

Požadavky komerční sféry na informace veřejného sektoru

Tento dokument vznikl na základě shrnutí požadavků, které byly zaznamenány v průběhu semináře **Využívání informací veřejného sektoru (PSI - Public Sector Information): Stav věcí v EU, v ČR a směry dalšího rozvoje** dne 12. září 2005 v Praze a na základě následné neformální diskuze, která probíhala ve 4. čtvrtletí roku 2005 mezi zainteresovanými stranami.

Úvod

Poskytovatelé komerčních informačních služeb využívají informací veřejného sektoru k vytváření dalších produktů a služeb, které pak – většinou doplněné o vlastní přidanou hodnotu - komerčně nabízejí na trhu. To je plně v souladu s pojetím „re-use“ informací veřejného sektoru definovaném ve Směrnici 2003/98/EC o opakovaném využívání informací veřejného sektoru.

Při své práci tito komerční poskytovatelé informačních služeb nestojí ani tak o to, aby byla informace dostupná jednorázově a pro jeden konkrétní účel, ale aby mohla posloužit jako „surovina“ pro další zpracování. Tím se požadavky poskytovatelů komerčních informačních služeb podstatně liší od potřeb občanů. Zatímco občan požaduje od instituce konkrétní informaci většinou pro vyřešení svého individuálního problému nebo k uspokojení své vlastní zvědavosti a dále s ní většinou nepracuje, komerční firmy na těchto informacích staví další služby. Na rozdíl od práva občana na informace, které řeší zákon č. 106 o svobodném přístupu k informacím („access“ – právo na přístup), právo dalších zpracovatelů na informace v podobě, kterou potřebují k „re-use“, není nikde explicitně zakotveno. Veřejné instituce většinou argumentují tím, že právo získat informaci má každá právnická či fyzická osoba, a tudíž je povinností veřejné instituce takovému požadavku vyhovět, ale právo získávat informace od veřejné instituce pravidelně de facto neexistuje, a proto nemají tyto instituce ani povinnost takovéto služby poskytovat.

Komerční firmy tak vycházejí z předpokladu, že co není zakázáno, je dovoleno, i když mnohdy používají k získání veřejně dostupných dat poněkud „partyzánské“ postupy. Je možno konstatovat, že firmy, které nějaké informace potřebují pro své komerční služby, je nakonec stejně získají, avšak jsou nuceny své aktivity různě maskovat, vytvářet poloilegální sítě kooperujících severů, pracně stahovat data řadu dnů apod.

Jsou například známy případy, kdy firma za účelem pravidelného získávání aktuálních údajů o obchodních společnostech zapojila do stahování dat desítky serverů, protože z každé IP adresy je možno stáhnout pouze 5000 záznamů, jinak hrozí nebezpečí postihu. Jiná firma, které se zdál takovýto postup v éře informační společnosti nedůstojný a nechtěla ho proto zvolit, se svou oficiální žádostí u provozovatele veřejného serveru neuspěla se zdůvodněním, že poskytování takovýchto služeb instituci zákon neukládá.

Některé veřejné instituce považují podnikání v oblasti „re-use businessu“ dokonce za neetické, protože je podle nich postaveno na „veřejných statcích“ a využíváno pro tvorbu zisku soukromých subjektů.

Je například znám případ, kdy státní organizace – držitel informací veřejného sektoru – odmítl informace poskytnout právě pouze komerčnímu subjektu, zatímco ostatním státním institucím a neziskovým organizacím informace poskytl.

Problémy při využívání informací veřejného sektoru

Problémy, se kterými se zpracovatelé a poskytovatelé komerčních informačních služeb při využívání informací veřejného sektoru setkávají, je možno rozdělit do několika oblastí:

- Obsahové
- Distribučně technologické
- Kompetenčně organizační

Obsahové problémy

Obsahové problémy lze shrnout do 4 klíčových slov:

- dostupnost
- úplnost
- aktuálnost
- konzistence

Problém dostupnosti nastává, pokud data vůbec neexistují, tj. veřejná instituce je neviduje, nesleduje. Takových případů je však poměrně málo. Mnohem častějším problémem je, že data v instituci sice existují, avšak nejsou zveřejňována. Pak jsou pro zpracovatele prakticky nedostupná, protože i kdyby poslal žádost o poskytnutí požadovaných dat, dostal by je v nejlepším případě jako jednorázový výstup. Třetím druhem nedostupnosti dat je jejich nečitelnost, což často nastává např. u naskenovaných podkladů. Nečitelná data jsou zpracovatelem vyhodnocena jako data nedostupná.

Z pohledu zpracovatelů je právě dostupnost dat zcela klíčovým bodem a je z hlediska priorit vždy na prvním místě. Ostatní problémy zmiňované v dalším textu vyplývají z neexistence jednotné integrační platformy mezi systémy státní správy a je možno se s nimi částečně vyrovnat "vnějším" integračním řešením. Pokud však některý datový zdroj není neexkluzivně dostupný vůbec (přestože nevede žádné citlivé osobní údaje), vzniká obtížně řešitelná situace při poskytování navazující služeb (viz např. registr vozidel, hospodářských zvířat, jmen a příjmení apod.)

Problém úplnosti nastává, když ve zpřístupněných datech chybí některé informace. Zpracovatelé, pokud na to vůbec přijdou, jsou nuceni dohledávat chybějící údaje v jiných evidencích, čímž jim vznikají dodatečné náklady. S tím úzce souvisí i **problém konzistence**. Řada stejných jevů je evidována v různých systémech, mezi kterými neexistuje úplná shoda. Velmi často se tento problém vyskytuje např. u adres (jeden subjekt má v různých systémech odlišné adresy), u údaje aktivity či neaktivity firmy apod. Při aktualizaci údajů u zpracovatele tak vznikají chyby, které je nutno – pokud je to vůbec možné – napravit složitými programovými algoritmy, které ale mnohdy nejsou moc platné, protože zpracovatel neví, u které instituce je údaj evidován správně a u které chybně.

Problém aktuálnosti vzniká v důsledku nesouladu reality s jejím odrazem v evidencích a informačních systémech. Každá instituce má jiný režim zpracování dat, mnohdy jedna instituce přebírá data od jiné. Na aktuálnost údajů se tak nelze spolehnout a je nutno opět konfrontovat více zdrojů, vyvíjet programové procedury a věnovat načítání dat mnoho času.

Pokud jsou data sice dostupná, avšak neúplná, nekonzistentní či neaktuální, jsou zpracovatelem souhrnně označena za **data chybná**, která je třeba pracně ošetřovat, čistit a opravovat, a pokud ani to z různých důvodů není možné, ztrácí poskytovaná komerční služba na kvalitě a může být koncovým zákazníkem odmítnuta.

Distribučně technologické problémy

Distribučně technologické problémy lze shrnout pod hesla

- forma
- metody
- distribuční cesta

Problémy s formou nastávají, když instituce není schopná nebo ochotná poskytovat data v takové formě, jakou uživatelé vyžadují. Například soubory jsou poskytovány pouze ve formátu pdf, který není možno jednoduchým způsobem dále datově zpracovávat, nebo ve formátu html, který je pro další zpracování mnohem méně vhodný než formát XML. Často jsou důležité informace (např. o změnách) poskytovány pouze ve formě textového sdělení, tiskových zpráv, zápisů z jednání apod.

Problémy s metodou přístupu k datům jsou identifikovány tehdy, nejsou-li požadovaná data dostupná takovými metodami, které uživatelé pro další zpracování potřebují. Lze například vyhledat soubor informací podle výběrových kritérií, avšak tato kritéria nelze podle potřeby kombinovat. Poskytovaná data tak plně vyhovují právu na „access“ – přístup (tj. jsou uzpůsobena převážně pro jednorázové použití, např. ověření údaje o firmě, stažení konkrétního detailu apod.), ale nepokrývají potřeby pro „re-use“ – opakované využití, které nezbytně vyžaduje použití metod hromadného stahování dat, filtrování podle kombinace různých kritérií apod.

Problémy s distribuční cestou lze vyjádřit tak, že data nejsou k dispozici takovými distribučními cestami, které uživatelé s ohledem na potřeby dalšího zpracování vyžadují. Jedná se přitom o běžné distribuční cesty, které odpovídají stupni rozvoje ICT. Například distribuce dat přes webové stránky v html formátu je jistě univerzální a snadno dostupná cesta, která ale opět mnohem lépe vyhovuje právu na „access“ než právu na „re-use“. Pro účely opakovaného využití PSI by bylo mnohem vhodnější distribuovat data jinými cestami, např. prostřednictvím ftp nebo webových služeb.

Kompetenčně organizační problémy

Tato kategorie problémů se týká zejména těchto oblastí:

- nedostatek koordinace
- nevyjasněné kompetence
- neexistence standardů a pravidel provozu

Nedostatek koordinace je patrný zejména poslední době, kdy se nabídka informací veřejného sektoru masivně rozšiřuje, čímž se stává nepřehlednou. Firmy, zabývající se profesionálně zpracováním dat si sice nakonec zdroje informací samy najdou, avšak mnohdy jsou nuceny reagovat ex post, až poté, je-li nějaký nový portál nebo webová stránka uvedena do provozu. Přitom příprava služeb s přidanou hodnotou je časově, kapacitně i finančně náročná. Pokud by zpracovatelé měli k dispozici nějaký ucelený katalog zdrojů informací veřejného sektoru případně i plány na rozvoj zpřístupňování dalších datových okruhů veřejného sektoru, určitě by mohli pracovat efektivněji a pružněji.

Nevyjasněné kompetence vedou k dublování, nekonzistentnosti, neúplnosti a neaktuálnost dat a rovněž stěžují komunikaci mezi veřejným sektorem a privátními subjekty. I když je do určité míry pochopitelné, že data veřejného sektoru se budou u různých institucí do určité míry překrývat a dublovat, přece jen by bylo praktičtější, kdyby kompetence za různé datové okruhy byly jednotlivým institucím jasně stanoveny, případně – pokud přebujelý resortismus neumožní dohodu – aby byl vytvořen „styčný důstojník“ pro centrální přijímání požadavků a řešení problémů. Vyjasnění kompetencí by prospělo i zkvalitňování dat a odstraňování chyb. V současnosti totiž často dochází k tomu, že pokud komerční zpracovatel odhalí chyby v datech, reportuje je sice správci zdroje, avšak nápravy se mnohdy nedočká, protože daná instituce přebírá data z jiného zdroje a necítí se být za ně zodpovědná.

Neexistence standardů a pravidel používání informací veřejného sektoru má negativní dopady na efektivnost procesu „re-use“. V praxi to znamená, že pro každý zdroj informací je nutné zpracovat podrobnou analýzu vstupních dat a naprogramovat unikátní proceduru pro jejich import. Nutno přiznat, že pro mnohé softwarové firmy je právě znalost tohoto know-how vítaným zdrojem příjmů, avšak v důsledku se jedná o plýtvání penězi klientů. Určitě by bylo přínosem, kdyby existovala jednotná integrační platforma alespoň pro systémy státní správy, standardy pro formáty souborů, XML tagy, aj. Nejednotnost regulí a podmínek, za kterých je možné informace veřejného sektoru používat pro „re-use“, pokrývá širokou škálu problémů od vymezení autorských práv a podmínek pro další použití poskytnutých informací přes technická pravidla provozu až po cenovou politiku. Jen při jejich jasném vymezení lze docílit rovných podmínek na trhu pro všechny soutěžitele v komerční sféře.

Jak může pomoci Iniciativa Czech PSI Watch?

Z neformální diskuze mezi poskytovateli komerčních informačních služeb využívajících informací veřejného sektoru vyplynulo, že po identifikaci nejdůležitějších problémů je na místě naznačit i určitá východiska vedoucí ke zlepšení stavu. Bude třeba zejména

- Formulovat problémy a požadavky různých zainteresovaných skupin (podrobněji než v tomto dokumentu, např. formou výběrového dotazníkového šetření u různých poskytovatelů služeb, jak se o to snažil např. projekt MEPSIR – www.mepsir.org)
- Provést procesní analýzu „re-use“ (zmapovat základní procesy při „re-use“, odhalit duplicity, neefektivní a brzdící kroky, odstranit nepotřebné části, najít styčné body při poskytování informací veřejného sektoru různými institucemi, navrhnout „hladný průběh“ aj.)
- Podílet se na návrzích věcného řešení, legislativy a metodiky (vyjadřovat se k návrhům zákonů, přinášet příklady z praxe, intenzivně komunikovat, aby bylo co nejvíce praktických poznatků a připravovaná opatření skutečně reagovala na potřeby praxe)
- Přispět k efektivnímu způsobu poskytování informací a k jejich racionálnímu využívání (využít zkušeností různých poskytovatelů komerčních služeb, zmapovat současná řešení např. pro odhalování duplicit a chyb, nové systémy uvádět do provozu až po oponentuře zástupců z praxe, navrhnout nové efektivní systémy a podílet se na jejich realizaci)

Závěr

Komerční sektor je tvůrcem hodnot a zdrojem ekonomického růstu. Podnikatelské aktivity v éře informační společnosti napříč všemi obory budou stále více závislé na informacích. Vytváření infrastruktury pro podporu a hladné fungování informačních systémů je významná aktivita, které se věnuje stále více firem – od systémových integrátorů a softwarových vývojářů až po informační agentury a poskytovatele obsahu. Pro efektivní poskytování těchto služeb nelze opomenout integraci již existujícího informačního obsahu, který byl vytvořen ve veřejném sektoru. To je ostatně smyslem směrnici 2003/98/EC o opakovaném využívání informací veřejného sektoru.

Cílový stav, kterého by se při vzájemné spolupráci komerčního, neziskového a veřejného sektoru mělo dosáhnout, je v podstatě velice prostý. Každý poskytovatel informačních služeb, který potřebuje informace veřejného sektoru, by asi chtěl

- mít přehled o datech, která jsou k dispozici a kdo je poskytuje (katalog informačních zdrojů a kompetencí k nim)
- mít možnost data si snadnou cestou „objednat“, tj. nikoliv jen získat jeden údaj, ale stáhnout či pravidelně stahovat data pro další zpracování (místo práva na „access“ dosáhnout práva na „re-use“)
- pokud zpracovatel narazí na chyby v datech, mít jedno centrální místo, kam bude chyby reportovat („styčný důstojník“)
- mít možnost zvolit si nejvhodnější způsob distribuce (formát souboru, přenosová cesta)
- znát jasné podmínky, za kterých je možno informace veřejného sektoru integrovat do dalších informačních systémů s přidanou hodnotou

Informace poskytli:

Česká kapitálová informační agentura, a.s.
EPMA – Agentura pro evropské projekty a management
Adastra, s.r.o.
Coface Intercredit Czechia, s.r.o.
Tovek, s.r.o.
a další.

V Praze dne 1. února 2006

Zpracovala:

Ing. Dagmar Vránová