativu, ovšem napojení na nadřazené slovníky může být vytvořeno mapováním stejně jako mezi pojmy Metropolitního plánu Prahy a stavebního zákona.

Podobným příkladem mohou být standardy, které jsou běžně používány v určitých oblastech, například AIXM v letectví nebo DATEX II v dopravě obecně. Oba standardy se chovají jako OFN, ale SSP o nich neví a tedy poskytovatelům nenabízí možnost jejich použití. Do SSP by mělo být možné takové standardy zahrnout, ovšem musí být formálně zavedeny jako součást SSP, především musí být popsány konceptuálními modely. Jedná se tedy o "obrácenou" úlohu tvorby datových schémat z formalizovaných slovníků, kde datové schéma existuje a je nutné vytvořit jeho konceptuální model.

Úředníci v české veřejné správě se zodpovídají českému legislativnímu rámci, ovšem používání mezinárodních standardů usnadňuje interoperabilitu. Napojením mezinárodních standardů na českou



legislativu získají výhody standardizace a splní závazky vůči legislativě.

4.3. Sémantická interoperabilita Veřejného Datového Fondu a Propojeného Datového Fondu

[Národní architektonický plán](#) (NAP) zavádí principy sdílení údajů mezi jednotlivými agendami prostřednictvím propojeného datového fondu (PPDF) a veřejného datového fondu (VDF). Aby bylo sdílení údajů hospodárné a efektivní, je nutné zajistit interoperabilitu informačních systémů, ve kterých jsou údaje vedeny. Dva systémy jsou interoperabilní, pokud jeden systém může od druhého systému přijmout a zpracovat data, aniž by došlo ke ztrátě dat nebo k jejich špatné interpretaci, a to bez nutnosti zásahu do jednoho či druhého informačního systému.

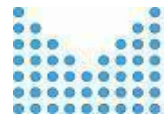
V rozsáhlém a složitém prostředí, kterým je i prostředí eGovernmentu ČR, je interoperabilita složitější, protože je nutno zajistit výměnu údajů v rámci mnoha informačních systémů. Jednotlivé informační systémy jsou vytvářeny nezávisle na sobě a mají svůj vlastní životní cyklus. Národní architektonický plán zavádí tzv. kontext, kterým se dle definice v NAP rozumí “soubor jednoznačně pojmenovaných a enumerovaných typů subjektů a objektů a jejich atributů evidovaných v agendě. V přeneseném smyslu je pak kontextem nazývána definice schématu (povolené datové struktury) datových nákladů zpráv pro publikování údajů o subjektech a objektech publikačním agendovým informačním systémem pro podporu výkonu dané agendy.”

Hlavní rozdíly mezi podporou poskytování otevřených dat (popsané v kapitole 4.2) a Veřejným nebo Propojeným datovým fondem jsou:

1. Otevřená data může publikovat i konzumovat kdokoliv, metodika využití konceptuálních modelů směřuje k tomu, aby data byla co nejvíce přepoužitelná, zatímco Veřejný nebo propojený datový fond cíleně integruje data ve známe struktuře od známých poskytovatelů.
2. Otevřená data jsou publikována tak, aby je každý mohl používat k libovolnému záměru, mají otevřený formát, jsou publikována bez omezení a jejich metadata jsou veřejně dostupná. Data Veřejného nebo Propojeného datového fondu jsou poskytována ke specifickým účelům ve formátu, který je k tomu nejvhodnější, dostupná jsou pouze systémům, které jejich využití vyžadují, ale metadata musí být veřejně přístupná ve stejné kvalitě jako u otevřených dat.

Historicky byly používány dvě techniky pro zajištění interoperability. Špagetová architektura, která řeší interoperabilitu mezi dvěma systémy tvorbou ad-hoc softwaru transformujícího data z jedné struktury do druhé, se historicky ukázala jako zcela nevyhovující a neudržitelná. Druhý způsob předpokládá centrální autoritu, která vydává komunikační standardy, které jsou jednotlivé systémy nuceny používat. Historicky se ukazuje, že tento systém funguje v omezené míře pro konkrétní domény, například letectví ([AIXM](#)) nebo zdravotnictví ([HL7 FHIR](#)). Integrace dat ve veřejné správě ale vyžaduje zajištění interoperability též mezi systémy z různých domén. Řešením je datově centrická architektura, která do středu staví konceptuální model zachycující koncepty reálného světa.

Použitím technik konceptuálního modelování se zajistí udržitelnost datových rozhraní, při změně legislativy automaticky identifikovat části datových rozhraní a datových schémat VDF, které je nutné revidovat. Konceptuální modely rovněž umožní dokumentovat význam datových rozhraní a datových schémat a tím zajistit správnou implementaci legislativy vývojáři a současně správné použití těchto



rozhraní a schémat jejich konzumenty.

Scénář č. 7 - Automatická tvorba dokumentace datové sady

Správně vytvořený konceptuální model legislativy, OFN a obecných slovníků umožňuje pro libovolnou datovou sadu, která je významem napojena na konceptuální modely, vytvářet automatizovaně dokumentaci a to jak v lidsky (HTML), tak strojově (DCAT, XML metadata...) čitelné podobě.

Dokumentace je automatizovaně i publikována a pojmy v ní použité jsou identifikovány pomocí perzistentních a dereferencovatelných IRI identifikátorů, které propojují dokumentaci s dalšími pojmy.

Pro datové sady Veřejného datového fondu se zároveň jedná o dokumentaci integrace jednotlivých systémů.

Scénář č. 8 - Konceptualizace modelu veřejných zakázek za účelem zefektivnění jejich kontroly

Konceptuální modelování umožní kontrolu souladu řešení veřejné zakázky se zadáním. Zadavatel veřejné zakázky poskytne jako součást zadávací dokumentace veřejné zakázky formální strojově čitelný konceptuální model agendy, které se zakázka týká. To poskytne oběma stranám strojově čitelnou specifikaci a zefektivní tak požadavky, řešení zakázky i následné předání výsledku. Jako součást veřejné zakázky je možné požadovat od dodavatele zpracování konceptuálního modelu dat rozšiřujícího konceptuální model agendy. Tím bude možné zefektivnit kontrolu souladu dodaného řešení s legislativou.

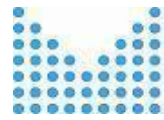
V praxi to samé platí i pro tvorbu datových formátů a poskytování dat obecně. Data musí být poskytována ve struktuře, která odpovídá konceptuálnímu modelu a tím je efektivně zamezeno tvorbě proprietárních datových formátů pro předávání dat ve veřejné správě a do určité míry předchází i vendor lock-inu, protože data ve zdokumentované otevřené podobě mohou být zpracována kýmkoliv.

Definice a procesy zadávání veřejných zakázek jsou popsány v zákoně 134/2016 Sb. o zadávání veřejných zakázek a ve vyhlášce 168/2016 Sb. Registr práv a povinností zavádí pro zadávání veřejných zakázek agendu A53. Pro plné porozumění této problematice je potřeba vytvořit konceptuální slovník veřejných zakázek s využitím pojmů ze všech tří slovníků tak, aby je funkčně propojoval podle jejich skutečného významu. Pouze jejich propojením vzniká komplexní systém vztahů mezi Zadavatelem a Dodavatelem, definující řízení, zadávání, smlouvy a popisuje požadovaný výsledek řízení. To je nezbytným předpokladem pro jednoznačnost a auditovatelnost zadávací dokumentace a výstupů veřejné zakázky. Tento postup může vést k lepší kontrole zadávání veřejných zakázek, ale i ke zlepšení a zefektivnění legislativy s ním spojené.

Využití konceptuálního modelu pro skutečné zadávání veřejných zakázek spočívá v identifikaci instancí konceptů modelu, jejich anotaci a využití propojení konceptů k nalezené souvislosti v konkrétních zadáních.

Scénář č.9 - Evidence údajů o agendách Registru práv a povinností

Zákon č. 111/2009 Sb., o základních registrech, vedle Registru územní identifikace, adres a nemovitostí, Registru obyvatel a Registru osob zavedl i Registr práv a povinností. V něm jsou vedeny údaje



o orgánech veřejné moci, soukromoprávních uživatelích údajů, agendách a právech a povinnostech právnických a fyzických osob. Orgány veřejné moci mají povinnost ohlašovat agendy obsahující definice údajů v agendě vedených.

Evidence údajů v Registru práv a povinností by měla být přes veřejné i neveřejné údaje propojená s datovými sadami publikovanými například ve Veřejném datovém fondu nebo propojeném datovém fondu, které obsahují tyto údaje. Z druhé strany - z Národního katalogu otevřených dat - jsou pak datové sady anotovány pojmy evidence údajů podle RPP. Pro ohlašování agend existuje metodika vydaná Ministerstvem vnitra, definující typy objektů, typy subjektů a typy údajů a jejich atributy. Ačkoliv jednotlivé typy údajů vycházejí z legislativy a věcně by měly odpovídat datovým položkám agendových informačních systémů sloužících k podpoře práce v dané agendě, nejednoznačnosti v metodickém postupu zavádění údajů způsobují komplikace při interpretaci datových položek a jejich vztahu k legislativě. Pojmy agend a poskytovaných údajů je vhodné mít ve formě glosáře a konceptuálního modelu. Jednak je tímto způsobem možné propojit agendy a identifikovat tak související datové sady a to ještě před jejich případnou publikací, zároveň to umožňuje efektivně vyhledávat údaje poskytované v rámci konkrétních agend. Konceptuální model registru práv a povinností umožní ještě před implementací agendových informačních systémů vyhodnotit vzájemné reference údajů mezi jednotlivými agendami, jejich souvislosti, překryvy a případné duplikace a tím zajišťuje konzistenci propojeného i veřejného datového fondu.

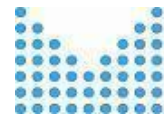
Evidence údajů pak může být použita například pro vyhledávání datových sad odpovídajících jednotlivým agendám, například při tvorbě centrální úřední desky pro zveřejňování doručenek a dokumentů osob s ohlášenou adresou na ohlašovně daného obecního úřadu.

Z konceptuálního modelu je navíc možné odvozovat technický způsob reprezentace údajů a zjednodušuje tak proces přípravy a publikace dat VDF a PPDF, podobně jako ve scénářích [Podpora tvorby a publikace otevřených dat](#). Publikovaná data jsou pak odvozena z konceptuálního modelu a je možné odvozené údaje přepoužívat v dalších agendách. Může nastat situace, kdy objekty, subjekty a typy údajů dané agendy nemají přesné zakotvení v legislativních dokumentech. V těchto případech může být na základě konceptuálního modelu agendy legislativa upravena, konceptuální modelování tedy slouží pro podporu legislativního procesu, jak je dále uvedeno ve scénářích [Podpora legislativních procesů](#).

4.4. Podpora legislativních procesů

Tvorba datových schémat, datových sad a metadat či datové dokumentace často vychází ze slovníků, které popisují legislativní dokumenty. Ty vznikají v legislativním procesu, do kterého vstupuje řada účastníků, od legislativců přes právníky až po politiky. Posuzování legislativního procesu s použitím konceptuálních modelů umožňuje snazší průběh a lepší kontrolu správnosti procesů.

Při tvorbě nové verze legislativního dokumentu (na úrovni veřejné či státní správy) konceptuální modely udržují přehled o stávajícím názvosloví v legislativním dokumentu používaném a o přesném významu těchto pojmů v kontextu tohoto dokumentu. Na základě vzájemných vztahů mezi pojmy pomáhá identifikovat konflikty, například pojmy se stejným názvem a různým významem v různých dokumentech



dokáže odlišit jako různé koncepty.

Kromě používání správných pojmů ve správných kontextech umožňuje konceptuální modelování při tvorbě legislativního procesu kontrolu ucelenosti znalostní báze, například zjištěním pojmů, které v legislativě nemají definici nebo hledáním závislostí mezi pojmy, jejichž význam se plánovanou úpravou mění.

Scénář č. 10 - Zefektivnění aktualizace Metropolitního plánu Prahy

Současná verze Metropolitního plánu Prahy prochází připomínkovým řízením, které navrhuje řadu úprav. Tyto úpravy ale nesmí narušit konzistenci celého dokumentu. Mezi pojmy vznikají v konceptuálním vztahy a ty se řídí určitými pravidly. Ty zajišťují propojování dat na základě významu, inference nových znalostí a zároveň kontrolují konzistenci modelu. Při změně významu pojmu by měl být autor změny upozorněn na pojmy související se změněným pojmem a zkontrolovat, že změnou definice nepopře definice nebo vztahy pojmů v jiné části legislativního dokumentu. Při aktualizaci verze je potřeba pokaždé projít změny a zkontrolovat, jaký mají dopad na konzistenci dokumentu. Při použití konceptuálního modelu může být kontrola nekonzistencí do značné míry automatizovaná.

Návrhy změn se týkají především definic pojmů a významových vztahů mezi nimi. Vzhledem k tomu je potřeba kontrolovat, zda změna tyto vztahy nemění. Například může být navržena změna definice pojmu, který je definován jako specializace jiného pojmu z nadřazeného zákona (který není předmětem právní úpravy). Změna definice ale bude odporovat definici nadřazeného pojmu. Odhalením této nekonzistence musí být změna přehodnocena - buď musí být pojem samostatný a nezávislý na nadřazeném pojmu, nebo jeho definice musí být upravena.

Scénář č. 11 - Odhalování nekonzistencí v legislativních dokumentech

Jedním z problémů, který se ve veřejné správě vyskytuje velice často je nejednoznačnost používaných pojmů. Ta je způsobena buď tím, že v různých zákonech mají stejná slova různý význam (například protikladné definice budovy v katastrálním zákoně č. 256/2013 Sb. a v zákoně č. 406/2000 Sb., o hospodaření energií), nebo tím, že zákony používají pojmy, které v nich (ani jinde) nejsou definovány. Tady je jedním z mnoha příkladů chybějící definice pojmů "táboření" a "nocování" v zákoně č. 114/1992 Sb., o ochraně přírody a krajiny, přestože je zákon sám v různých situacích zakazuje.

Konkrétně ve třetím odstavci § 37 zakazuje táboření mimo místa k tomu vyhrazená. Předpokládá se, že nocování tam zakázáno není, ale ani jeden z pojmů není vysvětlen tak, aby občan nebo úřední osoba, která má dodržování zákona kontrolovat věděl, v čem se liší.