

Jak a proč chráníme důvěrné statistické údaje

Povinnost dodržovat ochranu údajů, citlivých informací a dbát na **bezpečnost informačních systémů** má náš úřad danou ze zákona. Při tom se řídí i konvencemi doporučenými Eurostatem.



Možnost odhalení důvěrných údajů a chráněných informací reálně existuje.

Na úvod je třeba si položit otázku: Co vlastně chráníme? Za prvé důvěrné údaje a citlivé informace dané právním rámcem, za druhé bezpečnost informačních systémů. Vybalancovat ochranu soukromí respondentů a jejich obchodního tajemství není lehké. Například vládní organizace, vědecká a výzkumná komunita a ostatní odborná veřejnost požadují stále mohutnější a podrobnější informační toky. Možnost odhalení důvěrných údajů a chráněných informací však reálně existuje. Databázové systémy, datové sklady a metodika Da-

ta Miningu umožňují totiž velmi efektivně identifikovat statistické jednotky a jednotlivce.

Oblast metodiky ochrany důvěrných statistických údajů a řízení rizika prozrazení (SDC – Statistical Disclosure Control) úzce souvisí

s konkrétním účelem, ke kterému mají být uvolňovaná data použita. Od toho se odvíjí způsob uvolňování – například pro účely vědeckého výzkumu nebo publikaci statistických informací na CD/DVD, vzdáleným přístupem přes webové

rozhraní do veřejných databází nebo pro práci v zabezpečené datové laboratoři. Obecně platí, že existuje určitá akceptovatelná úroveň rizika prozrazení jako daň, kterou platíme za poskytování statistických informací. Naproti tomu strategie úplného odmítnutí rizika vede ke znehodnocení informačního obsahu publikovaných dat.

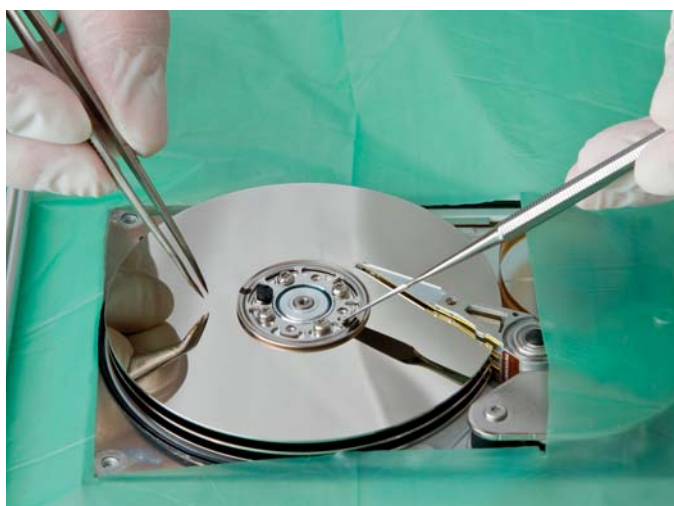
Příklady našich vnitřních norem

Směrnice č. 1/2008 stanoví zajištění ochrany dat v informačním systému evidence obyvatelstva a v databázi fyzických osob. **Směrnici č. 5/2008** se vydávají zásady pro poskytování informací z Českého statistického úřadu externím uživatelům.

Existence jedinečnosti

Pro účely praktické ochrany důvěrnosti v konkrétním šetření je nutné vymezit přesně pojmy mikro-

statistika a makrostatistika včetně odpovídajících metadat. V mikrostatistice mohou být statistickými subjekty osoby, domácnosti, podniky atd., jinak řečeno - mikrostatistika je složena z informací na úrovni respondenta. Každému subjektu přináší jednotlivý datový vektor (záznam, datová množina) kvalitativních nebo kvantitativních proměnných. I přesto, že odstraníme formální identifikační údaje, jako jsou například jméno a adresa, lze odhalit konkrétní údaj díky jedinečnosti v souboru. Ta je určena velikostí populace (vzorku), stupněm členění, například geografických informací, počtem a detailností poskytovaných charakteristik za každou jednotku v databázi apod. Makrostatistika je charakterizována jako množina tabulek individuálních dat, kde je každá buňka definována jako průsečík proměnných (atributů), ať již kvalitativních nebo kvantitativních. Ve statistice jsou běžné různé druhy tabulek. Často se jedná o tabulku četností (kontingenční tabulka) a tabulku agregací (tabulka hodnot veličin).



Snížení reidentifikace respondentů

Zatímco μ -ARGUS je specializovaný softwarový nástroj pro oblast ochrany mikrodat, který v případě použití metody globálního překódování způsobí transformaci několika kategorií proměnné do jedné (Collapsing categories of identifying variables). **Používá se i metoda lokálního potlačení** (Local suppression of individual scores on individual variables), aby byly jedna nebo více proměnných (které mohou být v kombinaci nebezpečné) nahrazeny symbolem „missing value“. Obě metody vedou k určité ztrátě informace snížením její detailnosti, ale tím **sníží riziko reidentifikace respondentů**. Naproti tomu systém τ -ARGUS produkuje zabezpečené tabulky. Účelem systémů ARGUS je minimalizovat riziko prozrazení za současné maximalizace vypovídací schopnosti výsledného zpracování.

Systémy ARGUS

Pro praktické provedení primární a sekundární ochrany důvěrných statistických údajů jsou u nás aplikovány softwarové systémy, které vyvinuli naši pracovníci – PRIMUT

a OKUSTA. Dále se u nás využívají programové systémy ARGUS, které vyvinuli vědci z předních univerzit a odborníci ze statistických úřadů z Nizozemska, Velké Británie, Itálie, Německa a Španělska. Výsledkem jejich společného úsilí jsou systémy,

které využívají osvědčené základní metody ochrany důvěrnosti statistických dat (SDC – Statistical Disclosure Control Methods).

Dva programové systémy (μ -ARGUS, τ -ARGUS) sice řeší na první pohled společnou pro-

blematiku ochrany postavenou na dvou základních metodách (Global Recoding, Local Suppression), ale ve skutečnosti jde o velmi odlišnou problematiku ochrany mikrodat (jakožto primárních záznamů), resp. ochrany tabulkových dat (jakožto agregovaných soustav).

Naše instituce má zájem o maximální zabezpečení důvěrných statistických údajů s tím, že metody ochrany chceme přizpůsobit konvencím doporučeným Eurostatem. Proto bylo přistoupeno k prověření systémů μ -ARGUS a τ -ARGUS a jejich postupnému nasazení v podmínkách statistických šetření prováděných našim úřadem. O nákupu systému τ -ARGUS bude rozhodnuto až po jeho odzkoušení. Součástí jeho demonstračního balíku nejsou totiž některé speciální šifrovací metody (XPRESS nebo CPlex). Jeho silnou stránkou je výpočet optimalizace rizika versus informační ztráty hyperkubickou metodou, což jiné softwarové nástroje, postavené na jednodušších metodách, nedovedou.

Daň za informace

Každodenní práce s důvěrnými údaji a citlivými informacemi může vést zaměstnance i nechtěně k aktivitám, které by mohly být v rozporu s uvedeným právním rámcem. Náš úřad se proto snaží své pracovníky maximálně chránit. V této souvislosti jsou rozpracovány základní vnitřní normy, které upravují pravidla s nakládáním dat. Jsou to: Bezpečnostní politika ČSÚ, Systém ochrany důvěrnosti a bezpečnosti informací, Metodika provádění analýzy rizik, Soubor krizových scénářů a také směrnice, řády a rozhodnutí.

Legislativní rámec ochrany a bezpečnosti informací

Pro celou Evropskou unii platí Nařízení Evropského Parlamentu a Rady č. 223/2009 ze dne 11. 3. 2009 o evropské statistice, které vymezuje problematiku statistické důvěrnosti. V české legislativě tomu odpovídá **zákon č. 89/1995 Sb., o státní statistické službě**. Ochrana osobních údajů respondentů je zaručena **zákonem č. 101/2000 Sb., o ochraně osobních údajů**. Legislativní prostor je dále vymezen **zákony č. 106/1999 Sb., o svobodném přístupu k informacím**, a **zákon č. 412/2005 Sb., o ochraně utajovaných informací a o bezpečnostní způsobilosti**. Z oblasti eGovernmentu jmenujme alespoň dva právní předpisy: **zákon č. 365/2000 Sb., o informačních systémech veřejné správy**, a **zákon č. 440/2004 Sb., kterým se mění zákon č. 227/2000 Sb., o elektronickém podpisu**.

Speciálním zákonem významným pro činnost ČSÚ v letošním roce, ze kterého je zřejmá přísná ochrana sbíraných a zpracovávaných údajů, je **zákon č. 296/2009 Sb., o sčítání lidu, domů a bytů v roce 2011**.

Vladimír Holý
oddělení ochrany důvěrnosti
a bezpečnosti informací