

Kvalita dat a lidský faktor v podnikových informačních systémech

Účinnost každého informačního systému je přímo závislá na datech¹, se kterými pracuje a také na jejich kvalitě. Co je vlastně kvalita dat? Podle jedné z definic jde o taková data, která jsou užitečná a použitelná pro své „spotřebitele“ [1]. Narážíme tedy na relativně často opomíjený problém – informační systém je zde pro své uživatele a právě na nich závisí jeho úspěch či neúspěch, protože uživatelé obvykle stojí jak na vstupech dat do systému (pokladní „markující“ zboží na pokladně) tak i na výstupu (manažer plánující v BI reklamní kampaně na základě dat o prodeji). Na lidech tedy nejvíce záleží, s jakými daty bude systém pracovat a jak budou výstupy interpretovány²

Pro účely se esejí nebudu zabývat chybným zpracováním dat v samotném informačním systému³. Za tohoto předpokladu je pak největším zdrojem nepřesností právě „lidský faktor“ na vstupu do IS. Proto je právě na vstupu dat do IS nasazována celá plejáda technologií typu čárového kódu (oběh materiálu, součástek), RFID (jak materiál, tak i třeba pohyb pracovníků po budově), číselníků (seznamy ulic, PSC, seznamy léků a kontraindikací), kontroly vůči správným datům či kontrolním součtům (platební karty, rodná čísla). Nemusí vždy jít o dokonalou eliminaci chyb (u platebních karet jen chyba v jedné číslici), ale daná opatření zlepšují kvalitu dat na „dostatečně dobrou pro dané použití“ [1].

Zajímavější ale je, pokud informační systém musí podporovat při prvotním zadání dat složitější proces – dejme tomu bankovní systém s prvky CRM při založení nového účtu (tj. zaevidování všech informací vyžadovaných ČNB, nastavení pravidelných úkonů jako je odesílání výpisů, nabídka doplňujících bankovních produktů atd.) Naplno jsou využity číselníky, kontrolní součty, formuláře a různí „přívodci“. Co se ale stane, pokud firemní proces požaduje něco, co IS neumí? (...) [2 věty]

Zde se projevuje hned několik podstatných faktorů: kvalitu dat ovlivňuje školení zaměstnanců a (...) [odstavec v poznámce pod čarou] způsob, jakým informační systém podporuje stanovené procesy a způsob, jakým informační systém nakládá s přístupem k datům. Nejde jen o utajení, některé údaje jsou pouze interního charakteru, a proto je zejména u CRM systémů důležité výrazně vyznačit (databáze, GUI), které zadávané údaje jsou pro veřejné a které ne. Jinak dojde k obtížně odstranitelné „kontaminaci dat“, která pochopitelně velmi snižuje jejich použitelnost v rámci informačního systému při jakémkoliv automatickém zpracování.

Na příkladu jsem vysledoval i určitou představu, která je až doposud čitelná i z mojí práce (zcela záměrně pro zjednodušení). Informační systém je uživateli v mnoha případech vnímám jako jakási černá skříňka, ve které pracují další lidé a jakákoliv zadaná data a provedené úkony přetřídí a kontrolují. Opak je pravdou – při integraci současných IS téměř každý uživatel přímo či nepřímo ovlivňuje data všech ostatních v systému a drtivá většina jeho akcí neprochází žádnou další kontrolou (zejména v CRM, ECM, v menší míře v SCM/APS). Takováto integrace je naprosto žádoucí a nezbytná (ostatně právě kvůli zpřístupnění dat se podnikový IS zavádí), přidává však navíc další požadavky na (kvalitní) data: musí je být schopen zadat každý zaměstnanec a IS musí brát v potaz značně rozdílnou kvalifikaci zaměstnanců napříč firmou a přizpůsobit jí jak terminologii, tak i vynucenou kontrolu dat⁴. V mém extrémním příkladu je očekávání další kontroly člověkem naprosto zřejmé, nikdy k ní ale nedocházelo.

¹Pro účely práce používám pro data i **informace** souhrnné označení „data“.

²Interpretace a vnímání dat v informačním systému je zajímavým způsobem popsána v jedné z povídek – The Evitable Conflict – v knize Isaaca Asimova I, Robot. Oproti fikci ale zatím v realitě neumí „Stroje“ do svého uvažování zahrnout nespolutracující uživatele.

³Například zaokrouhlovací chyby při použití výpočtů s plovoucí čárkou, ztráta dat.

⁴Nemusí jít o kontrolu všech dat, při dostatečně velkém vzorku postačuje namátková. Ověřování je nezbytné u dat nejasné kvality – od brigádníků, jiných firem, ...

Zmiňovaný proces v bance byl na vstupu dat a napomáhá vnímání IS jako černé skříňky. Pokud ale budeme chápat IS jako součást organizace, dochází ke vstupu dat prakticky neustále. Ve chvíli, kdy si tuto situaci uživatelé systému uvědomí, změní pohled na kvalitu dat jako celek. Už nejsou tvořena mytickými, (ne)dokonalými pracovníky z neznámého IT/vedoucího oddělení, ale někým, koho známe (a jehož chyby až příliš dobře známe). Zde se projeví metody udržování kvality dat v plné míře a v kvalitě začíná hrát roli i nový prvek – čas. Přímo odstrašující příklad je portál gov.cz – proběhlo úvodní zadání dat, pravděpodobně i kontrola a data jsou zadána ve zpracovatelných formátech. K čemuž už ale nedochází, je jejich pravidelná aktualizace. Původně kvalitní data postupem času ztrácí svoji přesnost, přestože každá informace má přiřazenou zodpovědnou osobu – není totiž nijak sledován stav již existujících dat. Celý informační systém tak ztrácí vypovídací hodnotu.

V praxi se stane ještě něco horšího. Zejména v rozsáhlejších IS (nemusí jít o celý systém, může se jednat o vybrané moduly, což je možná ještě horší) platí „teorie rozbitých oken“ [4] – pokud budu vědět, že v systému zodpovídám za určitá data, která (beztrestně) neudržuji „kvalitní“, pak budu totéž očekávat i od ostatních. Dojde k jakémusi kolapsu důvěry⁵ v data v informačním systému⁶. Tuto teorii je možné aplikovat i na systémy, které fungují jinak, než je požadováno firemními procesy⁷. Uživatelé jednoduše části systému začnou ignorovat (nebo používat jen formálním způsobem) a navrátí se k jiným, méně efektivním metodám předávání informací (například emaily obcházející ECM, ignorování úkolů obchodníky v CRM systému).

Dalším problémem je přesnost dat – u kvalitních dat je jednoznačně udána⁸. I drobná odchylka v datech na začátku (způsobená například jinými zvyklostmi v pobočkách v různých zemích[3]) způsobí obrovské rozdíly ve výsledné agregaci.

Do určité míry je možné připodobnit podnikový informační systém s takovýmto množstvím vad k centrálně plánované ekonomice. Také je zdánlivě řízený shora (v rámci jakési pyramidy), zde nástroje BI, MIS, pracující „nad všemi daty“. Prakticky je ale pyramida naprosto převrácená – uživatelé nedůvěřují datům, na kterých jsou založena rozhodnutí vedení (případně jsou data chybná), vedení nedůvěřuje datům od zaměstnanců, předávání informací (informační funkce peněz) je omezené a v každém kroku systému si uživatelé tvoří drobné rezervy oproti údajům v IS (rezervy oproti výrobnímu plánu), což vede ke značné neefektivitě.

U úvah o CPE se nabízí ještě jeden problém kvalitních dat v informačních systémech. Mnohé firemní procesy a znalosti obsahují nepsaná data „tady a teď“⁹, která platí jen pro konkrétní oblast, časový okamžik nebo třeba jen pro konkrétní osobu (to samé ostatně platí i pro samotné firemní procesy). V úvodu jsem ztotožnil data a informace pod jednotný pojem data a zde je myslím nutné toto ztotožnění připomenout. Kvalitu dat–informací v systému ovlivňuje i (ne)možnost zaznamenat právě tyto „lokální“ znalosti. V případě nemožnosti uživatelé začnou IS obcházet (například u CRM systému si obchodníci začnou budovat vlastní databázi klientů s vlastními

⁵Právě portal.gov.cz obsahuje data v takové kvalitě, že je nejen nepoužitelný pro vnější uživatele – občany, ale je tak vnímám i úředníky, kteří ignorují jakékoliv upozornění na chyby. Příčinou nemusí být jen nezáměr úředníků. Je možné, že změnit jakýkoliv údaj v systému představuje sérii administrativně velmi pracných kroků, které svou náročností převyšují „dobrý pocit“ z oprav jako je nová emailová adresa některých odborů.

⁶Stejně důsledky mohou mít výpočetní chyby, „dětské nemoci“ při zavádění nebo jen nepříliš důvěryhodné vypadající GUI

⁷Příkladem budiž „mini“ ERP systém jisté zážitkové agentury. První verze umožňovala rezervovat zážitek pomocí certifikátu přes web, spravovat rezervace, sestavovat objednávky u dodavatelů a sledovat spokojenost zákazníků. V praxi se ale ukázalo, že postup *dodavatel zadá volné termíny* → *zákazník si vybere termín* → *obchodník potvrdí u dodavatele* je neproveditelný. Většina dodavatelů nikdy nevyplnila, ve kterých termínech zážitky realizují.

⁸Kvalitní data nemusí být absolutně přesná, ale musí být zřejmé, jak jsou nepřesná. Mírně ne-IT příklad je hypoteční krize v USA. Několikerym promísením různých hypoték dohromady přes několik finančních instrumentů ztratilo jakékoliv udávané riziko vypovídací hodnotu.

⁹Zaznamenání a zpracování těchto dat je v rámci CPE neuskutečnitelné a představuje jeden z hlavních problémů. Viz statě od F. A. Hayeka, L. von Missese (napadá mě Ekonomická kalkulace v socialistické společnosti)

25. února 2011

poznámkami, pokud není možné uvést pro ně důležité podrobnosti – třeba kam chodí významný klient na golf). Na druhou stranu, nemělo by dojít k opačnému extrému – zahlcení celého informačního systému množstvím dat, která nemají vypovídací hodnotu pro nikoho jiného než jen několik málo uživatelů a která budou pravděpodobně velmi „denormalizovaná“. Oba případy opět vedou ke ztrátě původní funkce IS.

Závěrem bych rád zdůraznil myšlenku celého eseje – kvalita dat v informačním systému závisí na jeho uživatelích nejen při tvorbě nebo zpracování dat, ale také závisí na tom, jak samotní uživatelé data v informačním systému, často subjektivně, vnímají – nejen z hlediska správnosti, ale i z hlediska vlastních schopností s daty pracovat, porozumět jim a interpretovat je. Udržování kvalitních dat je dlouhodobý, nikdy nekončící a nikoliv jednorázový proces a měl by být zastoupen v každém prvku jakéhokoliv podnikového informačního systému.

Reference

- [1] Cinzia CAPPIELLO, Chiara FRANCALANCI, Barbara PERNICI, *Data quality assessment from the user's perspective*. IQIS '04: Proceedings of the 2004 international workshop on Information quality in information systems, June 2004
- [2] Diane M. STRONG, Yang W. LEE, Richard Y. WANG, *Data quality in context*. Str. 103 – 110, Communications of the ACM archive, Volume 40, Issue 5 (May 1997)
- [3] Leo L. PIPINO, Yang W. LEE, Richard Y. WANG, *Data Quality Assessment*. Str. 211 – 218, Communications of the ACM archive, Volume 45, Issue 4 (April 2002)
- [4] James Q. WILSON and George L. KELLING *Broken Windows*. The Atlantic Monthly, March 1982, http://www.manhattan-institute.org/pdf/_atlantic_monthly-broken_windows.pdf