VYSOKÁ ŠKOLA EKONOMICKÁ V PRAZE

FAKULTA INFORMATIKY A STATISTIKY

DIPLOMOVÁ PRÁCE

2007 PAVEL SKALICKÝ

Vysoká škola ekonomická v Praze

Fakulta informatiky a statistiky

Katedra informačních technologií

Studijní program: Aplikovaná informatika

Obor: Informační systémy a technologie

Diplomant: Pavel Skalický

Vedoucí diplomové práce: Ing. David Slánský, Ph.D.

Recenzent: Doc. Ing. Jan Pour, CSc.

TÉMA DIPLOMOVÉ PRÁCE

DATOVÁ KVALITA ŘEŠENÍ BUSINESS INTELLIGENCE & MASTER DATA MANAGEMENT

-

HOUSEHOLDING

školní rok 2006/2007

Prohlášení

Prohlašuji, že jsem diplomovou práci zpracoval samostatně a že jsem uvedl všechny použité prameny a literaturu, ze kterých jsem čerpal.

V Praze dne 4. srpna 2007

…….

podpis

Abstrakt

Diplomová práce se zabývá tématem householdingu. Householding je specifickou oblastí řešení datové kvality a je dále provázán s disciplínami Master Data Management, Customer Data Integration a dalšími. Jedná se o proces rozpoznávání ekonomicky významných skupin subjektů, kterých se využívá zejména v marketingu a risk managementu.

Cílem diplomové práce bylo vytvoření uceleného pohledu na householding. Ten zahrnuje nalezení či vytvoření jednotné definice, nastínění v praktickém životě využívaných možností použití, dále v identifikaci případných typů householdingu a zmapování postupu budování, včetně problémů provázejících tento proces. Takto je i práce strukturována. Cíle bylo dosaženo na základě zpracování a interpretaci existujících zdrojů zastřešujících tuto tématiku. Dále byly pro účely práce realizovány osobní konzultace se zástupci společnosti Adastra Corporation.

Přínosem práce je shrnutí a vlastní rozvedení důležitých aspektů householdingu do veřejně dostupného materiálu, který do té doby chyběl.

Klíčová slova: householding, household, HH, Business Intelligence (BI), Master Data Management (MDM), datová kvalita (DQ), identifikace skupin, Customer Data Integration (CDI), Customer Relationship Management (CRM)

Abstract

This thesis is focused on householding, a specific area that deals with Data Quality Solutions and is further interlaced with Master Data Management, Customer Data Integration and other disciplines. This process is about recognition of significant economical groups which are especially used in Marketing and Risk Management.

The aim of this thesis was to create an integrated view of householding. This comprises of creating a unified definition, outlining common usage, identifying possible types of householding and chartering progress construction including problems that are following this process. This is also how the thesis is structured. The aim was reached by elaboration and interpretation of available sources that are concerned with this topic. Furthermore, the aim was personally consulted with Adastra Corporation representatives.

The contribution of this work is in summarization and inherent expansion of important aspects of householding into a public access publication which has been missing until then.

Keywords: Householding, household, HH, Business Intelligence (BI), Master Data Management (MDM), data quality (DQ), group identification, Customer Data Integration (CDI), Customer Relationship Management (CRM)

Poděkování

Rád bych na tomto místě poděkoval panu Ing. Davidu Slánskému za vedení diplomové práce a za přínosné rady poskytnuté během její tvorby.

Dále bych chtěl poděkovat panu Vladimíru Kyjonkovi za odborná doporučení a čas, který mi během konzultací věnoval.

Své poděkování bych chtěl vyjádřit i svým rodičům za trpělivost a vstřícnost, kterou mi při tvorbě diplomové práce i během celého studia projevovali.

Osnova

1 Úvod ... 3

1.1 Cíle práce ... 3

1.2 Vstupy ... 4

1.3 Postup řešení ... 4

1.4 Výstupy ... 6

2 Co je householding? ... 7

2.1 Definice householdingu ... 7

2.1.1 Adastra ... 7

2.1.2 Odborné články a knihy ... 8

2.1.3 Významní vendoři ... 9

2.1.4 Shrnutí a definice použitá pro účely práce ... 10

2.2 Všeobecné využití householdingu ... 11

3 Typy a přístupy k householdingu ... 14

3.1 Typy ... 14

3.1.1 Osobní householding ... 14

3.1.2 Komerční householding ... 15

3.1.3 Kombinace osobního a komerčního householdingu ... 18

3.2 Přístupy ... 19

3.2.1 Maximalistický přístup ... 20

3.2.2 Minimalistický přístup ... 20

3.3 Závěrečné shrnutí kapitoly ... 21

4 Budování householdingu ... 22

4.1 Postup budování významnými společnostmi ... 24

4.1.1 Adastra ... 24

4.1.2 Harte-Hanks Data Technologies ... 31

4.1.3 Business Objects / Firstlogic ... 34

4.2 Problémy při budování ... 41

4.2.1 Obtížnost definice ... 42

4.2.2 Identifikace entit ... 43

4.2.3 Význam entit ... 44

4.2.4 Příslušnost k více householdům ... 45

4.2.5 Práce na základě odhadu (pravděpodobnosti) ... 46

4.2.6 Zákon 101/2000 – nakládání s osobními údaji ... 47
4.2.7 Určení adresového bodu ... 49
4.2.8 Stárnutí dat i householdu ... 50

4.3 Podmínky nutné a postačující ... 51
4.3.1 Kvalitní externí číselníky ... 51
4.3.2 Kvalitní a čistá data ... 55

4.4 Závěrečné shrnutí kapitoly ... 61

5 Závěr ... 63

6 Seznam použité literatury a informačních zdrojů ... 65

6.1 Literatura ... 65

6.2 Články z časopisů ... 65

6.3 Internetové zdroje ... 66

7 Seznam obrázků a tabulek ... 68

7.1 Obrázky a grafy ... 68

7.2 Tabulky ... 68

8 Seznam zkratek a terminologický slovník ... 69

1 Úvod

Diplomová práce se zabývá specificky zaměřenou tématikou – householdingem. Tento pojem spadá do oblasti datové kvality a je dále provázán s disciplínami, jako jsou Master Data Management, Customer Data Integration a další. Využití lze nalézt mj. v řízení vztahu se zákazníky a v bankovnictví, což jsou oblasti, o které se osobně zajímám. Téma práce bylo zvoleno také z důvodů subjektivního pocitu nedostatečného informačního pokrytí této problematiky, alespoň co se českých poměrů týče.

Konkurenční prostředí firem neúnavně roste, jednotlivé trhy se stále více nasycují. Pro udržení zákazníků je v dnešní době nezbytná jejich dobrá znalost, a to včetně jejich potřeb. V práci bude uveden přístup, díky kterému mohou firmy identifikovat jedince v kontextu skupiny – nebo lépe řečeno identifikovat skupiny, které mají něco společného. Tato znalost může být silným nástrojem v boji s konkurencí.

Informační systémy a datové sklady obsahují nespočetné množství dat, tak proč nevyužít lépe jejich potenciálu a nenechávat je vydělávat?

1.1 Cíle práce

Hlavním cílem práce je, na základě zmapování problematiky householdingu, vytvořit ucelený pohled na tuto tématiku. Ten spočívá v nalezení či vytvoření jednotné definice, v praktickém životě využívaných možností použití, identifikaci případných typů a zmapování postupu budování, včetně problémů provázejících tento proces.

Dílčím cílem je nalezení významných autorů (společností), kteří se touto problematikou zabývají a porovnání jejich přístupů s přístupem využívaným ve společnosti Adastra.

Dosažení těchto cílů předpokládá především práci s existujícími zdroji informací a dále osobní konzultace se zástupci společnosti Adastra.

1.2 Vstupy

Hlavními vstupy práce budou zejména internetové zdroje, na jejichž analýze, zpracování a interpretaci, bude práce založena. Pro získání teoretických znalostí v oblasti datové kvality poslouží publikace Englishe[1], Dyche[2], Novotný[3] a také prezentace z celosvětové konference TDWI (Orlando 2006). Jako zdroj informací týkajících se Householdingu budou využity materiály ze společnosti Adastra.

Nezanedbatelným pramenem informací jsou též přednášky Katedry informačních technologií Vysoké školy ekonomické v Praze, zejména z předmětu Business Intelligence, a dále také osobní konzultace se zástupci společnosti Adastra.

1.3 Postup řešení

Jak již bylo uvedeno, práce je založena na zpracování elektronických zdrojů. Na počátku jsem si tedy vytvořil plán, podle kterého bude práce tvořena. Ten obsahuje:

- identifikaci zdrojů
- výběr klíčových slov pro vyhledávání
- samotné vyhledávání
- zpracování a interpretace informací

Identifikace zdrojů

Pro vyhledání relevantních zdrojů byl zpočátku práce využit vyhledávací systém www.google.com. Toto předpřípravné vyhledávání posloužilo, mimo jiné, k získání velkého počtu klíčových slov, která byla dále využita.

---

[1] English, L.P.: Improving Data Warehouse and Business Information Quality: Methods for Reducing Cost and Incresing Profits, 1999

[2] Dyché, Jill, Levy, Evan: Customer Data Integration: Reaching a Single Version of the Truth, 2006

[3] Novotný, O., Pour, J., Slánský, D: Business Intelligence, 2004

Další zdroje přehledně ukazuje následující tabulka:

Odborné servery	Významní autoři	Významní vendoři
dmreview.com	Jill Dyche	Adastra
twdi.org	Aaron Zornes	Trillium
the-CDI-institute.com	Bill Inmon	Siperian.com
proquest.umi.com (ProQuest 5000)	Claudii Imhof	Purisma.com
…	David Loshin	Hyperion
	…	Cognos
		Business Objects / FirstLogic
		Harte-Hanks / Trillium

Tabulka č.1: Primární zdroje diplomové práce

Klíčová slova

Klíčových slov použitých při vyhledávání bylo mnoho. Některá vznikala systematicky, jiná z momentálních nápadů. Prostudování materiálů navíc postupně rozšiřovalo tento seznam.

Některá klíčová slova: householding, household, CDI (customer data integration), CRM (customer relationship management), MDM (master data management), data quality, corporate householding, BI (business intelligence), DWH (datawarehousing), marketing, risk management, group identification a další.

Jednotlivá klíčová slova byla navíc různě kombinována a doplňována.

1.4 Výstupy

Závěrem práce by mělo být zodpovězení následujících otázek:

Co je to householding?
Kdo píše o householdingu a co tím myslí?
Jaké jsou typy householdingu?
Jakým způsobem se jednotlivé typy využívají?
Jaký je postup budování?

Tabulka č.2: Cíle diplomové práce v otázkách

2 Co je householding?

Nalezení jednotného vymezení pojmu „householding“ není jednoduché. Nejednoznačnost definice vychází z různého chápání tohoto pojmu. Úvod práce se zabývá pohledy světových autorů na householding a tím, jakým způsobem jej vymezují.

Někteří lidé mohou householding chápat jako *„skupinu lidí se stejným jménem a adresou“*. Tato „definice“ je však pro účely práce nedostačující.

2.1 Definice householdingu

Jednotlivé definice jsou získávány z veřejných zdrojů, zejména vyhledáváním na internetu a z odborných knih a článků. Podle toho je také tato kapitola rozdělena do čtyř částí. V závěrečné části se nachází shrnutí jednotlivých definic a stanovení té, která bude pro mou práci určující.

2.1.1 Adastra

Společnost Adastra v praxi nejčastěji pracuje s householdingem jako s identifikací netriviálně definovaných skupin fyzických osob, mezi kterými existují vztahy různého typu. Tyto vztahy indikují příslušnost k domácnosti v kontextu bydliště a rodiny a předpokládají ekonomickou provázanost jejích členů.

Jednodušeji řečeno se jedná o seskupování klientů, kteří mají něco společného. Adastra se však neomezuje pouze na rodinu. Household chápe jako zástupný termín, který může být totožný například s **domácností, rodinou, zájmovou či profesní skupinou, ale také komerčními subjekty.** Komerční subjekty mohou mít mezi sebou vztahy vlastnické nebo partnerské. Lidé mohou být ve vztahu ke komerčnímu subjektu vedeni jako zaměstnanci, členové statutárních orgánů nebo například vlastníci. Tyto konkrétní vztahy jsou jen příkladem toho, co může být předmětem householdingu.

[Adastra]

2.1.2 Odborné články a knihy

Jacob Sacolick – specialista pracující v Harte-Hanks Data Technologies

Householding je proces organizování zákaznických účtů a různých dat o zákaznících. Mluvíme-li o HH, obvykle používáme tři typy entit:

1) household (domácnost)
2) jednotlivec
3) zákazník

Household (domácnost) může být myšlen jako rodina. V širším slova smyslu je Household chápán jako určitá jednotka vytvářející rozhodnutí. Za jednotlivce se považují fyzické osoby. Zákazníkem jsou jednotlivci a ostatní entity, které mají vlastní účet. Účet je vztah mezi zákazníkem a firmou, který je udržován finančními interakcemi.

Domácnosti, jednotlivci a zákazníci se neorganizují sami o sobě. Také nechodí s identifikačními značkami. Proto není jejich slučování jednoduché. Existují dvě základní cesty, jak mohou firmy jednotlivce poznávat a přiřazovat do householdů. První cesta je dotazováním a druhá vyhledáváním v datech, a to při vzniku či používání účtů.

Z toho, co bylo v této části zmíněno, lze shrnout pracovní definice určující HH:

- household je jedinečný identifikátor připojený k účtům, které jsou příslušné k tomuto householdu
- jednotlivec je jedinečný identifikátor připojený k účtům, které souvisí s tímto jednotlivcem
- zákazník je jednotlivec, který má „otevřený“ účet ve firmě
- příležitostí je jednotlivec, který nemá účet ve firmě, nebo ho má pouze „uzavřený“

[Sacolick]

Greg Mancuso and Al Moreno – ředitelé společnosti Sinecon, specializující se na konzultaci v oblasti BI, datovou integraci a design řešení BI/DW architektur.

Householding spočívá ve slučování dat z mnoha zdrojů. Jedná se o jednoznačné identifikování skupiny datových záznamů z jednoho zdrojového systému a jejich vztah k jiným datovým záznamům z jiného zdrojového systému.

Householding je proces identifikování souvisejících informací z mnoha zdrojů a jejich uložení do datového skladu tak, aby všechny informace související s danou skupinou byly dostupné při vyhledávání této určité skupiny.

V dnešní době se householdingem rozumí jakékoliv slučování (shlukování) informací o dané osobě, rodině, domácnosti či firmě. Může jím být tedy například rodina, nepříbuzní lidé žijící na stejné adrese, členové klubu, zaměstnanci firmy, její dceřinné společnosti či velké korporace.

[Mancuso]

Jill Dyché

Householding je konsolidace dat o zákaznících za účelem uspořádání jednotlivců do skupin (rodin), ve kterých žijí. To umožňuje firmám být více uvážlivý při komunikaci se zákazníky a také přesněji utvářet profil jednotlivců ve vztahu k ostatním.

[Dyché]

2.1.3 Významní vendoři

Business Objects – Firstlogic

HH je proces, ve kterém organizace hierarchicky strukturuje svá data za účelem získání kontextově citlivého pohledu na zákazníky, dodavatele, pacienty atd. Tyto hierarchické stromy jsou dále porovnávány mezi sebou navzájem a hledají se „větve“, které mají něco společného. Cílem HH je optimalizace obchodních operací.

[Firstlog]

SAS – DataFlux

Společnost DataFlux chápe HH jako proces konsolidace záznamů o zákaznících do identifikovatelných zákaznických skupin. Při tomto procesu používá sofistikované technologie využívající fuzzy logiku.

Bohužel více o definici HH na svých stránkách neuvádí.

[Flux]

2.1.4 Shrnutí a definice použitá pro účely práce

Při hledání definic householdingu jsem narazil na poměrně velké množství těch, které chápou HH pouze jako skupinu lidí se stejným jménem a adresou. To je pouze drobná část jistě komplexnější tématiky. Kvalita či podrobnost získaných definic nebyla mnohdy vyhovující.

Householding je většinou definován jako slučování „čehosi“. Jednotliví autoři se však liší v tom, co se vlastně slučuje. Zdali se jedná o data či záznamy (informace o konkrétních lidech, nebo jiných skupinách), nebo o určité konkrétní skupiny lidí. Adastra zjednodušeně definuje Householding jako proces seskupování klientů, kteří mají něco společného. Jde tedy o slučování konkrétních osob do skupin. Firstlogic, nebo také pánové Mancuso a Al Moreno, naopak nahlížejí na householding jako na slučování jednotlivých záznamů o lidech.

Dalším podstatným rozdílem v chápání householdingu je slučování či naopak rozdělování HH od čištění, případně konsolidace dat. Rozdílnost definice Adastry a většiny ostatních spočívá v tom, že Adastra odděluje proces konsolidace záznamů od vlastního householdingu. Householding chápe jako samostatnou disciplínu, která je následným krokem až po samotné unifikaci záznamů, tedy čištění a konsolidaci dat.

Jak je tedy patrné, možných pohledů a definic HH je více. **Pro potřeby této práce bude householding chápán jako proces rozpoznávání ekonomicky významných skupin subjektů.**

Za *subjekt* je považována fyzická osoba, právnická osoba nebo jakákoliv jednotka, se kterou má smysl dále pracovat. Více o konkrétních subjektech je popsáno v kapitole typy HH [3.1].

Ekonomicky významných znamená, že danou skupinu má smysl a význam sledovat pro dosažení konkrétních business cílů. Bez jasného stanovení business cílů postrádá budování HH smysl. Business cílem může být např. snížení počtu úvěrů osobám, které mají v rodině velké dlužníky se špatnou platební morálkou, nabízení produktů pro celou rodinu, nabídnout firmám speciální produkt pro jejich zaměstnance atd. (více viz [2.2]).

V této práci bude také zmíněn pojem „household“. Tímto pojmem se rozumí již určitá konkrétně definovaná skupina subjektů. Naproti tomu „householding“ je pro účely této práce chápan jako disciplína, proces, a v některých částech také jako činnost.

2.2 Všeobecné využití householdingu

Možných praktických použití householdingu je mnoho. Vpodstatě existují dvě významné oblasti, ve kterých je HH nejběžněji využíván. Jedná se o marketing a řízení rizik. HH by měl být vždy v souladu s business cíly. Mezi konkrétní cíle, viz [Vavruška], může například patřit:

- *lepší řízení rizik* - snížení počtu úvěrů osobám, které mají ve své rodině velké dlužníky se špatnou platební morálkou
- *přesnější marketingové zacílení* - poznání a ohodnocení existujících zákazníků a jejich potřeb
- *lepší ohodnocení existujícího zákazníka*
- *získání nových zákazníků*
- nabízení speciálních produktů firmám pro jejich zaměstnance
- nabízení produktů celým rodinám

Znalost rizika klienta je zásadně důležitá pro banky, úvěrové instituce a leasingové firmy. Neznalost toho, že klienti patří do jedné domácnosti, zvyšuje pravděpodobnost, že partnerovi rizikového klienta by mohl být, v rozporu se záměrem finanční instituce, poskytnut úvěr.

Podle Jindřicha Vavrušky lze využití HH nalézt v marketingu tím, že: „*Při marketingových kampaních směřují nabídky pouze na jednoho člena domácnosti, případně jsou diverzifikovány podle role jednotlivce v domácnosti. Tím se jednak šetří náklady na kampaně, jednak se předchází frustraci klientů z vícenásobných nabídek téhož produktu nebo služby.*“

„Při alespoň přibližné znalosti domácnosti lze zahrnout do nabídky i takové produkty, které pro klienta samotného nemají význam, avšak mohou oslovit jiného člena domácnosti.“

[Vavruška]

Znalost příslušnosti jednotlivců, z hlediska struktury domácnosti, umožňuje firmě přesněji vyhodnotit hodnotu každého zákazníka. Typickým příkladem mohou být manželé se samostatnými účty. Manželka na mateřské dovolené dostává na svůj účet dvanáct tisíc korun na běžné výdaje a chod domácnosti. Banka neví, že účet s měsíčním obratem sto tisíc korun, z něhož peníze každý měsíc dostává, patří jejímu manželovi. Jakákoliv žádost o zvýšení týdenního limitu na platební kartě bude zamítnuta.

Kde se householding používá?

- retail marketing: CRM (člen householdu)
- B2B marketing: CRM
- pošta: úspory ze zasílání pošty
- finančnictví: agregace rizik (úvěry, bankroty)
- SCM – řízení dodavatelských řetězců: konsolidace dodavatelů (úspory z rozsahu, zajištění lepších smluvních podmínek)
- péče o zdraví: fůze a akvizice (konsolidace dat o pacientech)
- sales management: (redukce konkurence o stejného zákazníka)

[Firstlogic]

Formou householdingu je též v telekomunikacích používaný pojem „Sphere of Influence“ (v překladu sféra vlivu). Pro marketingové účely analyzují telekomunikační

společnosti seznamy lidí, se kterými voláte. Sledují se příchozí i odchozí hovory. Vytvářejí se tak skupiny blízkých osob, pro které je možné přizpůsobit obchodní model.

Zajímavé možnosti využití HH se nabízejí také ve státní správě. Za příklad by mohl posloužit systém vyplácení sociálních dávek. Ty jsou vypláceny spíše plošně a často se tak stává, že dávku dostane někdo, kdo ji vlastně ani nepotřebuje, nebo již podobnou jinou dávku získal. Dochází tak k nevhodné redistribuci finančních prostředků a ve své podstatě k plýtvání. Householding by v tomto případě mohl sloužit jako pomocný nástroj pro odhalování případných nesrovnalostí. Využití householdingu tímto způsob by se však muselo opírat o příslušné zákony, a je tedy v současné době spíše hypotetické.

Další významnou oblastí, ve které může HH pomoci, je soudnictví a kriminalistika.

3 Typy a přístupy k householdingu

V této části práce jsou popsány charakteristické typy a přístupy k householdingu, které literatura popisuje, nebo o kterých se hovoří. U jednotlivých typů bude uvedena definice, praktický příklad využití, charakteristické vazby a jejich popis. Závěr kapitoly obsahuje shrnutí.

3.1 Typy

Dostupná literatura a zdroje[1] hovoří převážně o osobním (personal) a komerčním (corporate) householdingu. Praktické využití nabízí také kombinace těchto zmíněných typů. Té je věnována poslední část kapitoly.

3.1.1 Osobní householding

Osobní householding je možné definovat jako vztahy mezi konkrétními fyzickými osobami v kontextu (nejen) rodiny.

Na úvod této podkapitoly krátký příklad: Dva manželé mají každý svůj účet ve stejné bance. Manžel pracuje jako generální ředitel významné společnosti a jeho plat mnohonásobně převyšuje celorepublikový průměr. Je tak pro banku zcela jistě velmi důležitým klientem. Jeho manželka je žena v domácnosti. Sama vpodstatě žádný plat nemá a úspory „domácnosti“ jsou vedeny na účtu manžela. Jako jednotlivec může být tato žena pro banku méně významnou. Problém nastane, požádá-li žena o nějaký úvěr. Banka, pakliže neví, že tyto dvě osoby patří k sobě, může žádost manželky zamítnout s tím, že nejsou dostatečně kryté její zájmy. Tím však může velmi popudit manžela, který jistě nebude spokojen s tím, že jeho manželce nebylo vyhověno. Může to být důvod, aby oba změnili svou banku. Se znalostí finanční situace celé rodiny by banka úvěr manželce ochotně nabídla sama a hledala by cestu, jak obrat na účtu manželky zvýšit. V konečném důsledku by pak banka lépe prodala svůj produkt.

[1] Klofáčová; Madnick, Wang, Zhang

Tento příklad ukazuje, jak je pro banky důležité (a nejen pro ně) znát své klienty. Nedokáže-li rozpoznat vztah těchto lidí, resp. jejich příslušnost k určité skupině, může ztratit mnohem více, než jen nesplacený úvěr.

Mezi oblasti, ve kterých je možné využít efektu householdingu, tedy snížit náklady a zvýšit ziskovost, patří podle [Klofáčová]:

- **Profitabilita zákazníka** je pro firmy významným ukazatelem. Za předpokladu přijetí všeobecně uznávaného faktu, že 15-20% zákazníků generuje kolem 50% celkového zisku, se mohou firmy prostřednictvím využití HH zaměřit právě na ty zákazníky, kteří generují hodnotu. Firmy tak získávají realističtější pohled na své zákazníky, resp. přehled o svých zákaznících, a dokáží tak zvyšovat své zisky.

- **Úspora nákladů** na zasílání přímých nabídek poštou nebo přímé oslovení prostřednictvím direct-callu je další devizou householdingu. Oslovením pouze jednoho člena domácnosti lze odstranit náklady na několikanásobné oslovení. Také se zamezí tomu, aby různým členům domácnosti přicházely nabídky s různými cenami, což je jistě nežádoucí.

- **Cross-selling**[1] je další oblastí využití householdingu. O řadě finančních produktů rozhodují členové domácnosti společně. Může se jednat například o hypotéky, různé půjčky, životní pojištění atp. Vývoj produktů a jejich cílená nabídka pro cross-selling bez zohlednění householdingu ztrácí na efektivitě a může být i kontraproduktivní.

Možných oblastí využití osobního householdingu by se dalo jistě najít mnoho. Výše zmíněné však patří mezi ty nejběžnější.

3.1.2 Komerční householding

Komerční HH je dle [Madnick] možné definovat, *„určitou analogií k rodinné domácnosti, jako skupinu obchodních jednotek spojených nebo považovaných za jedno společenství.“* Může se také jednat o obchodní subjekty, které společně kooperují.

[1] Cross-selling = prodej souvisejícího zboží

Rychle se měnící obchodní prostřední má výrazný vliv na časté změny podnikových struktur a mezipodnikových vztahů. Do změn, znesnadňujících pochopení vzájemných vztahů mezi komerčními subjekty, patří: regulace, deregulace, akvizice, konsolidace, fůze, strategické partnerství, nové regionální centrály, nová odvětví, bankroty, franchizy atd.

Vztahy charakteristické pro komerční householding jsou například:

- vztahy uvnitř společenství, tj. oddělení, divize, odvětví atd.,
- vztahy představované organizačními schématy,
- vztahy se zákazníky, partnery, a spolupracovníky,
- vztahy s dodavateli,
- vztahy se zprostředkující třetí stranou, tj. obchodníky, distributory, dealery, atd.,
- vztahy s nadřízenými orgány

Zdroj [Madnick] uvádí typické využití komerčního householdingu. Patří mezi ně oblast finanční, právní a v neposlední řadě také obchodní.

Finanční nebezpečí

- **Úvěrová rizika** – Rozhodnutí o úvěru vstupuje stále čím dál více do různých finančních operací. Tím i komerční HH roste v této oblasti na významu. Oceňování bonity[1] jednotlivých klientů vyžaduje významné množství úsilí. Např. pro firmu plánující rozšířit úvěrovou linku společnosti HP Česká republika může být užitečná informace, že společnost HP má úvěrovou bonitu pouze stupně „AA“, přestože společnost HP Česká republika má tuto bonitu vyšší („AAA“). Jinými slovy řečeno, při oceňování úvěrového rizika může být nápomocná znalost toho, že některé entity mohou být považovány za tytéž.
- **Rizika bankrotu** - Riziko bankrotu je úzce spojeno s úvěrovým rizikem. Při rozhodování, zdali půjčit či nepůjčit finanční prostředky určité firmě, je pro

[1] Bonita = schopnost dlužníka dostát svým finančním závazkům

banku vhodné vědět, kdo by mohl ručit v případě krachu firmy, případně kdo může být za tento krach zodpovědný. Například krachuje-li dceřiná společnost, jakou odpovědnost (jestli vůbec nějakou) její "matka" má? Tato informace je využitelná v případě vymáhání případných pohledávek.

- **Mezinárodní rizika** - S rostoucím globálním dosahem firem rostou i rizika spojená s rozdíly v obchodních zvyklostech v různých částech světa. Např.: Zakládáme společnost v Brazílii, která je divizí nadnárodní americké společnosti. Nebo zakládáme firmu v USA, jejíž rodič sídlí v Japonsku (například Toyota má výrobní linku přímo v USA). Můžeme tyto firmy považovat za Americké, nebo jsou spíše pouze dceřinými společnostmi?

Konsolidace účetnictví

Komerční HH může být též nápomocný při vyhodnocování účetních informací větších organizací třetí stranou, která má vůči organizaci nějaké obchodní či marketingové záměry. Větší organizace mohou mít po celém světě více než stovku dceřinných společností a může být tak obtížné vyhodnotit účetní informace celé společnosti.

Právní oblast

V právní oblasti existuje mnoho aktivit spojených s komerčním HH. Patří mezi ně vydávání softwarových a patentových licencí. Problém mohou mít dodavatelé velkých celopodnikových řešení při účtování jednotlivých licencí, dojde-li k fůzi nebo akvizici mezi jejich zákazníky. Metody a nástroje HH mohou být nápomocné při jejich identifikaci.

Řízení obchodních operací

Uplatnění komerčního householdingu lze nalézt také v oblastech, jako je řízení vztahů se zákazníky (CRM), řízení dodavatelských řetězců (SCM), v prodeji, marketingu a také v business intelligence.

- CRM - Když je zákazníkem nadnárodní společnost, mívá prodejce ve svém informačním systému uloženo velké množství unikátních kontaktních údajů. Který kontakt je relevantní pro konkrétní událost?
- prodej a marketing - V této oblasti je stále rostoucí potřeba systémů pro identifikaci zákazníků, které umožňují vytvoření celistvého pohledu na B2B zákazníky a tím vyhledání těch, kteří mají vysoký obchodní potenciál.
- SCM - Pro pokles nákladů je pro firmy důležité jasně identifikovat vztahy mezi jednotlivými dodavateli materiálu. Některé výrobní pobočky mohou mít rozdílné dohody s týmž dodavatelem na stejný materiál.

3.1.3 Kombinace osobního a komerčního householdingu

Typem householdingu může být i kombinace obou výše zmíněných typů. Jde o přístup, kdy součástí householdu je jak osoba fyzická, tak i konkrétní obchodní subjekty nebo obchodní jednotky. Znalost členů takto vytvořené skupiny může být efektivně využita.

Mezi vztahy charakteristické pro tento typ householdu patří:

- vztahy mezi firmami a jejich zaměstnanci
- vztahy mezi firmami a vlastníky či majiteli
- vztahy mezi zákazníky a firmami
- vztahy společností a statutárních nebo kontrolních orgánů, členů dozorčích rad atp.

Důvodem sledování těchto vztahů může být například u vazby firma-vlastníci snaha o odhalení podvodného chování. V nemalém množství se stávají případy, kdy konkrétní osoba založí více společností a mezi nimi „přelévá“ finanční prostředky, závazky atp. Další neblahou praktikou některých podnikatelů je praní špinavých peněz. Householding výrazně přispívá k identifikaci tohoto chování a rozklíčování mnohdy velmi složitých vztahů. Banky tak získávají informace, které slouží jako varovné signály těchto podvodných aktivit.

Kombinace osobního a komerčního HH je také využívána při identifikaci vazeb ukazujících na střet zájmů. Střetem zájmů se rozumí situace, ve které je člověk odpovědný dvěma nebo více různým organizacím či autoritám, jejichž požadavky na jeho chování se více či méně zásadním způsobem rozcházejí. V některých případech je střet zájmů nezákonný.

Využití lze nalézt pro marketingové účely. Například banka, mající svého klienta, pana Nováka, ředitelem společnosti Xyz s.r.o., může usilovat o to, aby měl pan Novák svůj osobní účet ve stejné bance, jako je účet společnosti. Banka je pak schopna nabízet panu Novákovi individuální služby, které zvýší jeho komfort i spokojenost. Efekt pro banku je zřejmý.

3.2 Přístupy

Z výše uvedeného vyplývá, že využití výsledků identifikace HH lze spatřit ve dvou tradičních sférách. Jedná se o:

1) Risk Management
2) Marketing

U rizikově orientovaného HH je v některých případech možné připustit, aby konkrétní klient příslušel do více konkrétních typů householdů. Je to dáno tím, že se snažíme zachytit a vyhodnotit co nejvíce rizikových faktorů, tudíž i zapojení do různých householdů (zájmových skupin, atp.). U marketingově orientovaného typu bude klient příslušet nejspíše právě do jednoho householdu. Právě při marketingu chceme konkrétní household co nejefektivněji a nejsnadněji oslovit s pokud možno optimální nabídkou – ta by měla být právě jedna.

Uvedené oblasti se liší ve způsobu využití informace, což vede i k odlišnostem v definici a způsobu identifikace HH. Vpodstatě lze k HH přistupovat ze dvou odlišných směrů. Jednak v širším slova smyslu, tzn. maximalisticky, nebo úzce, tzn. minimalisticky.

3.2.1 Maximalistický přístup

U risk managementu se očekává využití HH pro účely odhalení co největšího počtu možných vazeb a asociací, které mohou představovat riziko při schvalování úvěru. Jde o to naleznout co nejvíce možností a variant. Jedná se tedy o široce chápaný přístup k HH, který je „maximalistický“. Současně je také tento přístup *„liberální“*, jelikož jsou při něm uplatňována volnější pravidla pro identifikaci podezření na existenci potenciálních (přímých i nepřímých) vazeb.

3.2.2 Minimalistický přístup

Pro marketingově orientovaný HH je charakteristický spíše konzervativní přístup. Ten je „minimalistický“. To je zapříčiněno dvěma důvody. Chybné oslovení konkrétní domácnosti může zcela negovat účel kampaně a naopak vyvolat negativní odezvu. Druhým důvodem je zachování co nejstabilnější definice HH pro analytické účely. Důsledkem je, že výsledná skupina by měla být identifikována spíše na základě přímých a jednoznačně dohledatelných vazeb – např. by měla mít společnou kontaktní adresu, na kterou je marketingová kampaň směřována.

Minimalistický přístup se, dle [Adastra] využívá zejména v marketingu, a to pro dva základní účely:

a) Jednak jako podklad pro přípravu a realizaci kampaní cílených právě na domácnosti. V tomto případě poskytuje výstup této úlohy seznam identifikovaných domácností, na které je kampaň zaměřena, tj. jako metodu oslovení nejpřesnější skupiny potenciálních zákazníků. Tento způsob využití je mj. charakterizován tím, že je zapotřebí zamezit vícenásobnému oslovení členů jedné domácnosti, případně dokonce s nabídkami za různých podmínek, a současně zohlednění potenciálu domácnosti jako celku i při různém individuálním hodnocení jejich příslušníků. Důsledkem je konzervativní přístup s cílem co nejpřesněji identifikovat HH a eliminovat možné chyby.
b) Druhou oblastí je využití v oblasti marketingových analýz. To vyžaduje především stabilní a relativně obecně použitelnou definici HH, která umožní získat „co nejpravděpodobnější“ informace o chování zákaznického portfolia

Tento přístup umožňuje pracovat relativně spolehlivou informací. Cenou za konzervativnost je fakt, že méně zjevná propojení klientů nebudou zřejmě odhalena.

3.3 Závěrečné shrnutí kapitoly

V této části práce byly identifikovány a blíže popsány tři typy householdingu, které se v praxi využívají a jsou též popsány dostupnou literaturou. Jedná se o osobní householding, tedy vztahy mezi konkrétními fyzickými osobami v kontextu (nejen) rodiny, dále householding komerční, jakožto skupinu obchodních jednotek spojených nebo považovaných za jedno společenství nebo také subjekty, které vzájemně kooperují a v neposlední řadě kombinaci obou těchto typů.

Dostupné zdroje se nejvíce věnují komerčnímu householdingu, který je patrně v popředí zájmu. Zajímavou oblastí je kombinace osobního a komerčního typu, zejména z důvodů pestrosti vztahů a jejich obtížnější identifikaci. Mou snahou bylo se právě o této oblasti co nejvíce dozvědět, ale bohužel není toto téma, alespoň ve vyhledaných materiálech, příliš zdokumentováno.

K householdingu existují dva základní přístupy. Jedná se o šířeji chápaný household (v práci nazván maximalistický přístup), který se snaží pojmout co nejvíce variet vazeb. Toho se využívá zejména při identifikování a kvantifikaci možných rizik. Pro tento přístup je charakteristická kvantita (vztahů) a nižší kvalita (jasná identifikace householdu). Druhým, jemnějším přístupem, je přístup minimalistický. U toho přístupu je naopak velice důležitá kvalita, tedy přesné určení householdu. To umožňuje dle [Kyjonka]: *„Pracovat s relativně spolehlivými informacemi. Úskalím tohoto přístupu je fakt, že méně zjevná propojení klientů nemusí být odhalena.“*

4 Budování householdingu

V postupu budování householdingu lze, u dále zmíněných společností, nalézt společné fáze, které jsou pro householding charakteristické. Společnosti při householdingu hovoří o oblastech (Adastra), krocích (FirstLogic), modulech (Hart-Hanks), ale v podstatě lze tyto „postupné kroky“ zastřešit třemi fázemi.

	Adastra [4.1.1]	**FirstLogic** [4.1.2]	**Hart-Hanks** [4.1.3]
1. fáze – úvodní	definování business cílů	stanovení cílů projektu	parsing jmen a adres
	definice vlastního HH	definice použitých termínů a dohoda o funkčních přínosech	doplnění poštovních údajů
		definování obchodních pravidel k dosažení podnikových cílů	geocoding
2. fáze - budování householdů	konfigurace pravidel a zařazení do HH	vytvoření matice pravidel	utváření dílčích přiřazovacích pravidel
		ověření matice pravidel	
		vytvoření aplikačních parametrů / parametrů aplikace	kombinování přiřazovacích pravidel
	ladění pravidel	kontrola konzistence pravidel	
		ověření všech přiřazovacích sad	aplikování jednoho nebo více dílčích přiřazovacích pravidel
		ověření konsolidačních kritérií	
3. fáze – dokončovací	implementace	spuštění Householdingu	load dat do databáze (DWH, DM)
	realizace	dokončení nahrání dat do systému	

Tabulka č.3: Fáze budování householdingu

V první fázi, úvodní, dochází k ujasnění toho, co se od projektu očekává, definují se cíle, kterých se chce dosáhnout a sjednocuje se vzájemná terminologie. Dále se stanovují termíny a další dílčí podmínky. Velice důležité je definování business účelů, pro které se householding vytváří (více o business účelech viz kapitola [2.2] - Všeobecné využití householdingu). Odlišný přístup lze nalézt u společnosti Hart-Hanks [4.1.2], která se během této fáze zabývá řešením datové kvality (parsing jmen a adres,

doplnění poštovních údajů, geocoding). Naopak Adastra i FirstLogic předpokládají práci s již unifikovanými záznamy[1], tedy čistými a kvalitními daty [4.1.1 a 4.1.3]. V této kapitole bude vysvětleno, proč je dobré projekt řešení datové kvality od samotného householdingu oddělit. Problematika datové kvality je však i přesto pro householding velice významná. Bude tedy blíže popsána v podkapitole [4.3.2] – Nutné podmínky budování householdingu / Kvalitní a čistá data.

Během druhé fáze, budování householdů, se řeší utváření pravidel, která jsou určující pro zařazení subjektů do householdu. Tato fáze je dle mého názoru stěžejní, protože právě zde dochází k tvorbě vnitřní logiky věci. Adastra i FirstLogic vkládají jednotlivá pravidla do matic (viz tabulka č. 6), se kterými následně pracují. Forma těchto informací zapsaných do matice je vhodná jednak pro určitou názornost, ale také pro pozdější automatizované zpracovávání. Při utváření pravidel může dojít ke zjištění, že některé potřebné údaje chybějí. Tím vznikne požadavek na jejich dodatečné doplnění, umožňuje-li to daná situace. V této fázi se dále vymýšlí (nebo aplikuje již vymyšlené), jakým způsobem se budou řešit případné problémy či anomálie. Podle postupu Adastry se zde například řeší rozdílnost skupin v malém či velkém městě, různá četnost jmen, stárnutí dat atp. Význam jednotlivých pravidel může být též různý, a proto dochází k jeho kvantifikaci. Při kontrole je zapotřebí spolupráce se zadavatelem, který ověří správnost jednotlivých pravidel.

Třetí fází, dokončovací, se řešení HH nasazuje. Dochází k ověření funkčnosti nad reálnými daty. Součástí této fáze by mělo být školení budoucích uživatelů systému vedoucí k osvojení obsluhy. Závěrem by měla být shoda zúčastněných stran na naplnění vytyčených cílů, které byly na začátku projektu stanoveny.

Pro každý projekt, a ani householding není výjimkou, musí být splněny určité předpoklady. Charakteristiku a popis těchto předpokladů lze nalézt v podkapitole [4.3] Nezřídka projekty provázejí možné neduhy a problémy. Identifikované problémy budou v této kapitole blíže popsány [4.2] a to včetně návrhu jejich řešení.

---

[1] Unifikovaným záznamem se rozumí záznam, který je odvozen z instancí procesem unifikace. [pojem převzat s metodiky identifikace implementované v Purity - Adastra]

4.1 Postup budování významnými společnostmi

Postupem budování householdingu se zabývají jen některé vyhledané zdroje. Tato část práce se zabývá těmi nejvýznamnějšími.

4.1.1 Adastra

Společnost Adastra rozděluje projekt HH do šesti oblastí, které na sebe navazují.

1) definování business cílů
2) definice vlastního HH
3) konfigurace pravidel a zařazení do HH
4) ladění pravidel
5) implementace
6) realizace

Definování business cílů

Pod definováním business cílů si můžeme zjednodušeně představit vyjádření důvodů, pro který se celý proces identifikace householdu provádí. Tato část je pro projekt HH klíčovou, protože bez jasného stanovení cílů postrádá budování HH smysl.

Mezi tradiční sféry uplatnění HH patří risk managament a marketing. Proto nejběžnější business cíle budou vycházet z těchto oblastí. Příklady business cílů jsou uvedeny v kapitole [2.2 - Všeobecné využití householdingu].

Jasné stanovení cíle, kterého se chce v projektu householdingu dosáhnout, umožní předejít případnému zklamání.

Definice vlastního HH

Cílem této části je definování obsahu householdingu a také určení, jakými oblastmi se projekt zabývat nebude.

Pro úspěch v householdingu je důležité, aby si jednotlivé zúčastněné strany rozuměly. Proto dochází na počátku projektu k vlastní definici householdu a jeho obsahu. Obsahem této definice je i určení toho, kdo je zákazníkem, co je zakázkou atd. Pro samotný projekt HH je nezbytné, mít jednotlivé pojmy vyjasněné.

Přesná definice Householdu se pro jeho jednotlivé typy může v detailech lišit. Společnost Adastra definuje[1] HH jako netriviálně definovanou skupinu fyzických osob, mezi kterými existují vztahy různého typu, jež indikují příslušnost k domácnosti v kontextu bydliště a rodiny a předpokládají ekonomickou provázanost jejích členů.

Z praktických důvodů a ze zkušeností ověřených z praxe rozděluje Adastra HH podle principu dvouúrovňové identifikace. Ta spočívá ve vytvoření tzv. kandidátské[2] a klientské skupiny.

Kandidátská skupina je podle [Adastra] *„skupina všech unifikovaných klientů, kteří se shodují s jinými na základě přímé i nepřímé vazby (shody přes existující mezičlánek) alespoň v jednom z možných párovacích atributů. Tato skupina je relativně široká, protože zahrnuje všechny unifikované klienty, kteří mají nebo mohou mít něco společného.“*

Tento pohled umožňuje vysledování i nepřímých vazeb, které mohou indikovat i na první pohled neviditelné propojení mezi klienty. U kandidátské skupiny je důležité uvědomit si, že jednotlivé vazby by měly být chápány jako indikativní a používány spíše jako podklad pro další šetření. Nejsou tedy vhodné například pro automatizované zpracování citlivých operací (například automatické neschválení úvěru).

Klientská skupina[3] je tvořena kolem každého člena kandidátské skupiny na základě přímých a přísnějších vazeb. Jednotlivé vazby jsou oceněny stupněm příslušnosti k danému HH vyjádřeném na škále od 1 do 100.

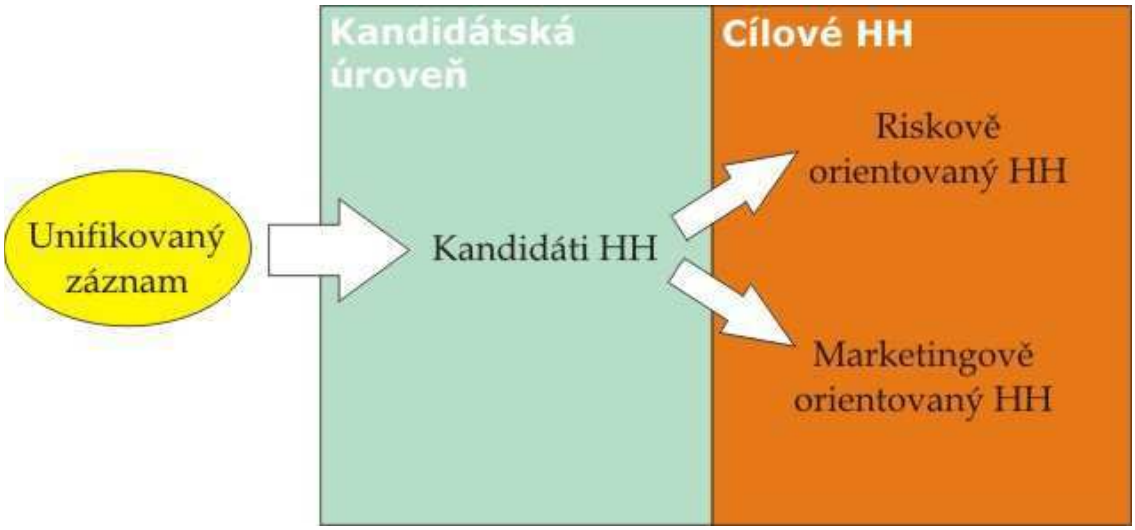
---

[1] Tato definice byla převzata z jednoho konkrétního příkladu. Dle Vladimíra Kyjonky se definice mění v závislosti na oblasti, kterou HH pokrývá – více viz 2.1.1

[2] Pojmy kandidátská a klientská skupina byly převzaty z metodiky Adastry pro unifikaci a konsolidaci dat, vyvinuté prvotně pro identifikaci klientů

[3] Možné chápat též jako cílovou skupinu

Dochází tak k identifikaci HH s co největší mírou jistoty (v rámci předem definované tolerance). Tento pohled umožňuje pracovat s informací jako s relativně spolehlivou. Částečným nedostatkem klientské skupiny je fakt, že méně patrná propojení klientů nemusí být odhalena.



[Adastra]

Obrázek č.1: Schéma principu dvouúrovňové identifikace

Konfigurace pravidel a zařazení do HH

V této fázi dochází k určení toho, podle jakého klíče se budou jednotliví klienti slučovat a jaká pravidla budou pro household určující. Takových pravidel může být mnoho. Limitujícím faktorem jsou dostupná data (interní i externí zdroje). V některých případech může nastat situace, kdy potřebný údaj není evidován. V této fázi tak může vzniknout požadavek na jeho dodatečné doplnění.

Mezi nejtypičtější sledované ukazatele, které vstupují do utváření pravidel, patří:

- příjmení
- trvalá adresa
- kontaktní adresa
- telefonní číslo
- email

- horizontální a vertikální vztah[1]

Při utváření pravidel Adastra zohledňuje také četnost příjmení či velikost města, ze kterého daná osoba pochází. Například Nováků je v ČR evidováno přibližně 35 tis. a lehce tak může dojít k záměně konkrétních osob. Na druhou stranu Běhunků je evidováno jen kolem 30 a je tedy veliká pravděpodobnost, že jsou-li v malém městě Běhunkové dva, budou s největší pravděpodobností příbuzní.

Pro názornost pravidel a pochopení jejich významu se vytváří matice pravidel.

	MP	VP	MTA	VTA	MKA	VKA	TEL	HOR	VER
I.	X		X						
II.	X			X					
III.		X	X						
IV.		X		X			X		
V.	X				X				
VI.	X					X			
VII.		X			X				
VIII.		X				X	X		
. . .									

Tabulka č.4: Příklad matice pravidel

Legenda:

- **MP** méně časté příjmení (např. Kmínek)
- **VP** více časté příjmení (např. Novák)
- **MTA** méně četná trvalá adresa – ukazuje na malou obec
- **VTA** více četná trvalá adresa – ukazuje na velkou obec
- **MKA** méně četná kontaktní adresa – ukazuje na malou obec
- **VKA** více četná kontaktní adresa – ukazuje na velkou obec
- **TEL** telefonní číslo pevné linky
- **VER** vertikální vztah (např. otec-syn)
- **HOR** horizontální vztah (např. manžel-manželka)

[1] vztah je horizontální, právě tehdy, když jsou dané subjekty na stejné úrovni – například manžel x manželka; vertikální je vztah například mezi matkou a dcerou

Z této tabulky je patrné, že jsou jednotlivá pravidla tvořena prostřednictvím příjmení, včetně rozlišení četností, dále trvalé i kontaktní adresy s rozlišením malé nebo velké obce, telefonního čísla, horizontálního i vertikálního vztahu.

Jednotlivých pravidel může být v zásadě mnoho. Například na základě interpretace I. pravidla z matice jsou do kandidátské skupiny zařazeni klienti, kteří mají shodné málo četné příjmení a zároveň mají shodnou trvalou adresu na malém městě. Podle II. pravidla jsou do kandidátské skupiny zařazeni klienti, kteří mají shodné málo četné příjmení a zároveň mají shodnou trvalou adresu ve velkém městě. Podle III. pravidla jsou do kandidátské skupiny zařazeni klienti, kteří mají shodné hodně četné příjmení a zároveň mají shodnou trvalou adresu na malém městě. Podle IV. pravidla jsou do kandidátské skupiny zařazeni klienti, kteří mají shodné hodně četné příjmení a zároveň mají shodnou trvalou adresu ve velkém městě. Záznamy musí mít navíc shodné i telefonní číslo. Podle jiného pravidla mohou být klienti slučování na základě existence deklarovaného horizontálního vztahu (manžel-manželka) atp.

Ladění pravidel

Nastavení a vyladění pravidel je závislé na dohodě se zadavatelem. Při ladění dochází ke kontrole a případné opravě správné funkce jednotlivých pravidel.

Adastra mj. dále zohledňuje **stárnutí informací**[1]**,** což vpodstatě znamená další pravidlo. Dochází tak ke kvantifikaci stárnutí – tzv. skóre stáří. Významné je sledovat stáří záznamů a také horizontální vztahy, tedy manžel-manželka. Naopak nemá smysl hodnotit stáří vertikálního vztahu, protože například vztah otec-syn je nezrušitelný.

Kvantifikované stáří informací (skóre stáří) může vypadat například takto:

[1] Více viz kapitola 3.2.6 Stárnutí dat i householdu

Skóre stáří informace	Intervaly stáří v letech					
	0-1	1-2	2-4	4-8	8-16	16 a více
Stáří horizontálního vztahu	0	20	40	60	80	100
Stáří unifikovaných záznamů	0	20	40	60	80	100

Tabulka č.5: Příklad tabulky skórování stáří informací

Hodnoty skóre stáří jsou definovány v rozsahu 0 až 100. Čím starší hodnota, tím vyšší skóre.

I jednotlivá pravidla mají různou vypovídající schopnost. Proto se hodnotí a nastavují se jim váhy – tzv. váhy přislučovacích pravidel. Víme-li například, že určití klienti jsou ve vztahu otec-syn (vertikální vztah), je jisté, že tyto osoby do konkrétního HH přísluší. Naopak vztah, na který aplikujeme pravidlo vícečetných příjmení společně s velkým městem, už tak jistý být nemusí. Pravidla jsou aplikována v pořadí podle skóre od nejnižšího po skóre nejvyšší, dokud se jedno z nich neuplatní.

Aplikované pravidlo	Váha pravidla
VER	100
HOR	100
MP+MTA	75
MP+MKA	70
VP+MTA+TEL	70
VP+VTA+TEL	65
TEL	55
MP+VTA	50
MP+VKA	50
VP+MKA	25
VP+MKA	25

Tabulka č.6: Příklad vah přislučovacích pravidel

Tabulka č. 3 ukazuje příklad ocenění „kvality“ pravidel. Jednotlivé váhy jsou nastaveny v rozmezí 0 až 100. 100 znamená jistotu, že je záznam přiřazený do householdu tímto pravidlem správně zařazen. 0 znamená minimální důvěryhodnost správnosti zařazení.

Započtením stáří informace a kvality použitého pravidla lze vyjádřit celkovou **míru příslušnosti klienta k householdu**.

Implementace

Implementace je závěrečná, avšak důležitá část nasazení řešení HH, při které dochází k postupnému zavádění jednotlivých komponent do provozu. Součástí implementace je také určité osvojení obsluhy tohoto systému jednotlivými uživateli. Implementace je velice důležitou částí procesu zavádění HH, protože ačkoli je daný systém sebelepší, chyby v implementaci mohou mít takřka katastrofální účinky.

Čas implementace je závislý na mnoha faktorech. Zájmem každé firmy je mít tuto část pokud možno co nejdříve za sebou. Je však velice žádoucí dát si pozor na to, aby rychle odvedená práce nebyla na úkor kvality. Doba implementace je individuální projekt od projektu.

Klíčové pro všechny úlohy řízení datové kvality (kam HH patří) je nastavení business procesů v podniku.[1] Úlohy DQ vpodstatě mění informaci, se kterou se pracuje (obsah, relevanci, spolehlivost …) Jednotlivé výstupy mají smysl pouze tehdy, když se dále využívají. Při HH vzniká nová business entita – household. Ta by měla mít svého vlastníka a mělo by být dáno, kdo a jakým způsobem s ní může zacházet.

Realizace

Po nasazení řešení householdingu začíná jeho provoz. Je možné, že se v této fázi zjistí, že některé části neprodukují očekávané výstupy. Vpodstatě až uživatelé ověří funkčnost systému. Mohou tak vznikat připomínky a požadavky na konkrétní změny či doplnění. Dodavatel systému by měl být schopen na tyto požadavky reagovat.

Householdy se v čase vyvíjí. HH definované dnes nemusí odpovídat požadavkům budoucnosti. Systém by měl být nastaven tak, aby bylo možné jednotlivé změny zapracovávat a systém tak dále rozvíjet.

V této fázi by též mělo dojít k zhodnocení projektu. Splnila se očekávání do něj vložená? Byly naplněny business cíle stanovené v úvodu projektu? Jaký je finanční přínos householdingu? Změřit skutečný efekt householdingu v penězích není

[1] Problematika business procesů je velice rozsáhlá a není součástí této práce. V textu je uvedena pouze jako důležitá zmínka a to z důvodů častého opominutí tohoto aspektu v úvodní fázi.

jednoduché. Je to dáno zejména tím, že se tyto efekty neprojeví hned po nasazení řešení. Efekty by měly být poměřovány na základě metrik, stanovených v úvodní fázi.

[Adastra]

4.1.2 Harte-Hanks Data Technologies

Harte-Hanks je společnost s celosvětovou působností poskytující přímá a zacílená marketingová řešení. V únoru 2004 zakoupila produkt Trillium a vytvořila samostatnou divizi Trillium Software. Trillium Software System je produkt pro řízení kvality zákaznických dat, který kombinuje funkce čištění dat a párování zákaznických dat dle domácností (householding).

[Sacolick] spatřuje komplexitu (náročnost) householdingu zejména v: *„používání různých přiřazovacích pravidel, u nichž se pracuje s neurčitostí, dále v tom, že data mohou obsahovat různé formáty jmen a adres a v neposlední řadě v časové změně HH pod vlivem externích faktorů.“*

Harte-Hanks [Sacolick] hovoří o householdingu jako o souboru jednotlivých dílčích částí (modulů), které na sebe navazují. Definice HH dle Harte-Hanks je uvedena v první kapitole této práce. Jednotlivé moduly jsou:

- parsing jmen a adres
- doplnění poštovních údajů
- geocoding
- utváření dílčích přiřazovacích pravidel
- kombinování přiřazovacích pravidel
- aplikování jednoho nebo více dílčích přiřazovacích pravidel
- load dat do databáze (DWH, DM)

Tyto moduly se skládají z mnoha dalších kroků a i mezi nimi může docházet k různým přeformátováním dat, tříděním dat atp.

Pro názornost [Sacolick] uvádí, že se householding obyčejně skládá ze 4-5 přiřazovacích pravidel (kritérií), 3 úrovní výstupů (householdy, jednotlivce, přiřazené adresy), 30-40 postupných kroků (od parsingu až po nahrání dat do databáze) a u větších finančních společností z desítek až stovek miliónů záznamů, které se v každém kroku zpracovávají.

Parsing jmen a adres

Jména a adresy bývají často do systému vkládány v nenaformátované podobě. To způsobuje problémy při jejich automatizovaném zpracování. Parsing zajišťuje rozdělení nestrukturovaných vstupních údajů do požadovaných skupin. Parsing je jedním z nejobtížnějších úkolů při HH. Vyžaduje využití pomocných tabulek (číselníků), pomocí kterých se jednotlivé komponenty jmen a adres identifikují. Například Trillium obsahuje stovky tabulek, ve kterých jsou vedena běžná jména i příjmení, obchodní názvy, názvy měst, běžné tituly atd.

Parsing nenaformátovaných jmen a adres je potřebný pro:

- zvýšení přesnosti a věrohodnosti jmen a adres pro poštovní účely
- vyloučení vícečetných pojmenování stejného účtu
- zvýšení přesnosti přiřazovaní

Doplnění poštovních údajů

Tento modul využívá externích číselníků měst, ulic, čísel popisných a PSČ k doplnění chybějících údajů, případně jejich opravě nebo identifikaci těch, se kterými by mohly být potenciální problémy. Více o externích číselnících viz kap. [4.3.1]

Příklad: U adresného záznamu pana Nováka chybí údaj PSČ. Ten je možné dohledat na základě města, které u kontaktní adresy pana Nováka uvedeno je. Správnost uvedeného města lze zkontrolovat tím, že ověřím přítomnost ulice v daném městě.

Geocoding[1]

Hart-Hanks využívá modul nazývaný Geocoding. Jedná se o snahu přesné lokalizace jednotlivých adresných bodů v kontextu zeměpisné šířky a délky. Tyto údaje jsou dále využívány pro účely získání demografických informací, které znamenají pro householding přidanou hodnotu. Je tedy možné například vyhledat oblasti s vysokou koncentrací obyvatel atp.

Utváření dílčích přiřazovacích pravidel

Zde dochází k utváření dílčích pravidel, podle kterých jsou jednotlivé záznamy přiřazovány (slučovány). Výchozími údaji jsou:

- jmenná data
- adresní data
- jednoznačné identifikátory (RČ, čísla pojištění, občanských průkazů atp.)
- vazební data, která jsou již v systému evidovány

Po rozparsování záznamů se například vytvoří pravidlo, které spojí oficiální název jména se jménem méně oficiálním či přezdívkou (Honza = Jan, Tomáš = Tom atp.). Při spojování záznamů se lépe pracuje s číselnými daty, protože jsou, na rozdíl od textových, více konzistentní a přesná.

Všechna pravidla jsou aplikována na dílčí případy, to tedy znamená, že sada pravidel se používá znovu a znovu na každý účet. Proto se doporučuje některá pravidla sloučit a snížit tím počet datových operací.

Bližší údaje k tomuto, jistě důležitému, modulu zdroj neuvádí.

[1] Dle slovníku na www.camcad.cz znamená geocoding proces přiřazení zeměpisné polohy bodu podle adresy, ulice, města a PSČ

Kombinování přiřazovacích pravidel

Jednotlivá dílčí pravidla jsou slučována a kombinována pro získání dalších znalostí. Těmi může být: Víme-li, že účet A je podle nějakého pravidla přirazen k účtu B a účet B podle jiného pravidla k účtu C, můžeme uvažovat o tom, že účet A patří též k účtu C.

Tedy například jsou-li účty 1-5 sloučeny na základě shodného čísla pojistné smlouvy, účty 6-10 sloučeny podle stejného jména a adresy a účty 4-7 podle stejného čísla účtu, potom 1-10 náleží stejnému householdu.

Využitelnost tohoto modulu je omezena tím, že během něj dochází k vytváření velikých skupin. To má za následek vysoké požadavky na systémové prostředky (vysoké zatížení databáze atp.). Proto je možné tento modul využít jen v některých speciálních případech.

Aplikování jednoho nebo více dílčích přiřazovacích pravidel

Zde jsou jednotlivá pravidla použita a vznikají tak konkrétní vazby a householdy. Obyčejně se používají jednotlivá pravidla na dílčí případy.

Když jednotlivec přísluší, na základě určitého pravidla, pouze do jednoho HH, je možné tohoto jednotlivce k HH skutečně přiřadit.

Load dat do databáze (DWH, DM)

Po vytvoření HH jsou výstupní data nahrána do databáze. Projekt householdingu vyžaduje svůj určitý čas a bohužel i během něj data stárnou [4.2.8]. Může se tak stát, že data nahrávaná do databáze nemusejí již odrážet skutečný stav. Nejběžněji se jedná o změnu jména během svatby, změnu adresy i čísel účtů.

[Sacolick]

4.1.3 Business Objects / Firstlogic

Firstlogic je celosvětovým poskytovatelem řešení a služeb datové kvality pro firmy. Dne 8.2.2006 došlo u Firstlogic k akvizici společností Business Objects, jež patří mezi celosvětově nejvýznamnější hráče na poli BI.

Společnost Firstlogic vytvořila a zveřejnila dokument[1] s názvem Householding: An 11 – Step Approach, který je asi neobsáhlejším veřejně dostupným materiálem zabývajícím se touto tématikou.

[Firtstlog] hovoří o 11 krocích, které jsou součástí postupu tvoření HH. Jedná se o:

1) stanovení cílů projektu
2) definice použitých termínů a dohoda o funkčních přínosech
3) definování obchodních pravidel k dosažení podnikových cílů
4) vytvoření matice pravidel
5) ověření matice pravidel
6) vytvoření aplikačních parametrů / parametrů aplikace
7) kontrola konzistence pravidel
8) ověření všech přiřazovacích sad
9) ověření konsolidačních kritérií
10) spuštění Householdingu
11) dokončení nahrání dat do systému

V souvislosti s těmito 11 kroky hovoří Firstlogic o tzv. „GFS vendors“ (global financial software vendor), což je zástupné jméno pro konkrétního zákazníka a o „SME“ (subject matter expert), expertnímu týmu zajištujícímu samotný projekt HH. Tento tým expertů napomáhá při budování jednotlivých vazeb, budování samotné architektury řešení atd. Jednotlivé kroky napomáhají též dodavatelům podnikových systémů ke zlepšení svých produktů a služeb.

[1] http://www.csdk.dk/B6473516-7FA1-4104-A6D4-A189A65D2DFF

1. krok - stanovení cílů projektu

Toto je první rozhodující krok v jakémkoliv projektu. Cílem firem je získat jednotný pohled na své současné zákazníky. Na rozdíl od procesu HH se dodavatelé podnikových IS nezajímají o konsolidaci dat „nahoru“ či „dolů“ v organizačním schématu zákazníků. Spíše se zaměřují pouze na konsolidaci zákaznických účtů napříč každým odvětvím struktury firmy.

Firmy mají data a SME znalosti týkající se "rodičovských" vztahů jeho zákazníků. Jednotlivá nákupní rozhodnutí jsou stimulována dílčími obchodními odděleními. Firmy však mají zájem na tom, aby byla tato zákaznická nákupní rozhodnutí poháněna „z vrchu“ celé společnosti, tedy centrálně řízena.

V tomto kroku jde o získání jednotného pohledu na vzory chování jednotlivých zákazníků. Pro firmy je dále důležité rozpoznat, že různé názvy mohou patřit shodné společnosti. Např. Lotus i Lotus Development Corp. je název té samé společnosti.

2. krok – definice použitých termínů a dohoda o funkčních přínosech

Jednotné pojmenování, resp. standardizace definic meta dat, je velmi důležité pro jednotné a všem srozumitelné dotazování a reportování informací, a to napříč celou firmou. To je také užitečné pro nalezení odchylek v datech.

Vzájemné porozumění ohledně jednotlivých pojmů je důležité. V tomto kroku se například definuje, jak firma vnímá svého zákazníka, zakázku, míru spolehlivosti a další. Každý může pod těmito pojmy vidět něco jiného. Pro samotný projekt HH je nezbytné mít jednotlivé pojmy vyjasněné.

3. krok – definování obchodních pravidel k dosažení podnikových cílů

V případě firem to znamená získat jasný přehled o pravidlech, která podporují jednotlivé business cíle. Pravidla vznikají při vzájemné spolupráci zástupců dodavatele řešení a zákazníka. Tato pravidla obsahují:

- pravidla identifikování duplicit

- pravidla, podle kterých se ověřuje stupeň důvěryhodnosti (míra spolehlivosti) duplikátních záznamů
- příklady přípustných duplicit
- příklady různých stupňů míry spolehlivosti
- pravidla, podle kterých se řídí logika konsolidace záznamů
- příklady přípustných zkonsolidovaných záznamů
- vnitřní logika slučování speciálních polí. Tímto pravidlem se určuje, co se stane, když mají být sloučeny záznamy, u kterých je například různé telefonní číslo nebo číslo účtu.

Pakliže jsou všechna pravidla vypsána, vzniká prostor pro odbornou diskuzi. Výsledkem diskuze by mělo být ustanovení společné řeči – speciální syntaxe – jednotlivých dílčích pravidel. Součástí diskuze by také mělo být hledání duplicitních nebo nadbytečných pravidel.

Pro vysvětlení příklad:

Z debaty vzešly tři pravidla identifikace duplicitních záznamů o zákazníkovi firmy

1) duplicitní záznam = stejné jméno na stejné adrese
2) duplicitní záznam = stejné jméno, adresa a e-mail
3) duplicitní záznam = stejné jméno a e-mail

První pravidlo říká, že jako duplicitní záznamy budou brány všechny takové, které se shodují ve jméně a adrese. Podle třetího pravidla jsou záznamy duplicitní, když se shodují ve jméně a emailu. Nemusí tedy mít stejnou adresu.

Sdružením duplicitních záznamů podle 1. a 3. pravidla a společného jména lze získat všechny duplicitní záznamy v celých datech. Z tohoto důvodu je 2. pravidlo nadbytečné. 2. pravidlo by sice rozpoznalo určité duplicity, ale stane se tak méněkrát, než při 1. a 3. pravidlu.

Jméno	Adresa	E-mail
Jon Smith	100 2nd St.	Smith@abc.com
John Smith	100 2nd St.	Smith@Abacuscomputers.com
J Smith	100 Front St.	Smith@abc.com

Tabulka č.7: Příklady duplicitních záznamů

Bylo předem definováno, že pokud je u jména příjmení shodné, jsou i křestní jména Jon, John a J. stejná. 2nd St. a Front Street jsou také fyzicky identické adresy, s tím, že „Front“ zní „prestižněji“. „Abacus Computers“ byl původní název firmy, ze kterého se postupem času vniknul název „ABC“.

Tento jednoduchý příklad ukazuje, že pravidlem číslo dvě (vzniklým z debaty), nebude identifikován žádný duplicitní záznam. Pouze kombinací dvou více obecných pravidel, 1. a 3., a následným sloučením záznamů podle příjmení, mohou být nalezeny duplicitní záznamy s poměrně vysokou mírou spolehlivosti. Navíc použití 2. pravidla zpomalí proces přiřazovaní záznamů a to bez jakéhokoliv přínosu.

Samotné vypisování jednotlivých pravidel podněcuje firmy k úvaze nad konkrétními pojmy, a tím, co a jak chtějí dosáhnout. Až když jsou jednotlivá pravidla na papíře, zjistí, že text ne zcela popisuje věci tak, jak si myslely.

Vypsání pravidel je důležitým testem. Jestliže se zapojené strany nemohou domluvit, ukazuje to na to, že ještě není pravidlo hotové, nebo že došlo k zásadnímu nepochopení cíle tohoto pravidla. Chceme-li získat jednotný pohled na zákazníka, musí být tato neshoda vyřešena.

4. krok – vytvoření matice pravidel

Vypsaná pravidla, která jsou již schválena a podepsána příslušnými stranami, jsou uložena do tzv. matice pravidel. Účelem matice je shromáždění veškerých pravidel určujících HH na jedno místo. Tato business pravidla budou dále součástí různých dokumentů, jako například: plán projektu, dokument požadavků, zápis o práci nebo v jiných dokumentech poskytovaných klientem.

Pravidla určující HH bývají často uložena na mnoha stránkách plných textu. Vytažení jednotlivých pravidel do jedné matice umožňuje snadnější pohled na souvislosti, odhalí případná nadbytečná pravidla nebo možné vady v logice HH.

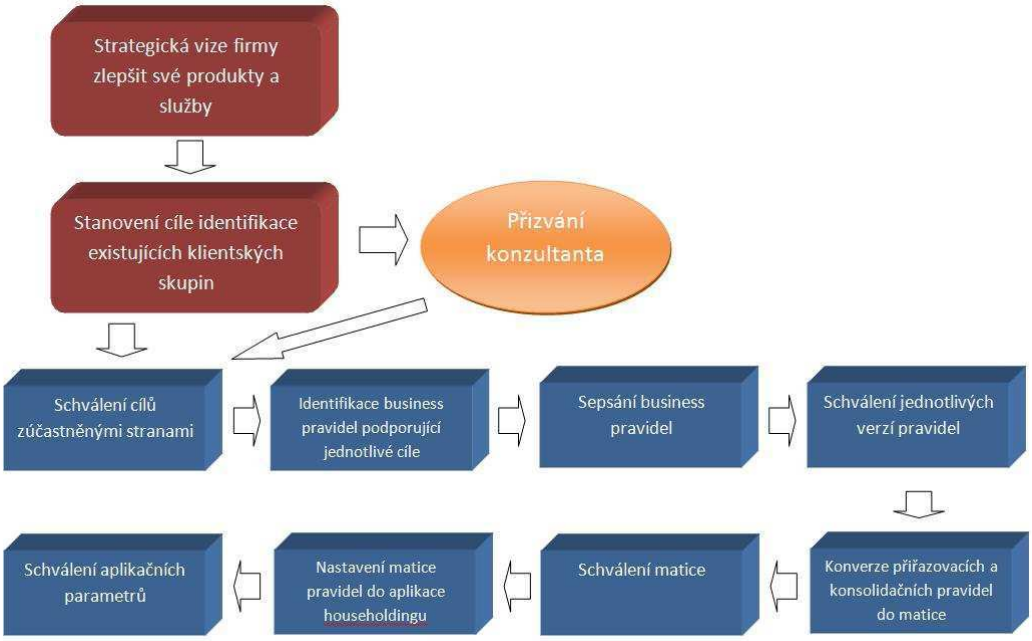
Ve struktuře matice řádky určují pravidla, zatímco sloupce použitelná datová pole. V průsečíku řádků a sloupců jsou specifické podmínky. Těmi může být například při určení HH podle adresy „odpovídá přesně“ nebo „je stejná, pokud jsou stejné první čtyři písmena z města“ atd.

5. krok – ověření matice pravidel

Odpovídající a sjednocená matice musí být ověřena a schválena. Přebytečná pravidla jsou označena, aby mohla být zrušena a případná nová pravidla zvýrazněna, aby bylo jasné, že jsou nová.

6. krok – vytvoření aplikačních parametrů / parametrů aplikace

Následně po schválení matice pravidel se tato pravidla převádějí na aplikační parametry. Firstlogic k tomu využívá svůj speciální grafický nástroj zvaný IQ Suite. Každý řádek matice představuje sadu kritérií, které IQ Suite přebírá jako parametr řízení úlohy



Obrázek č.2: Schéma projektu householdingu dle Firstlogic [Firstlog]

7. krok – kontrola konzistence pravidel

Aplikace používá vlastní vnitřní verifikační logiku na kontrolu soudržnosti a provozuschopnosti pravidel. Vyjasní se všechna varování a chybová hlášení a každé pravidlo je individuálně testováno, a to nezávisle na ostatních. Zákazník poté odsouhlasené výsledky porovná.

Během této fáze mezi sebou jednotlivé strany komunikují, a to proto, aby přesněji vyladily jednotlivé parametry business pravidel. Je třeba se domluvit, jaké varianty jednotlivých kritérií budou akceptovatelné, resp. které různé názvy budou brány za stejné. Například názvy firmy „Hlavatý s.r.o.“ a „Hlavatý spol. s. o.“ jsou vloženy (definovány) jako názvy téže firmy. Po odsouhlasení předpokládaných variant jednotlivých kritérií se kritérium označí jako hotové a pokračuje se dalším testováním.

8. krok – ověření všech přiřazovacích sad

Po kompletním testování každého přiřazovacího kritéria (a jeho různých kombinací) se kritéria spustí, aby se ověřilo, že nějaké business pravidlo nenarušuje ostatní.

Může nastat případ, kdy podle jednoho pravidla daný záznam do skupiny nepatří, i přesto, že podle jiných 5 pravidel určitě ano. V tomto případě je třeba zvážit, jestli není toto pravidlo nadbytečné nebo chybné.

9. krok – ověření konsolidačních kritérií

Poté, co dojde ke konsolidaci duplicitních záznamů, vzniká příležitost přizpůsobit business pravidla nově uzpůsobeným strukturám dat. Přiřazovací kritéria identifikovala záznamy jako duplicitní, ale tyto duplicity existovaly již v samotných zdrojových datech.

Například záznam o zákazníkovi uvádí dvě různé adresy. Jednu pro dodávku zboží a druhou pro fakturaci. Marketingové oddělení považuje adresu fakturace za místo sídla firmy. Zákazník však sídlí na adrese, která je uvedena pro dodávku zboží. Tento problém se řeší tím, že se vytvoří třetí adresní záznam. To v konečném důsledku znamená, že si firma vede tři záznamy pro adresu zákazníka: fakturační, dodávkovou a místo sídla zákazníka.

10. krok – spuštění Householdingu

Po dokončení všech předchozích kroků se spouští proces HH nad reálnými daty.

11. krok – dokončení nahrání dat do systému

Householding byl v projektu počátečním nákladem při vytváření nového CRM systému. Přínosy takovéhoto "předčištění" CRM systému lze spatřovat ve:

- snížení počtu záznamů, které musí systém obsahovat
- schéma záznamů bylo upraveno a ověřeno
- došlo k vyčištění dílčích datových polí (adresy, telefonní čísla, jména)

Toto „předčištění“ výrazným způsobem zvyšuje vypovídající schopnosti CRM systému. Dále napomáhá rychlejšímu a širšímu využití toho systému. Firma tak konečné získá jednotný pohled na vzory chování svých zákazníků.

4.2 Problémy při budování

Žádný projekt a ani householding není výjimkou, se bohužel neobejde bez dílčích problémů. Mezi typické problémy patří:

- obtížné definování householdu
- problémy při identifikaci entit
- problémy s určením významu entit
- příslušnost k více householdům
- práce s odhadem a pravděpodobností
- zákon o nakládání s osobními údaji (zákon 101/2000 Sb.)
- jasné určení adresového bodu
- stárnutí dat i householdů
- a další

Jednotlivé problémy nejspíše nemají univerzální řešení, přesto se pokusím vlastní řešení nabídnout.

4.2.1 Obtížnost definice

Popis problému
Jak je již z úvodu této práce patrné, definování householdingu není jednoduchou záležitostí. Jednotlivých pohledů existuje mnoho. Vpodstatě každá ze z účastněných stran může mít svou vlastní představu. Například zodpovězení, zprvu možná jednoduché otázky, „Kdo tvoří domácnost?“, nemusí být až tak jednoduché. Náleží do domácnosti tchýně, synovec na podnájmu, vnoučata atd.? Tímto se původně jednoduchá otázka stává stále komplexnější. Nalezení konkrétního příkladu, které záznamy do příslušné skupiny náleží, co je vlastním householdem i stanovení jednotné definice, může být obtížné. Pro zdárné započetí projektu je to však nezbytné.
Navrhované řešení
1) vymezit v úvodní fázi dostatečné množství času, protože případné chyby či nedostatky učiněné během této fáze mohou vést až k nezdaru celého projektu 2) vybrat určité modelové situace, které mohou nastat (např. synovec v podnájmu a mnoho dalších), a tyto situace důkladně prodiskutovat, patří-li opravdu do sledované skupiny; jednotlivé situace evidovat pro možné využití v dalších projektech 3) na základě definovaných business účelů stanovit (konzultací se zadavatelem) určité metriky, podle kterých by se hodnotilo, zdali byly konkrétní osoby, firmy apod. přiřazeny do správného HH.
Zodpovědnost za řešení
Ad 1) zadavatel projektu, resp. business jednotka zadavatele, protože

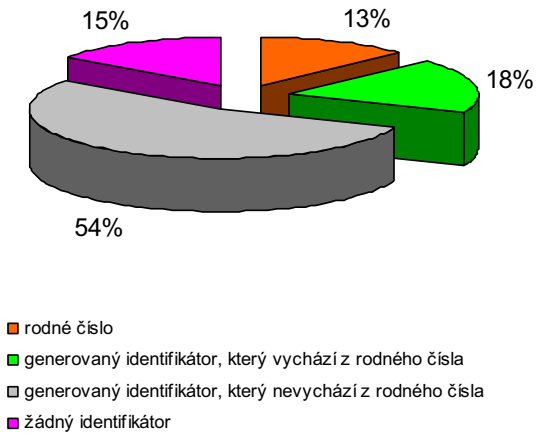
právě zadavatel projekt financuje a délka úvodní fáze je přímo úměrná množství finančních prostředků na householding vynaložených

Ad 2) při výběru modelových situací by měli být zúročeny zkušenosti řešitele HH, za odsouhlasení je zodpovědný zadavatel

Ad 3) za správnost nastavení metrik zodpovídá zadavatel

4.2.2 Identifikace entit

Popis problému
Samotné rozpoznání jednotlivých entit je složité. Problém spočívá zejména v určité nejednoznačnosti nebo lépe řečeno nejednotnosti názvů entit. Například entita „IBM“ je reprezentována názvy „International Business Machines Corporation“ nebo „I.B.M.“. Všechny tyto názvy odpovídají jedné a téže entitě. Jedna firma tak může být vedena ve zdrojových datech pod různými názvy. To činí identifikaci složitější. [Madnick] Problémem, souvisejícím s později zmíněným zákonem na ochranu osobních údajů, je identifikace klienta podle rodného čísla. Z průzkumů vyplývá, že téměř 31% organizací využívá pro identifikaci klientů rodná čísla či jejich odvozeniny. Zdroj [Adastra] uvádí, že legislativní úpravy do budoucna směřují k tomu, že možnost využití rodného čísla bude omezena. Téměř 80% společností to vnímá jako velmi významný problém.



Obrázek č.3: Graf - poměr jednoznačných identifikátorů klienta [Adastra]

Navrhované řešení
1) vytvořit a spravovat externí číselník, ve kterém by se evidovaly možné tvary názvů jednotlivých záznamů. Číselníků by mělo být více, a to podle oblastí, které pokrývají. Zdrojová data pro tyto číselníky by byla získávána ze zdrojových systémů (IS) zákaznických firem a dále doplňována s veřejně dostupných externích číselníků (viz kap. [4.3.1]) 2) vytvoření vlastního identifikátoru jednotlivých záznamů, který by usnadnil práci při dalším zpracovávání
Zodpovědnost za řešení
Ad 1) zadavatel projektu, případně příslušná business jednotka zadavatele Ad 2) zadavatel projektu

4.2.3 Význam entit

Popis problému
Po zjištění, že názvy „IBM“, „International Business Machines Corporation“ i „I.B.M.“ jsou názvy pouze jedné entity, je zapotřebí zjistit, co přesně tato entita znamená, nebo které další entity s ní

souvisejí. [Madnick] spatřuje problém též ve značné složitosti firemních struktur. Ty obsahují například různé dceřiné společnosti, obory, divize, větve atd. Například při zjišťování počtu zaměstnanců IBM je nezbytná znalost toho, že je Lotus „dcerou“ IBM a tedy i jeho zaměstnanci by se měli započítávat.
Navrhované řešení
1) evidence klientů (záznamů) prostřednictvím stromových struktur. Tímto způsobem se získá dodatečná informace (vazba) napomáhající efektivnějšímu utváření HH. 2) využitím externích číselníků (viz [4.3.1]) je možné ověřit, zdali se u konkrétní entity jedná o název ulice nebo jméno nadnárodní společnosti
Zodpovědnost za řešení
Ad 1) zadavatel projektu Ad 2) řešitel projektu

4.2.4 Příslušnost k více householdům

Popis problému
Z definice householdu vyplívá, že určitý subjekt může patřit do více householdů. Například při uplatnění pravidla horizontálního vztahu (otec – syn) přísluší subjekt do jednoho HH, a při uplatnění pravidla telefonního čísla do jiného. To může být u některých typů HH žádoucí, ale například chceme-li doručit marketingový materiál na konkrétní adresu, nemůže být oslovená osoba ve více householdech.

Navrhované řešení
vytvoření různých typů HH podle potřeby jejich využití (více viz kap. [3.2])
Zodpovědnost za řešení
řešitel projektu

4.2.5 Práce na základě odhadu (pravděpodobnosti)

Popis problému
Přiřazení záznamu k householdu není vždy zcela jednoznačné. Proces HH se opírá o odhady na základě různých informací evidovaných o klientech. Ty mohou, ale také nemusejí, věrohodně vypovídat o vztazích mezi klienty. Důsledkem tohoto faktu je to, že: 1) zařazení konkrétního klienta do HH je vždy pouze více či méně kvalifikovaným odhadem 2) výsledek bude pokrývat existující HH pouze z části. Z existujících dat se nenechají, při požadavku na „rozumnou“ spolehlivost, všechny HH odhalit **Snahou při householdingu je dosahovat přesnějších výsledků. Je však vhodné si uvědomit, že nic není stoprocentní a úspěšnost celého procesu HH by vždy měla být poměřována v kontextu business účelu a náklady na realizaci HH.**
Navrhované řešení
1) vyjádřit míru pravděpodobnosti - společnost Adastra pracuje s pojmem „míra příslušnosti k householdu“. Zjednodušeně to tedy znamená, že se při HH pracuje s pravděpodobností, a tedy že existuje jistá míra nejistoty či nejednoznačnosti. Cílem je tuto nejistotu či nejednoznačnost co nejvíce eliminovat a dosahovat tak přesnějších výsledků práce.

Celková míra příslušnosti klienta k householdu je vyjádřena na škále 0 - 100 bodů. Sestává se z příspěvků zohledňujících stáří informace a kvalitu použitého pravidla. Při hodnotě 100 jsme si jisti, že záznam do daného householdu patří. Hodnota 0 znamená minimální důvěru, že daný záznam ještě do householdu patří. 2) smířit se s tím, že householding nevytváří 100% přesné údaje, výstupy mají určitou míru jistoty a jsou tedy „přibližně přesné“ 3) stanovit, jaká míra jistoty bude vyžadována (ta může být pro různé typy HH odlišná, viz kapitola 3) 4) určit, jaké dopady může mít případné neodhalení HH či chybné zařazení daných subjektů
Zodpovědnost za řešení
Řešení tohoto problémů vzniká postupnou diskuzí zúčastněných stran.

4.2.6 Zákon 101/2000 – nakládání s osobními údaji

Popis problému
Dnem 1. července 2000 nabyl účinnosti zákon 101/2000 Sb. (zákon o ochraně osobních údajů). Tento zákon do jisté míry zkomplikoval práci mnoha firmám. Znamená pro ně mnohdy dodatečné administrativní i finanční náklady. Pro správce dat představuje řadu povinností, které znesnadňují práci s daty. Ovlivněn tímto zákonem je tedy i householding. Osobní údaje znamenají data, z nichž je možné zjistit identitu osoby, jíž se týkají, typicky tedy půjde třeba o jméno, adresu, telefon, rodné číslo, číslo občanského průkazu, řidičského průkazu, SPZ. Každý provozovatel systémů, které obsahují osobní údaje, musí splnit určité základní povinnosti. Mezi tyto povinnosti například patří: • stanovit účel, ke kterému mají být osobní údaje zpracovány

- shromažďovat osobní údaje odpovídající pouze stanovenému účelu a rozsahu nezbytnému pro naplnění stanoveného účelu
- uchovávat osobní údaje pouze po dobu, která je nezbytná k účelu jejich zpracování
- shromaždovat osobní údaje pouze otevřeně; je vyloučeno shromaždovat údaje pod záminkou jiného účelu nebo jiné činnosti, pokud zvláštní zákon nestanoví jinak
- nesdružovat osobní údaje, které byly získány k rozdílným účelům, pokud zvláštní zákon nestanoví jinak

Všechny organizace, které zpracovávají osobní údaje, se navíc musejí registrovat na Úřadě pro ochranu osobních údajů.

Další povinností je nutnost získat souhlas subjektu údajů ještě před zpracováním (i když existuje řada výjimek). HH většinou pracuje již s existujícími daty. Ty navíc byly původně získávány za jiným účelem. Pro marketingové a obchodní účely je sice možné za určitých podmínek tyto data zpracovávat, znamená to však další komplikace.

Pro HH tento zákon představuje významný problém, protože **právě HH pracuje s osobními údaji, které byly mnohdy pořízeny za jiným účelem.**

[Sagit]

Navrhované řešení

1) důkladně se seznámit se zákonem 101/2000 Sb. a minimalizovat tak dopad možných postihů za aktivity, které nejsou v souladu s tímto zákonem

2) householding je možné do jisté míry chápat jako marketingový nástroj. Zákon umožňuje právě pro tyto účely daná data zpracovávat.

3) **vyžádat si souhlas (i dodatečný) pro nakládání s osobními údaji pro účely HH – například vložením klauzule do všeobecných**

podmínek při podpisu smlouvy atp.
Zodpovědnost za řešení
Ad 1) řešitel projektu zodpovídá za řešení, zadavatel projektu za data poskytnutá v souladu se zákonem Ad 2) zadavatel projektu

4.2.7 Určení adresového bodu

Popis problému
Identifikaci domácnosti znesnadňují jistá národní specifika České republiky. Jedná se zejména o chybějící údaj o čísle bytu při zápisu adresy (např. v USA tento údaj povinný je). Z tohoto důvodu je nutné hledat jiné cesty, které vedou k identifikaci domácnosti. Důležité je také rozlišovat adresu bydliště (sídla) a adresu kontaktní. Ty se v mnohých případech liší. Větší firmy mají více adres, což může být v některých případech problém.
Navrhované řešení
1) doplnění chybějící informace o adrese klienta je možné získat prostřednictvím **dodatečné informace o vzájemném vztahu**. Tyto informace lze získat z různých smluv, kde bývá uveden například pojistník – pojištění (v případě, že se jedná o dvě různé osoby), dlužník – věřitel atd. Chybějící údaj o čísle bytu je do jisté míry možné nahradit získáním telefonního čísla hlavní telefonní stanice klienta. 2) požádat klienta o dodatečné doplnění chybějících údajů. 3) využitím externích zdrojů – například data získaná na Českém úřadu katastrálním a zeměměřickém. (Pozn. přístup k těmto datům není v současné době nejjednodušší a je zpoplatněn - pro širší využití nemusí být tedy zcela optimální) 4) v úvodu projektu stanovit, které adresy budou pro konkrétní typ

householdu relevantní
Zodpovědnost za řešení
Ad 1) dodatečné informace o vzájemném vztahu může identifikovat řešitel projektu, za ověření jejich platnosti však zodpovídá zadavatel Ad 2) zadavatel případně uživatel transakčních systémů, ze kterých se získávají data pro HH Ad 3) za využití externích číselníků všeobecně zodpovídá řešitel projektu Ad 4) zadavatel případně uživatel výstupů HH

4.2.8 Stárnutí dat i householdu

Popis problému
Problémem při householdingu je stárnutí dat, ke kterému dochází již během projektu. Stárnutí ovlivňuje vypovídající schopnost HH. Čím je informace při tvorbě householdu starší, tím nižší je míra jistoty, že záznam stále do daného householdu patří. Během života lidí dochází k mnohým změnám. Přejmenovaní při svatbě, změna bydliště nebo dokonce změna názvu ulice, a také výměna telefonního čísla patří mezi nejběžnější. Tyto změny bývá obtížné zachytit a také promítnout do HH. Stárnou i vytvořené householdy. Například jasně definovaná skupina dnes nemusí být tou, kterou by chtělo oslovit marketingové oddělení za 2 roky. Stejně jako život kolem nás, tak i householdy se mění.
Navrhované řešení
1) již v úvodu projektu stanovit, jak časté budou reporty (výstupy) z HH systému a tomu přizpůsobit aktualizaci dat a tedy i householdů. Jedná se o nastavení „vnitřního mechanismu“ zajištujícího oživovaní platnosti dat.

2) určit, jaká úroveň neměnnosti dat bude vyžadována, a také co se stane, když dojde ke změnám v těchto datech. 3) určit míru významnosti těchto změn s ohledem na kvalitu požadovaných výstupů 4) uplatnit „míru příslušnosti“ k HH a její automatická aktualizovatelnost viz řešení 1) v kap. 4.2.5
Zodpovědnost za řešení
Ad 1) zadavatel projektu případně osoby, pro které budou výstupy householdingu vypracovávány Ad 2) a 3) zadavatel projektu

4.3 Podmínky nutné a postačující

Postačující podmínku lze spatřit ve kvalitních externích číselnících. Jejich využití není sice pro householding nezbytné, práci však výrazným způsobem zjednodušují. Naopak nutnou podmínkou, bez které se householding neobjede, jsou kvalitní data.

4.3.1 Kvalitní externí číselníky

Při householdingu se využívají externí zdroje dat. Jedná se o různé číselníky, rejstříky či seznamy. Kvalita těchto zdrojů je různá stejně jako možnost jejich aktualizace. Některé číselníky jsou ze zákona spravovány státem, jiné komerčními subjekty. Od toho se také odvíjí fakt, že některé zdroje nemusejí být zdarma.

Mezi externí informační zdroje patří například:

číselníky jmen a příjmení
referenční databáze adresných míst
PSČ
referenční databáze ekonomických subjektů
telefonní seznam
a další

Tabulka č.8: Externí zdroje informací

Číselníky jmen, příjmení a titulů

Číselníky jmen a příjmení jsou spravovány Ministerstvem vnitra ČR a vystavovány na stránkách ministerstva[1]. V sekci „statistiky“ lze nalézt odkaz na „četnost jmen a příjmení“ (http://www.mvcr.cz/sprava/informat/cetnost/index.html), který obsahuje příjmení a jména vyskytující se s četností 5 a více. Ty jsou dále rozdělena podle pohlaví. Na této adrese lze také nalézt číselníky četnosti titulů, jmen cizinců.

Perioda aktualizace těchto číselníků není známa.

Referenční databáze adresných míst

Jako referenční databázi adresných míst v ČR je využíván Územně identifikační registr adres (UIR-ADR) vydávaný Ministerstvem práce a sociálních věcí ČR. Na stránkách ministerstva[2] (http://forms.mpsv.cz/uir/popis/popis.jsp) lze nalézt podrobnější informace o přístupu k těmto datům.

V současné době obsahuje struktura UIR-ADR následující celostátní číselníky:

Oblastí	Správních obvodů
Krajů	Městských částí a městských obvodů
Okresů	Částí obce
Obvodů ORP	Ulic a veřejných prostranství
Obvodů POÚ	Stavebních objektů
Obcí	Adresných míst
Pražských obvodů	Adresných pošt
NUTS4-obvodů	Správních obvodů

Tabulka č.9: Struktura číselníků UIR-ADR

[1] www.mvrc.cz

[2] www.mpsv.cz

UIR-ADR je aktualizován přibližně 1x týdně, přičemž jsou číselníky poskytovány zdarma.

PSČ organizací

Pro kontrolu a doplňování PSČ organizací se využívá číselník České pošty – PSČ organizací. Přístup k tomuto číselníku je zpoplatněn částkou 1000 Kč za rok využívání. Objednat jej lze na adrese http://www.cpost.cz/ . S každým zájemcem o tyto data je sepisována řádně vyplněná smlouva a objednávkový list.

Aktualizace probíhá jedenkrát měsíčně.

Česká pošta poskytuje i jiné využitelné číselníky.

Referenční databáze ekonomických subjektů

Jako referenční databází ekonomických subjektů lze pro prostředí České republiky využít Registr ekonomických subjektů (RES) vytvářený a distribuovaný na základě §20 zákona č.89/1995 Sb. Českým statistickým úřadem.

Do registru ekonomických subjektů se zapisují, pro účely householdingu možná ne všechny, využitelné údaje:

identifikační číslo (IČO)	**u fyzické osoby rodné číslo**
obchodní firma (název)	počet zaměstnanců
u právnických osob a organizačních složek státu sídlo, u fyzických osob místo podnikání	výše příjmu (výnosu)
územní identifikace sídla a místa podnikání, poštovní adresa, popřípadě telekomunikační spojení	výše základního kapitálu nebo nadačního jmění či kmenového jmění
předmět podnikání nebo jiné činnosti	výměra obhospodařované zemědělské a lesní půdy
právní forma	výměra obhospodařované orné půdy
datum vzniku právnické osoby a organizační složky státu; u fyzické osoby datum vydání povolení nebo dokladu o registraci nebo zápisu opravňujícího ji k podnikání	**skutečnost, zda subjekt skutečně podniká, či nikoliv**
datum a příčinu zániku právnické	hodnota zboží přijatého za příslušné

osoby a organizační složky státu; u fyzické osoby datum a příčinu zániku oprávnění podnikat	kalendářní období z ostatních členských zemí Evropských společenství
kategorie velikosti ekonomického subjektu	hodnota zboží odeslaného za příslušné kalendářní období do ostatních členských zemí Evropských společenství
sektor pro národní účty (zařazení do klasifikace)	

Tabulka č.10: Informace obsažené v registru ekonomických subjektů

Cena těchto dat však bohužel není nezanedbatelná. Základní jednorázová cena RES – ČR kompletní databáze, zahrnující všechny fyzické a právnické osoby, tj. včetně osob nezapsaných v obchodním rejstříku, činí 25000 Kč.

Informace o RESu lze nalézt na internetové adrese http://www.czso.cz/csu/redakce.nsf/i/registr_ekonomickych_subjektu . Aktualizace probíhá 1x měsíčně.

Telefonní seznam

Data telefonního seznamu živnostníků a firem poskytuje společnost Mediatel spol. s r.o. Za určitý poplatek, který se odvíjí od počtu záznamů, poskytuje společnost přístup do své databáze. Maximální počet záznamů je však omezen. Bližší informace o této službě lze získat na infolince společnosti 800 12 12 12 a to přímo u paní Boudníkové. Se zájemci o tyto data jsou dohodnuty individuální podmínky.

Telefonní seznam fyzických osob, známý též jako „Zlaté stránky“, v elektronické podobně, resp. ve formě databáze, k dispozici není. V roce 2004 bylo pro malý zájem zrušeno vydávání CD nosiče, který tyto data obsahoval.

Pozn.: S rostoucím počtem mobilních telefonních čísel postupně klesá informační význam tohoto externího zdroje dat. Mobilní čísla se, i přes mnohé snahy, nedaří evidovat. Proto je využití telefonního čísla, jakožto externího zdroje informací, omezené a má spíše doplňující informační funkci.

4.3.2 Kvalitní a čistá data

Předpokladem uplatnění principů householdingu je práce s vyčištěnými a kvalitními daty. Bez nich by mohl projekt householdingu velice rychle ztroskotat. Samotný proces čištění zákaznických dat a identifikace zákazníka je tedy nezbytně nutnou podmínkou pro úspěšný householding. Některé zdroje považují projekt čištění dat za součást householdingu, jiné jej oddělují. Tak či tak je datová kvalita s householdingem provázána.

Když kvalitně nepřipravíme data, výsledek celého procesu householdingu nebude též kvalitní.

Pojem „kvalita dat“ může být pro někoho neuchopitelný. Významní autoři, z jejichž myšlenek mnozí čerpají, definují datovou kvalitu následovně:

[Redman] uvádí, cituji: *„Data jsou vysoké kvality tehdy, pokud jsou vhodná pro jejich zamýšlené použití v provozu, rozhodování a plánování. Data jsou vhodná k zamýšlenému použití tehdy, když jsou prostá defektů a mají požadované vlastnosti.“*

[Olson] definuje datovou kvalitu prostřednictvím šesti základních vlastností. Data jsou kvalitní tehdy, jsou-li přesná, dostupná, relevantní, úplná, srozumitelná a důvěryhodná.

- přesná – znamená, jaké procento případů záznamů obsahuje skutečnou hodnotu odpovídající realitě (například jaké procento rodných jmen je v systému správně uloženo).
- dostupná – znamená, jsou-li data uživateli přístupná (například při inventuře je žádoucí, aby ten, kdo inventuru vykonává, měl možnost nahlédnout do seznamu zboží, jinak by samotná inventura postrádala smysl).
- relevantní – neboli platná, znamená použitelnost dat v rozhodovacím procesu, a to z hlediska jejich vztahu k popisovanému objektu (například v databázi koček nemá smysl sledovat počet jejich křídel).
- úplná – úplnost udává procentní poměr mezi zadanými a nezadanými údaji.
- srozumitelná – tato charakteristika je dosti subjektivní. Různí uživatelé dat mohou mít různé zkušenosti, požadavky.

- důvěryhodná – odráží míru spolehnutí se na daná data

Mezi synonyma nebo pojmy, které mají k datové kvalitě blízko patří: kvalita dat, čistota dat, profilování dat, audit dat, správnost a úplnost dat.

Příčiny „nekvalitních" dat

Zdrojů chyb a dalších typů „nekvalit" může být v zásadě mnoho. Při jejich analýze byl pro účely této práce využit odborný průzkum společnosti Markent, který byl proveden v roce 2006.

Průzkum byl zaměřen na kvalitu dat v informačních systémech českých společností. Bylo osloveno přibližně sto společností s obratem vyšším než 250 mil. Kč s více než sto zaměstnanci. Oslovené firmy působí ve všech významných segmentech trhu - finanční sektor, utility, telekomunikační společnosti a obchodní řetězce.

Průzkum ukázal, že zdrojem chyb a dalších typů „nekvalit" jsou v zásadě dvě skutečnosti. Nejvýznamnější je lidský faktor a nekonzistence dat v různých autonomních systémech.

Lidský faktor

V souvislosti se sníženou kvalitou dat se nejčastěji uvádí lidský faktor, a to ze 70 %. Data, se kterými pracují počítačové systémy, do nich často vkládají lidé. Ti často dělají různé chyby. Mezi nejběžnější patří například překlepy, zápisy do nesprávných polí a neznalost. Dále také nezanedbatelná tvořivost a „invenčnost", které však často nebývají žádoucí.

Do oblasti vlivu člověka lze zahrnout i další faktory. Je to například nevyškolenost personálu, nedbalost, stres, čas a tlak na rychlost při pořizování záznamů. Vedoucí IT pracovníci vidí dále jako závažný problém chybějící motivaci pro zadávání kvalitních dat.

Integrace systémů

Druhý důvod je dán tím, že informační systémy podniků dnes sestávají z mnohdy autonomních agend, zaměřených na různé oblasti, založených na různých technologiích, vytvořených různými dodavateli v různých dobách. Problém dále vzniká v případech,

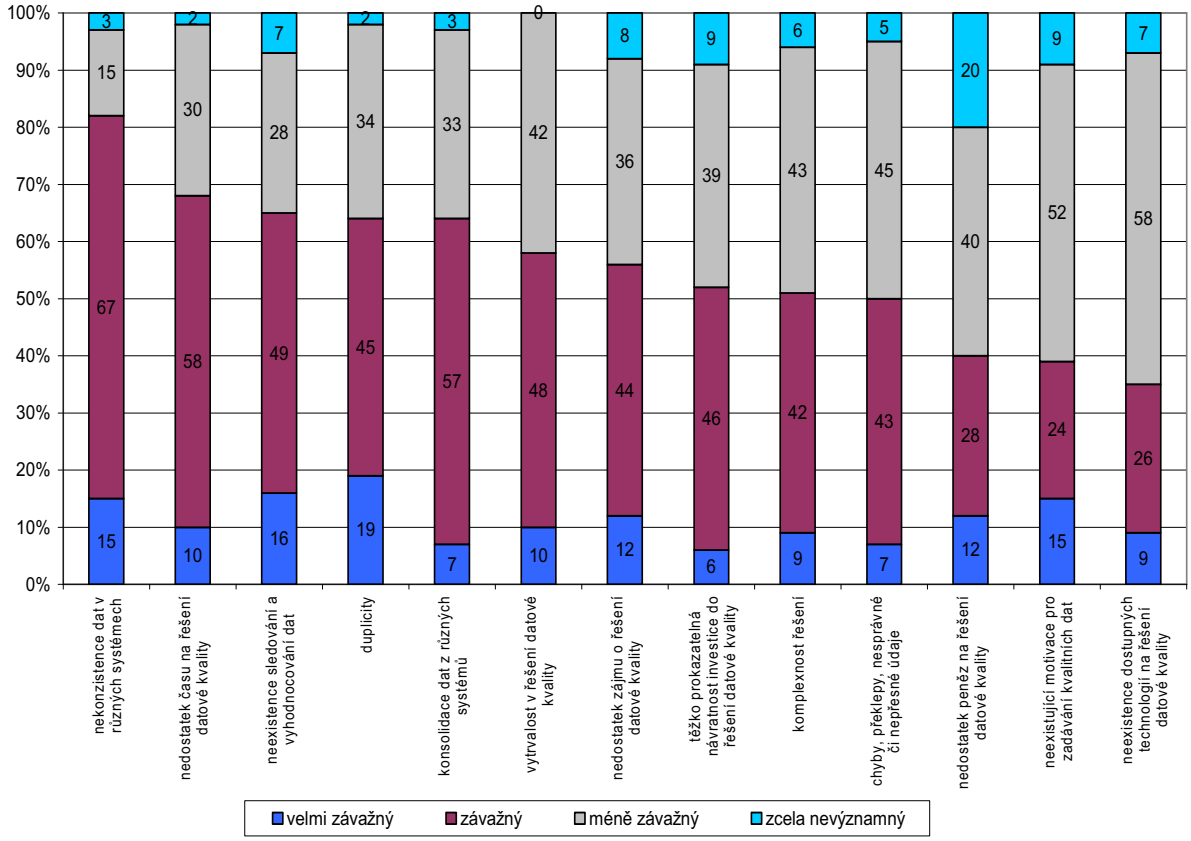
kdy v různých systémech platí různá pravidla pro zápis některých hodnot. Tím se rozumí například zápis rodného čísla jednou s lomítkem a podruhé bez, používání diakritiky či velká a malá písmena.

Mezi další nezanedbatelné stavební kameny datových „nekvalit“ podle [Adastra, Civín, Kyjonka] patří:

- nedostatek času na řešení datové kvality
- nedostatek peněz
- chybějící zájem řešení datové kvality
- chyby HW/SW případně špatně navržené datové struktury
- neexistence kontrol při zadávání
- neexistence jednoznačného identifikátoru
- zastarávání systémů včetně nedostatečné podpory
- a další

Závažnost problémů charakterizují čísla v následujícím grafu[1]:

[1] Graf převzat z materiálů společnosti Adastra [www.adastra.cz]

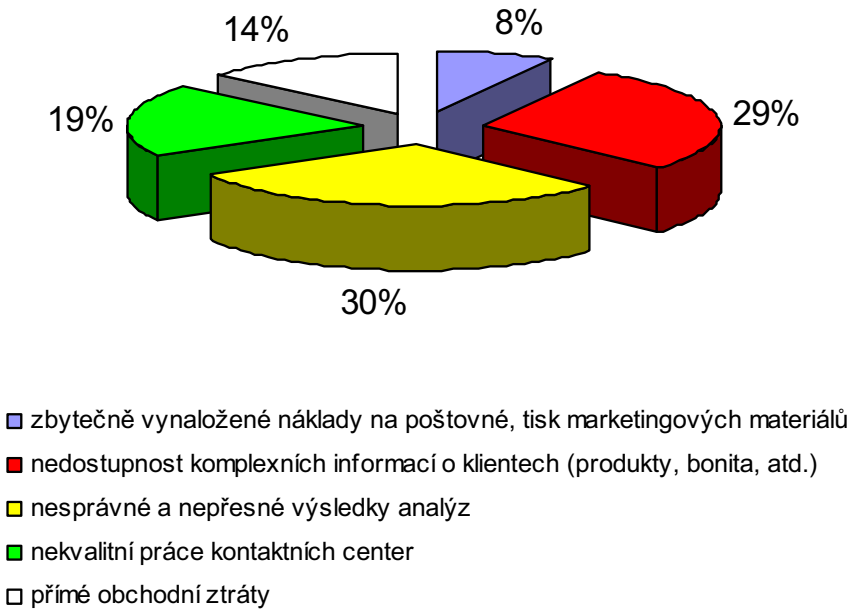


Obrázek č.4: Závažnost problémů v oblasti kvality dat

Důsledky datové „nekvality"

Datová „nekvalita" má ve fungování firem mnoho důsledků. Může to znít paradoxně, ale firmy, dle [Zavoral], „*věnují mnohem více času a prostředků analýze dat než péči o jejich kvalitu a správnost. Lidé tak používají analytické nástroje pro datové rozbory a automaticky předpokládají, že pracují se správnými daty*".

Nejvýznamnější důsledky datové „nekvality" popisuje následující graf:



[Adastra]

Obrázek č.5: Graf - důsledky „nekvality“ v datech

[Zavoral] uvádí, že dle zprávy analytické společnosti Gartner Inc[1]. je **25% dat používaných velkými korporacemi vadných**. Důvodem je jejich chybné zadávání způsobené lidským faktorem, dále změnami v zákaznických profilech (např. změna adresy, pozice atd.) a nedostatkem vhodných datových standardů v korporacích. Důsledkem jsou pak chybné a klamné statistiky, někdy i klesající tržby. Gartner navíc předvídá, že by se tato situace v nejbližší době zlepšovat neměla.

Na finanční ztráty způsobené datovou nekvalitou poukazuje i studie poradenské společnosti PricewaterhouseCoopers[2], která hovoří o tom, že **75% z 599 sledovaných firem má problém s datovou kvalitou,** která jim způsobuje přímé finanční ztráty.

Data Warehousing Institute[3] odhaduje, že **tyto ztráty mohou v amerických organizacích činit celkově více než 600 miliard USD.**

---

[1] Gartner Inc. (květen 2005) – www.gartner.com

[2] PricewaterhouseCoopers (2002) – www.pwc.com

[3] Data Warehousing Institute – www.dw-institute.com

Zajímavá čísla poskytuje průzkum realizovaný společností Adastra.

- 57 % společností považuje vliv nekvalitních informací na fungování organizace za kritický a velmi vážný
- 82 % respondentů uvádí jako nejzávažnější problém v kvalitě dat nekonzistentnost dat
- 80 % dotázaných si uvědomuje problém s legislativou a využitím rodného čísla jako identifikátoru zákazníka
- ve více než 70 % společností panuje ochota a chuť managementu řešit problematiku datové kvality

[Adastra]

Situace kolem datové kvality by se dala shrnout slovy společnosti Gartner Inc.: *„**víme stále více, ale rozumíme stále méně.**“*

Řešení datové „nekvality"

Již samotné IS se snaží s datovou kvalitou vypořádat. Jejich funkce jsou však omezené. Pro řízení datové kvality je tedy zapotřebí nastavit správné procesy a definovat strategické přístupy do budoucna. V zásadě existují dva přístupy. Jedním z nich je řešit datovou kvalitu vlastními silami a vlastní metodikou. Druhou cestou je procesní zajištění DQ a využívání specializovaného nástroje

Problematikou kvality dat se dlouhodobě zabývá několik mezinárodních společností. Mezi významnější patří například Trillium, Firstlogic/Business Objects či Adastra.

Shrnutí

Chyby v datech se objevují od samotného počátku jejich automatizovaného zpracování. Neřešení těchto chyb významným způsobem snižuje správnou použitelnost těchto dat a to nejen pro účely householdingu.

Pro využití efektů householdingu je nezbytným předpokladem práce s čistými a kvalitními daty. Téma datové kvality je velice rozsáhlé a vydalo by jistě na mnoho

diplomových prací. Pro účely této práce bylo vybráno, dle mého názoru, to nejpodstatnější.

4.4 Závěrečné shrnutí kapitoly

V této kapitole se podařilo strukturovat postup budování householdingu tak, jak jej významné společnosti definují. Z různých částí postupů byl vytvořen průnik a vzniklo tím rozdělení do tří fází. Detailnější popis jednotlivých fází obsahuje úvod kapitoly. Dále se podařilo identifikovat problémy, se kterými se household ing potýká.

Materiálů, týkajících se budování householdingu nebo popis vlastní charakteristiky tohoto procesu, není mnoho. Důvodem může být určitá obava z „odkrytí karet“ a prozrazení těžce získaného know how. Dalším potenciálním důvodem může být fakt, že jednotlivé firmy samotný postup ani sepsaný nemají.

K nejobsáhlejšímu materiálu, v práci uvedeném, jsem se dostal díky vedoucímu diplomové práce, panu Ing. Davidu Slánskému, který je Lead Business Consultant ve společnosti Adastra. Dalším velmi obsáhlým dokumentem, co se informací týče, je materiál společnosti Firstlogic - Household ing: An 11 – Step Approach. Některé zdroje obsahují pouze kusé informace.

Asi největším rozdílem v chápání householdingu a jeho budování je fakt, že některé společnosti (Adastra, Firstlogic) jasně oddělují proces čištění dat a samotný household ing. Naopak např. společnost Harte-Hanks tyto dvě části spojuje. U větších projektů, je dle mého názoru výhodnější přístup první. Rozdělení je výhodné zejména z důvodů značné časové, finanční a lidské náročnosti obou „subprojektů“, a také proto, že household ing chápu jako určitou nadstavbu řešení datové kvality. U menších projektů může docházet ke sloučení částí, otázkou však je, jestli ne na úkor kvality. To nejsem schopen, na základě svých současných znalostí, posoudit.

Nejzajímavější, ale také bohužel méně zdokumentovanou, součástí householdingu je vytváření pravidel, která určují zařazení do household u. Postup vytváření je vhodně popsán Adastrou [4.1.1] a částečně též u Firstlogic [4.1.3]. Zajímavá je snaha o kvantifikování rozdílností, které mohou nastat. Například household v malém či velkém městě, různá četnost jmen, stárnutí dat, atd. Váha jednotlivých pravidel může být různá.

Klíčovým momentem je úvodní fáze, ve které se ujasňuje, co se od projektu očekává. Zúčastněné strany si vzájemně sjednotí terminologii, dohodnou případné termíny a stanoví podmínky. Kritickým faktorem úspěchu je definování business účelů, pro který je householding budován. Tento bod se může zdát být samozřejmým a možná i banalitou, praxe a také časté zmínky v literatuře však ukazují opak.

Implementaci a realizaci householdingu někteří autoři odbudou slovy „nahrání dat do systému“.

5 Závěr

Householding je jedním z přístupů, který napomáhá společnostem lépe využít potenciál skrytý v jejich datech. Fascinovala mě možnost využití znalosti různých skupin ekonomických subjektů, a to zejména pro marketing a prodej.

Mým přáním a cílem bylo napsat takovou práci, která by poskytovala ucelený pohled na problematiku householdingu. Ne příliš složitou, aby měla přínos i pro laika, ale zároveň s dostatečným množstvím informací, které by mohly využít ti, kteří se o tuto problematiku zajímají. Má diplomová práce může posloužit jako podklad pro další práce studentů. Věřím, že se mi podařilo vytyčený cíl splnit.

Práce je rozdělena do 4 kapitol. Ucelený přehled se bezpochyby neobejde bez jasného vydefinování samotného pojmu householding. Tomuto je věnována 2. kapitola, ve které se zabývám pohledy a názory světových autorů. Jednotlivé názory se odlišují, a to zejména v oblasti zájmů spadajících do této problematiky. Někteří autoři chápou householding jako slučování klientů, jiní jako slučování záznamů o těchto klientech. Existují však určité jednotící prvky. Průnik názorů i definic mi umožnil vytvořit definici vlastní: **Householding je proces rozpoznávání ekonomicky významných skupin subjektů** (více k definici viz kapitola 2.1.4). V práci je dále vysvětleno, kde se dají hledat ekonomické přínosy householdingu (kapitola 2.2 a 3.1). Jádrem práce je 4. kapitola, ve které popisuji postup budování, problémy, které tento postup provázejí a také postačující a nutné podmínky, bez kterých se householding neobejde. Rozdílnost v pojetí budování householdingu se projevila u řešení kvality dat. Některé společnosti řeší kvalitu během projektu, jiné householding na kvalitu dat navazují. Práce obsahuje porovnání těchto přístupů. Rád bych na tomto místě podotknul, že **bez kvalitních dat není možné docílit kvalitních výsledků householdingu.**

Dílčím cílem bylo nalezení významných autorů a společností, zabývajících se touto problematikou. U českých autorů a firem se mi potvrdila úvodní hypotéza o nedostatečném informačním pokrytí. Nevylučuji to, ale kromě Adastry a odborníků s ní spjatých, se mi nepodařilo nalézt veřejně dostupný materiál popisující detailněji tuto problematiku. Zahraniční autoři jsou uvedeni v seznamu použité literatury. Mezi ty nejvýznamnější, ze kterých v práci čerpám, patří FirstLogic, Sacolick, Madnick, Mancuso a Moreno.

Závěrem bych rád uvedl, že **projekt householdingu je úspěšný tehdy, přesáhnou-li jeho ekonomické přínosy náklady na jeho realizaci a dojde-li k naplnění business účelů, pro který byl budován.**

6 Seznam použité literatury a informačních zdrojů

6.1 Literatura

[Dyché] Dyché, J.: The CRM Handbook: A Business Guide to Customer Relationship Management. Addison Wesley. 2005. ISBN 0-201-73062-6

Dyché, Jill, Levy, Evan: Customer Data Integration: Reaching a Single Version of the Truth, 2006. ISBN 0471916978

[English] English, L.P.: Improving Data Warehouse and Business Information Quality: Methods for Reducing Cost and Incresing Profits, 1999, ISBN 0471253839

[Kimball] Kimball, R., Caserta, J.: The Data Warehouse ETL Toolkit. Wiley Publishing, Inc, 2004. 491s. ISBN 0-7645-6757-8

[Novotný] Novotný, O., Pour, J., Slánský, D: Business Intelligence, Grada, 2004, 254s. ISBN 80-247-1094-3

[Olson] Olson, J.: Data Duality – The Accuracy Dimension. Morgan Kaufman Publisher, 2003. 294s. ISBN 1558608915

[Redman] Redman, T.: Data Duality – The Field Guide. Digital Press, 2000. 256s. ISBN 1-55558-251-6

6.2 Články z časopisů

[Civín] Civín, L.: Kvalita dat v českých společnostech; In: IT Systems 8/2006, strana 47

[Klofáčová] Klofáčová, Š.: Householding v bankovnictví; In: Bankovnictví, 9/2003

[Kyjonka] Kyjonka, V.: Důvěřujte svým datům – datová kvalita pod lupou; In: IT Systems 6/2006, strana 18

[Zavoral] Zavoral, P.: Data jako zdroj konkurenční výhody; In: Moderní řízení 7/2006, strana 19

6.3 Internetové zdroje

[Adastra] Adastra s.r.o. *[online]*, URL: www.adastra.cz

Materiály a interní dokumentace Adastry

[Firstlog] FirstLogic, *[online]*, URL: www.firstlogic.com

Firstlogic: Householding - An 11-Step Approach *[online]*. Dostupný z WWW: <http://www.csdk.dk/B6473516-7FA1-4104-A6D4-A189A65D2DFF>

[Flux] DataFlux: Data Householding *[online]*. Dostupný z WWW: <http://www.dataflux.com/Technology/Methodology/Data-Integration/Data-Householding.asp>

[Socolick] Sacolick, J.: Householding and Its Relationship to Customer-to-Customer Marketing and Data Mining *[online]*. Dostupný z WWW: <http://www8.sas.com/scholars/05/PREVIOUS/1999/pdf/085.pdf>

[Sagit] www.sagit.cz: Zákon č. 101/2000 Sb. *[online]*. Dostupný z WWW: <http://www.sagit.cz/pages/sbirkatxt.asp?zdroj=sb00101&cd=76&typ=r>

[Madnick] Madnick, S., Wang R., Xian X.: The Design and Implementation of a Corporate Householding Knowledge Processor to Improve Data Duality *[online]*. Dostupný z WWW: <http://ssrn.com/abstract=461001>

Madnick, S., Wang R., Zhang X.: A Framework for corporate householding *[online]*. Dostupný z WWW: http://web.mit.edu/tdqm/www/tdqmpub/CHH-ICIQNov02fixed.pdf

[Mancuso] Mancuso, G., Moreno A.: Data Integration: The Common Problem – Working with Merge/Purge and Householdingu *[online]*. Dostupný z WWW: <http://www.dmreview.com/article_sub.cfm?articleID=1006332>

[Vavruška] Vavruška, J.: ETL a kvalita dat *[online]*. Dostupný z WWW: <http://www.systemonline.cz/site/data-warehousing/kvalita4.htm>

[Zornes] Zornes, A.: Taking Customer Data Integreation: Master Data Management Milestones, *[online]*. Dostupný z WWW: <http://www.dmreview.com/article_sub.cfm?articleId=1046694>, citován 23.8. 2007

7 Seznam obrázků a tabulek

7.1 Obrázky a grafy

Obrázek č.1: Schéma principu dvouúrovňové identifikace 26

Obrázek č.2: Schéma projektu householdingu dle Firstlogic [Firstlog] 39

Obrázek č.3: Graf - poměr jednoznačných identifikátorů klienta [Adastra] 44

Obrázek č.4: Závažnost problémů v oblasti kvality dat 58

Obrázek č.5: Graf - důsledky „nekvality“ v datech 59

7.2 Tabulky

Tabulka č.1: Primární zdroje diplomové práce 5

Tabulka č.2: Cíle diplomové práce v otázkách 6

Tabulka č.3: Fáze budování householdingu 22

Tabulka č.4: Příklad matice pravidel 27

Tabulka č.5: Příklad tabulky skórování stáří informací 29

Tabulka č.6: Příklad vah přislučovacích pravidel 29

Tabulka č.7: Příklady duplicitních záznamů 38

Tabulka č.8: Externí zdroje informací 51

Tabulka č.9: Struktura číselníků UIR-ADR 52

Tabulka č.10: Informace obsažené v registru ekonomických subjektů 54

Tabulka č.11: Terminologický slovník 72

8 Seznam zkratek a terminologický slovník

Termíny přejaty z terminologického slovníku katedry informačních technologií (www.cssi.cz) a dále z odborné literatury.

Termín	Význam
Termíny přejaté z terminologického slovníku katedry informačních technologií	
Business Intelligence (BI)	Sada postupů, procesů a technologií, jejímž cílem je účinně a účelně podporovat rozhodovací procesy ve firmě. Představuje komplex aplikací IS/ICT, které podporují analytické a plánovací činnosti podniků a organizací a jsou postaveny na specifických, tzv. OLAP (On-Line Analytical Processing) technologiích a jejich modifikacích.
Customer Relationship Management (CRM)	Řízení vztahů se zákazníky (CRM) zahrnuje pracovníky, podnikové -> procesy a technologii -> IS/ICT s cílem maximalizovat loajalitu zákazníků a v důsledku toho i ziskovost podniku. Je součástí podnikové strategie a jako takové se stává součástí podnikové kultury. Technologicky stále více využívá potenciálu a možností -> Internetu.
Datový sklad (Data Warehouse, DWH)	Integrovaný, subjektově orientovaný, stálý a časově rozlišený souhrn dat, uspořádaný pro podporu řešení analytických úloh.
Datové tržiště (Data Mart, DM)	Podmnožina datového skladu adresující potřeby specifických předmětných oblastí podniku nebo uspokojující potřeby jednotlivých částí organizace.
Enterprise Resource Planning (ERP)	ERP je charakterizován jako typ aplikace, resp. -> aplikačního software v -> informačním systému, který umožňuje řízení a koordinaci všech disponibilních podnikových zdrojů a aktivit s cílem zajištění potřeb

	trhu i vlastního podniku. Aplikace ERP pokrývají všechny základní oblasti podnikového řízení: prodej, nákup, sklady, finanční účetnictví, controlling, majetek, lidské zdroje, práce a mzdy, technickou přípravu výroby, plánování výroby a podporují operativní řízení včetně dílenského řízení výroby. Mezi hlavní vlastnosti ERP patří schopnost automatizovat a integrovat klíčové podnikové procesy a data, zpracovávat je v rámci celé firmy a zpřístupňovat pracovníkům podniku detailní i komplexní informace. ERP aplikace jsou primárně, nikoli výlučně, určeny pracovníkům střední a nižší (operativní) úrovně řízení. Tyto systémy jsou orientovány převážně transakčně, to znamená, že umožňují vytvářet a aktualizovat rozsáhlé -> datové báze - zboží, dodavatelů, zákazníků, pracovníků, majetku, účtů. Podporují procesy operačního charakteru (řízení obchodních případů – nákup, prodej), zpracování obchodních dokumentů (objednávka, kontrakt, faktura, celní deklarace…) a poskytují požadované výstupy, přehledy, statistiky a základní analýzy ve všech vymezených oblastech řízení. Data ERP systémů jsou důležitým zdrojem dat pro aplikace kategorie -> Business Intelligence.
ICT (Informační a komunikační technologie)	-> Hardwarové a -> softwarové prostředky pro sběr, přenos, ukládání, zpracování a distribuci dat. Mezi hardwarové (technické) prostředky patří - servery, stacionární a přenosné personální počítače, tiskárny, komunikační a síťová zařízení (vysílače, směrovače, přepínače) a specializovaná koncová zařízení (myš, tablet, scanner, kamera, PDA, mobilní telefon apod.).

	Mezi softwarové (programové) prostředky patří základní software (operační systém, databázový systém, komunikační systém), -> aplikační software a software pro modelování a -> vývoj informačních systémů.
Metodika	Doporučený souhrn fází, etap, přístupů, zásad, postupů, pravidel, dokumentů, řízení, metod, technik a nástrojů pro tvůrce informačních systémů, který pokrývá celý životní cyklus informačního systému. Určuje, kdo, kdy, co a proč má dělat během vývoje IS.
Ostatní termíny	
Čištění dat	Přístup zvyšování kvality dat formou odhalování a korigování chyb ve vstupních datech IS.
Customer Data Integration (CDI)	Přístup centralizace managementu dat o zákaznících; procesy a řešení pro analýzy a pochopení zákazníka [Zornes]
Datová kvalita (DQ, Data Quality, kvalita dat)	Datová kvalita je definovatelná prostřednictvím šesti základních vlastností. Data jsou kvalitní tehdy, jsou-li přesná, dostupná, relevantní, úplná, srozumitelná a důvěryhodná. [Olson]
Deduplikace	Výběr nejlepšího záznamu, který bude nadále reprezentovat konkrétní subjekt. [Adastra]
Fuzzy logika	Specifický přístup práce s neurčitostí.
Geocoding	Přesná lokalizace jednotlivých adresových bodů v kontextu zeměpisné šířky a délky. [Sacolick] proces přiřazení zeměpisné polohy bodu podle adresy, ulice, města a PSČ [www.camcad.cz]
IS/IT	Informační systém/Informační technologie

Komerční householding (Corporate)	Přístup identifikace skupin obchodních jednotek spojených nebo považovaných za jedno společenství. Může se také jednat o obchodní subjekty, které společně kooperují. [Madnick]
Master Data Management (MDA)	Sada procesů, přístupů, metodologií, nástrojů, technologií a řešení, které pomáhají organizacím jednoznačně identifikovat a integrovat celopodnikově sdílená klíčová data. Slouží ke zvýšení kvality dat, jejich konzistence, dostupnosti a důvěryhodnosti. [Kyjonka]
Osobní householding (Personal)	Přístup identifikace vztahů mezi konkrétními fyzickými osobami v kontextu (nejen) rodiny. [Mancuso]
Parsing	Rozdělení nestrukturovaných vstupních údajů do požadovaných skupin. [Adastra]
Risk management	Soustavná, opakující se sada vzájemně provázaných činností, jejichž cílem je řídit potenciální rizika, tzn. omezit pravděpodobnost jejich výskytu nebo snížit jejich dopady.
Suply Chain Managament (SCM)	řízení dodavatelských řetězců je komponentou IS, která zahrnuje komplex aplikací a technologií pro řízení dodavatelských řetězců
Unifikace	Sloučení záznamů o osobách do jednotlivých skupin, které reprezentují záznamy o jednom klientovi. [Adastra]

Tabulka č.11: Terminologický slovník