



Kvalitní data – – kvalitní agendy

*Kvalita dat a její zajišťování
v agendových systémech veřejné správy*

Připraveno pro konferenci ISSS 2010

Ing. Jiří Vácha

Hradec Králové, 13.4.2010

© Adastra Group

Agenda

- Základní teze datové kvality
- Datová kvalita a základní registry
 - Data a zprovoznění základních registrů
 - Procesy datové kvality
 - Plnění dat pro ZR
- Zajištění kvality dat pro agendy (AIS)
 - Legislativní rámec
 - Měření datové kvality
 - Příklady
- Dotazy





Datová kvalita – základní teze

Základní teze datové kvality

V datech jsou chyby (a vždy tam byly...)

**Míra nekvality
dat představuje
riziko
významného
snížení jejich
použitelnosti**

Zpracování nesprávných, nesmyslných
a nadbytečných dat
- neefektivní zajištění veřejné služby

Nedostatečné fungování interních procesů
- negativní hodnocení veřejné služby

Proč jsou data nekvalitní?

Technika

Mnoho autonomních agendových IS

Různí dodavatelé v různém čase

Různé aplikace a technologie

Lidský faktor

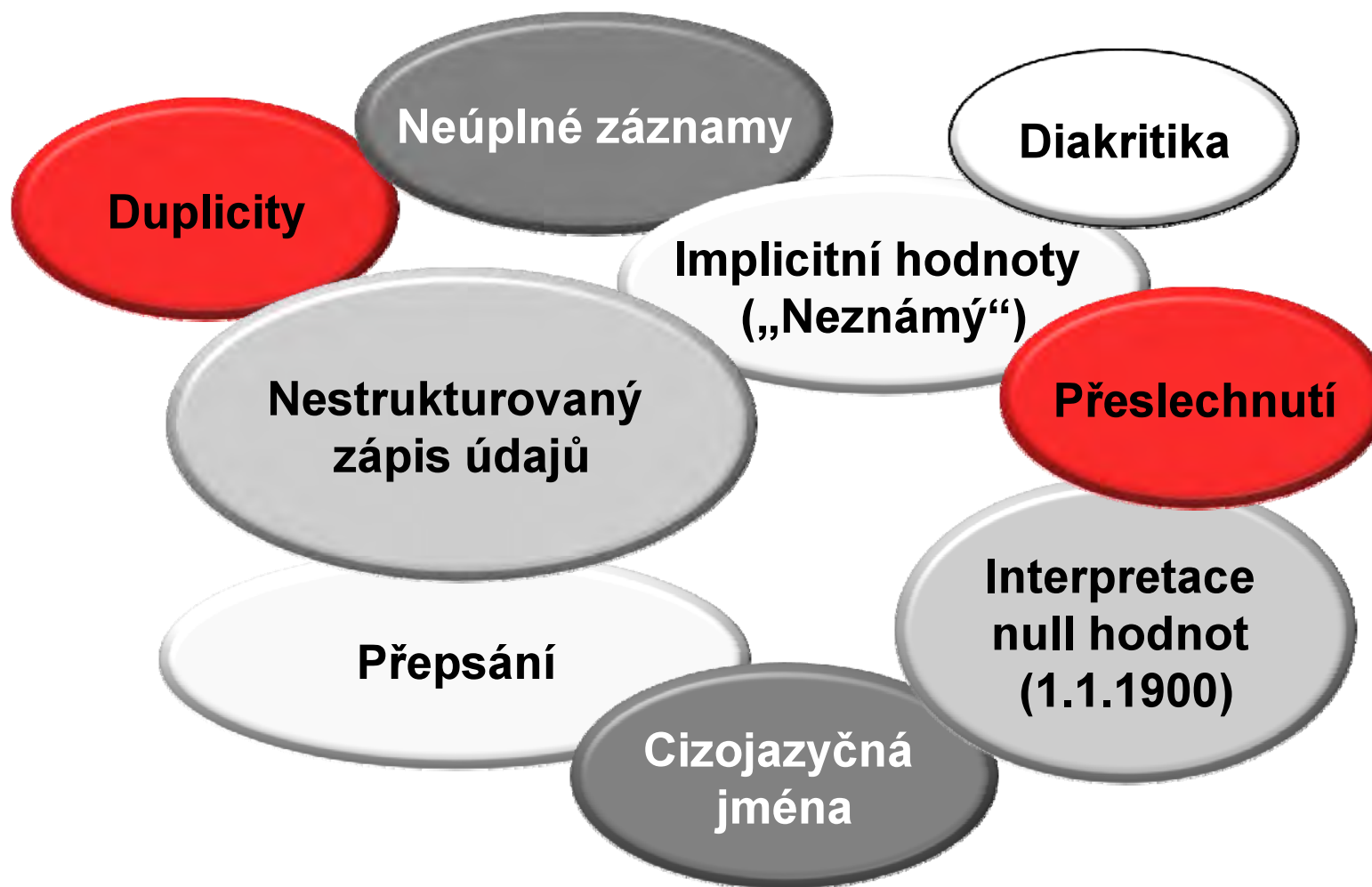
Překlepy

Zápisy do jiných polí

„Lidová tvořivost“

Změna legislativy

Typické problémy s kvalitou dat





Datová kvalita a Základní registry

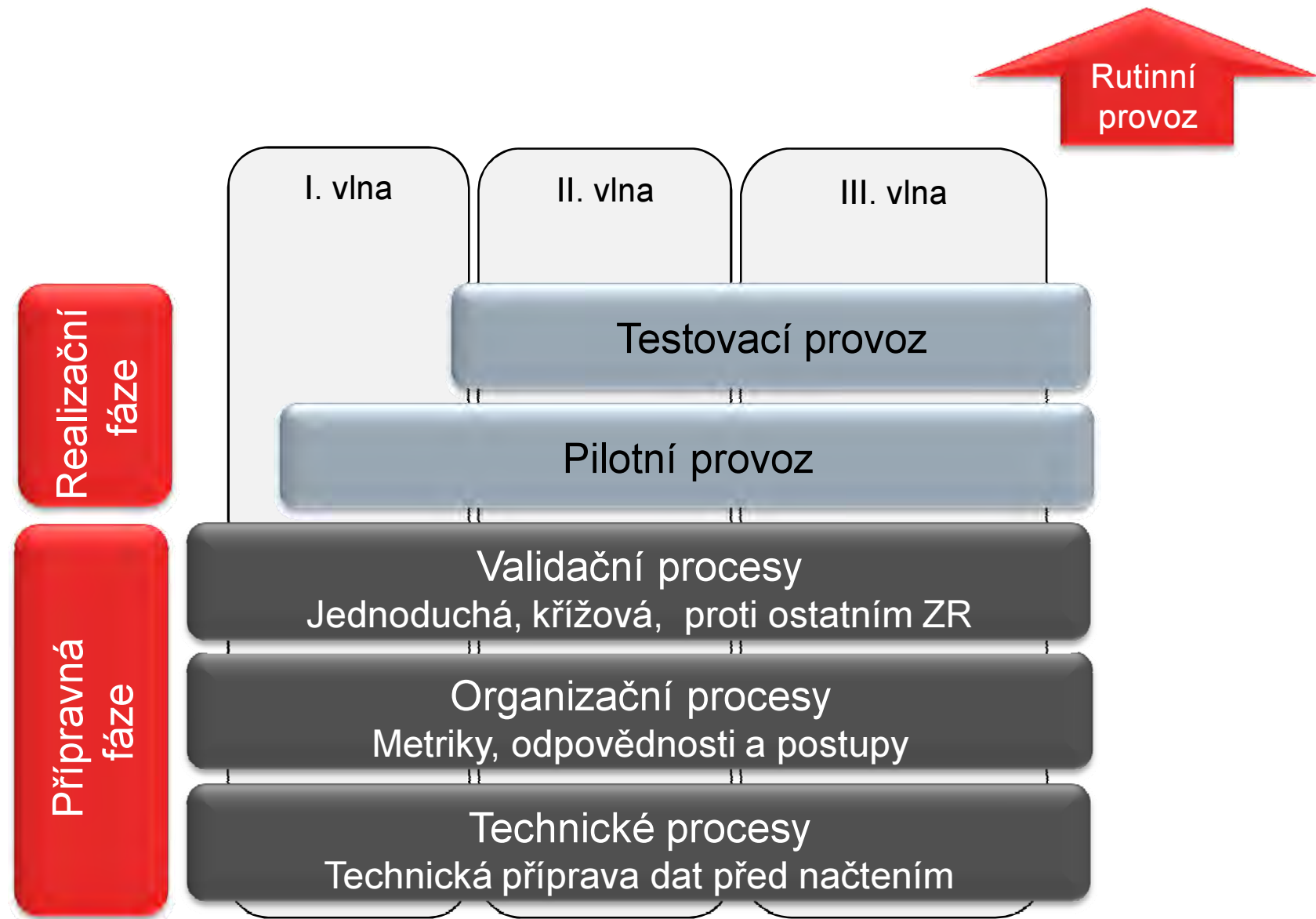
Kde vznikají data?



Zajištění kvality dat v ZR



Data a zprovoznění ZR



Technické procesy datové kvality



**Rozdělení záznamu
do datových položek**

*Rozpoznání obsahu datových
položek (např. název obce, ulice)*

Standardizace

*Převod dat na jednotný formát
(např. malá na velká písmena)*

Obohacení

*Doplnění chybějícího (např. doplnění
PSC, pokud je známo adresní místo)*

Unifikace

*Seskupení podobných
záznamů do jedné skupiny*

Deduplikace

*Nahrazení skupiny podobných
záznamů jedním – nejlepším*

Identifikace

*Nalezení již existujícího,
unifikovaného subjektu v (etalonu)*

Organizační procesy datové kvality



Vlastnictví dat

data (registry) mají jednoznačně určeného vlastníka, který zodpovídá za jejich strukturu, dostupnost, kvalitu a bezpečnost

Správa dat

tzv. data stewardship – průběžné měření datové kvality, analýza příčin nekvality, návrh a realizace nápravných opatření

Data Governance

*souhrnný koncept
Data Governance (adaptovaný
na podmínky výkonu státní správy)*

Validační procesy



Jednoduchá validace

Základní kontrola datových elementů (atributů) a jejich případné obohacení

Křížová validace

*Porovnání dat mezi agendami navzájem
prevence přepisu čistých dat „špinavými“*

Validace proti dalším registrům

*Ověření celkové
konzistence základních registrů*

Plnění dat a provoz registrů



RUTINNÍ PROVOZ
plné spuštění

PILOTNÍ PROVOZ

Úspěšné naplnění jednotlivých vln, průběžně ověřovat funkčnost řešení v omezeném rozsahu

TESTOVACÍ PROVOZ

Ověření celkové provázanosti, integrace a funkčnosti, ověření kvality dat a napojení agendových IS

Možná rizika

Chyby v datech AIS

*Byly tam, jsou
a budou - jde o to, znát výchozí stav*

Opravy dat

*AIS jsou zatíženy svou vlastní
činností a mohou nastat kapacitní problémy*

Práce s agendami

*Každá agenda má jiné poslání a z toho
odvozené potřeby, je třeba získat jejich součinnost*

Napojení AIS

*Úskalí integrace
(různí dodavatelé, různé konvence, různé standardy...)*

Definice pravidel

*Je třeba počítat s velkým
množstvím neshod vůči pravidlům a řešit je citlivě*

Složitost projektu

*Jednotlivé registry jsou samostatnými
projekty, je třeba dobře koordinovat jejich postup*

DQ v kontextu základních registrů

1

**Provedením opatření na
vstupu se vždy zvyšuje
celková úspěšnost**

2

**Kvalitu dat je potřeba
soustavně hodnotit a
provádět korekce**

3

**Kvalita registrů je závislá na
kvalitě agendových
systémů**



Zajištění kvality dat agendových systémů (AIS)

Proč zajišťovat kvalitu dat v agendách?

Efektivita

**Snížení pracnosti zpracování agend
Snížení nákladů na výkon veřejné správy**

Legislativa

**Naplnění požadavků příslušných
legislativních instrumentů**

Veřejné služby

**Poskytování kvalitních služeb občanům
Zvýšení prestiže orgánů veřejné správy**

Hospodárnost

**Účelné hospodaření s prostředky
daňových poplatníků**

Legislativní rámec pro datovou kvalitu



Měření kvality

Přesnost

Přesnost je definována jako shoda uvedené hodnoty se správnou hodnotou (tj. jedná se o existující osobu)

Úplnost

Úplnost je definována jako existence očekávaných hodnot na vstupu (daný subjekt např. musí vždy mít alespoň jednu adresu).

Konzistence

Konzistence znamená vzájemnou konzistenci jednotlivých položek (druhů údajů) v rámci jedné entity i mezi entitami.

Validita

Validita - data splňují kritéria doménové integrity, tzn. že odpovídají definovanému datovému typu, formátu a jsou v definovaném rozsahu přípustných hodnot.

Správné měření datové kvality

Anonymizované

*Může být prováděno
nezávisle na systému i vlastníkovi dat*

Automatizované

*Snižují se náklady
a zvyšuje efektivita procesů*

Srozumitelné

*Zjištěné neshody
jsou srozumitelně prezentovány*

Opakovatelné

*Umožňuje sledovat
trendy a efektivitu nápravných opatření*

Konsolidace dat

Stěžejním prvkem správnosti dat v organizaci je teprve konsolidace dat mezi agendami

Konsolidace = čištění+unifikace

Příklad: *Matka s dcerou ve 2 různých systémech. Dcera dostudovala (titul Ing.), provdala se (Zapletalová má **stejné RČ** jako Štergeníková) a odstěhovala se do Hradce Králové (řádky 3 a 6)*

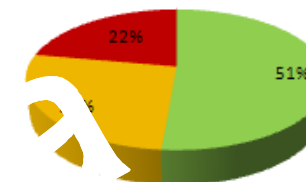
a	b	c	d	e	f	g	h	i	j
1	B	796008xxxx	8.10.1979	F	Jana	Štergeníková	16000	Praha 6	Malá 3
2	A	796008xxxx	8.10.1979	F	JANA	STERGENIKOVA	16200	PRAHA 6	MALA 542/3
3	B	796008xxxx		F	IngJana	Zapletalová	50003	Hrdec Králové	Úzká 2
4	B		8.10.1979	F	ŠTERGENÍKOVÁ	Jana	16000	Praha 6	Malá 3/542
5	A	495722yyyy	22.07.1949	F	JANA	STERGENIKOVA	16000	PRHA	MALA 3
6	B		8.10.1979	F	Jana	Zapltalová	50001	Hra Král	Úzká 2

Reporty datové kvality (příklad 1.)

Sum of hodnota	typ subjektu	entita	atribut	kategorie	Ověřené	Opravitelné	Neověřené	Ne vyplněné	Grand Total
PO	Právnícká osoba	IČ			15925		4530		20455
			Název firmy		10501	5424	4523	7	20455
			Právnícká osoba		10501	5424	4530		20455
FO	Fyzická osoba	Jméno			43796		98	5	43899
			Příjmení		43643	48	208		43899
			Titul		5391	194	52	38262	43899
			RČ		43892		7		43899
			Fyzická osoba		43322	235	342		43899
	Adresa provozovny	Obec			2367	5	1		2373
			Část obce		699	1633	27	14	2373
			Ulice		1998	334	38	3	2373
			Číslo popisné/orientační		2332		37		2373
			PSČ		2353	15	5		2373
Fyzická osoba podnikající	IČ			16671				16671	
		Název firmy		428	16243		1291	17970	
		Fyzická osoba podnikající		428	16243			17970	
Společné	Adresa	Obec			11253	1526	175		64354
			Část obce		2862	322	3202	273	64354
			Ulice		517	9812	3012	463	64354
			Číslo popisné/orientační		601	1	3047	455	64354
			PSČ		622	1454	582	31	64354
			Adresa		8801	52051	3502		64354

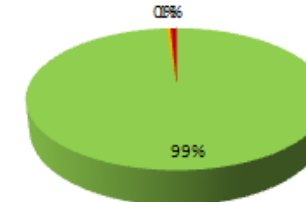
Právnícká osoba

Ověřené Opravitelné Neověřené



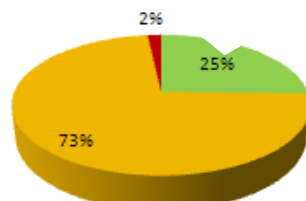
Fyzická osoba

Ověřené Opravitelné Neověřené



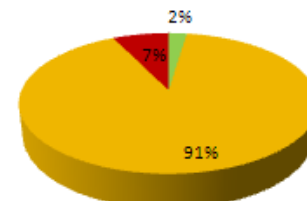
Adresa provozovny

Ověřené Opravitelné Neověřené



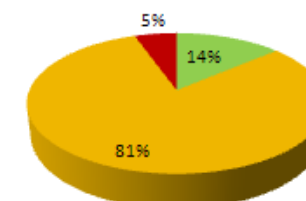
Fyzická osoba podnikající

Ověřené Opravitelné Neověřené



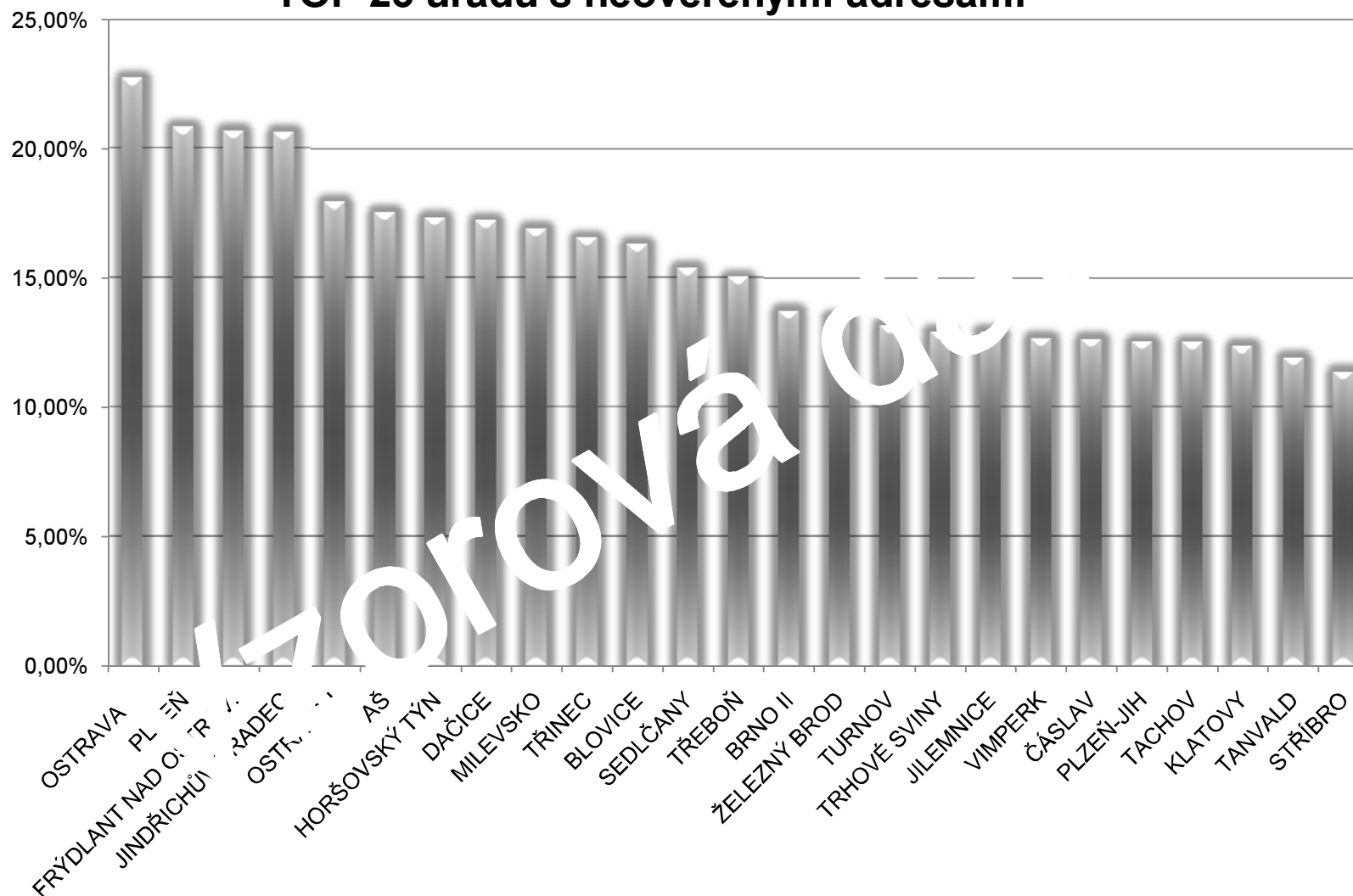
Adresa

Ověřené Opravitelné Neověřené



Reporty datové kvality (příklad 2.)

TOP 25 úřadů s neověřenými adresami



DQ agendových systémů

1

Agenda je primární zdroj informací i v ZR

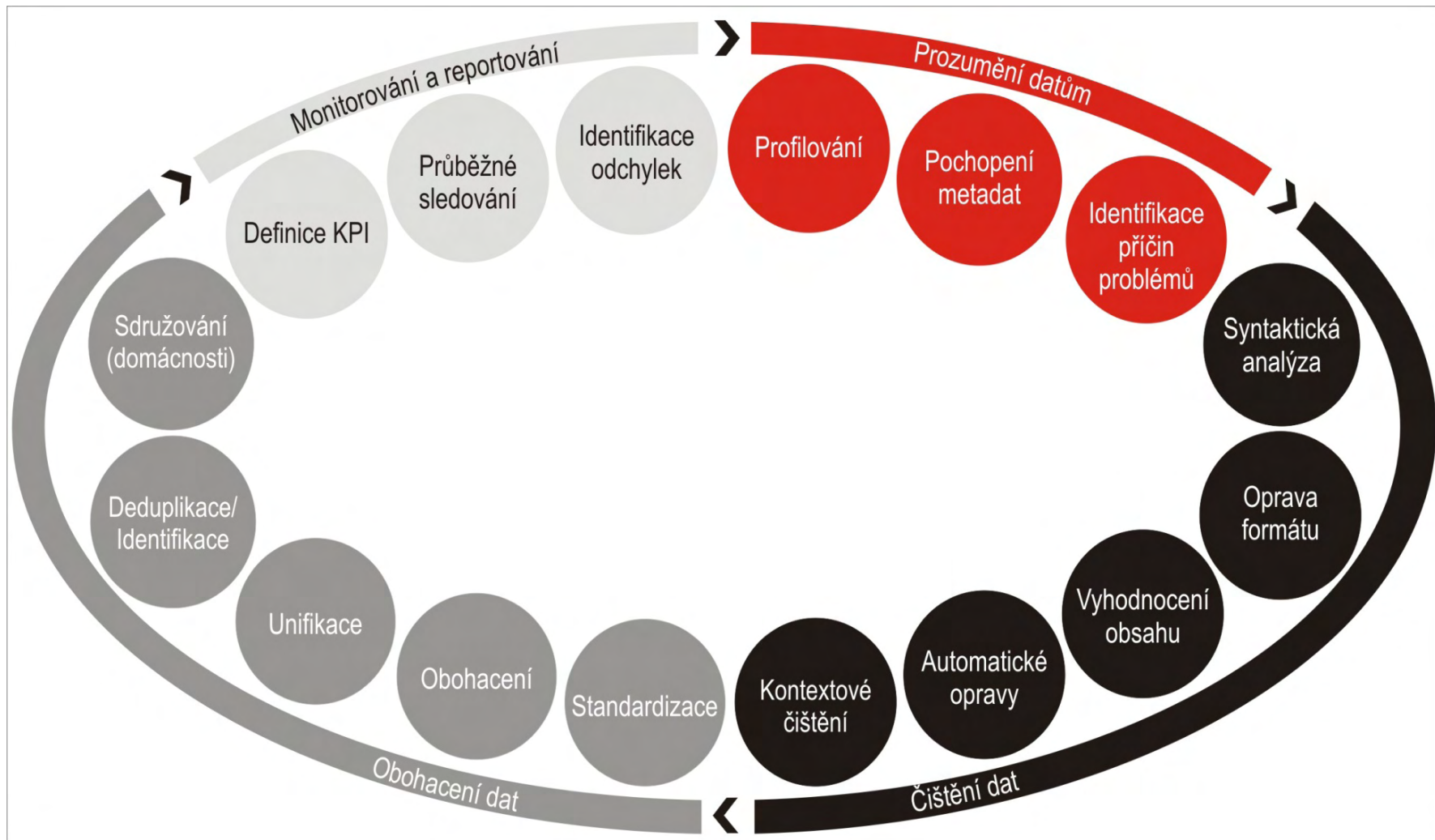
2

Datová kvalita není cíl, ale prostředek k správnému provozu agendy

3

Nejde však jen o samotná data, ale i o lidi a procesy s agendou související

Cyklus řízení datové kvality





ADASTRA CZ
Nile House
Karolinská 654/2
186 00 Praha 8

Tel.: +420 271 733 303
info@adastragrp.com
www.adastra.cz



www.adastragrp.com

Děkujeme za pozornost

ADASTRA GROUP North America
8500 Leslie St.
Markham, Ontario, L3T 7M8
Canada
Tel: +1 905 881 7946
info@adastragrp.com

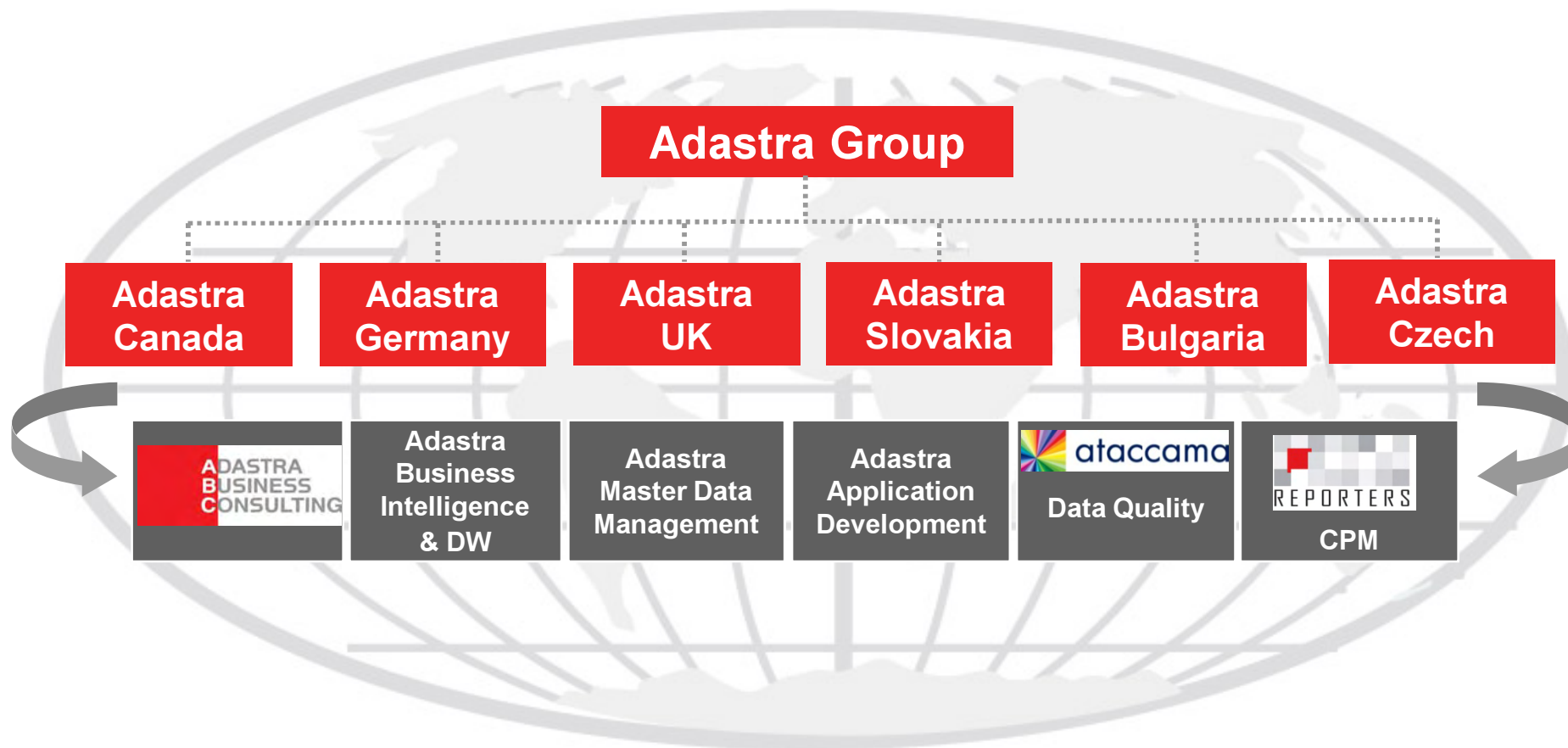
ADASTRA GROUP Europe
Karolinska 654/2
186 00 Praha 8
Czech Republic
Tel.: +420 271 733 303
info@adastragrp.com





O společnosti Adastra

Adastra Group – struktura společnosti



Adastra v číslech

1 cíl – pomáháme firmám přeměnit jejich data v konkrétní obchodní příležitost a reálnou konkurenční výhodu

14 let na trhu

43,6 mil EUR – obrat ve fiskálním roce 2008 (98% služby)

7 kanceláří – vedení společnosti v Kanadě (Toronto) a České republice (Praha), další kanceláře: Německo (Frankfurt), Velká Británie (Londýn), ČR (Ostrava), Slovensko (Bratislava), Bulharsko (Sofie)

600+ zaměstnanců po celém světě

220+ úspěšných projektů v **16** zemích

75% výnosů od stávajících zákazníků

16 významných mezinárodních ocenění

Vybraní zákazníci společnosti Adastra

- Česká správa sociálního zabezpečení
- Generální ředitelství cel
- Ministerstvo práce a sociálních věcí ČR
- Fakultní nemocnice Královské Vinohrady
- Nemocnice Kutná Hora
- Nemocnice Milosrdných Bratří, Brno
- Bawag Bank CZ
- Feron
- Global Payments Europe
- Mediatel
- Raiffeisenbank/eBanka
- Tipsport
- Vodafone Czech Republic
- Canadian Institute for Health Information
- Loblaw Companies Limited, Ontario, Kanada
- Ministerstvo zdravotnictví Kanada
- Ontario Energy Board
- Centennial College, Ontario, Kanada
- Deutsche Bank New York
- Raiffeisen Bank Aval, Kyjev
- KMB-Bank, Moskva
- Raiffeisen International, Vídeň



Ocenění a certifikáty

- Adastra patří mezi TOP 10 ICT firem roku 2008
- Adastra je mezi TOP 10 systémovými integrátory roku 2008
- Adastra prošla recertifikačním auditem (Moody International) a obdržela certifikát ISO 9001 do roku 2010
- Umístění v žebříčku Technology Fast 50 CE a Technology Fast 500 EMEA, sestavovaném společností Deloitte
- Jan Mrázek získal ocenění „Podnikatel roku 2008“ v kategorii Technology pro region Ontario v Kanadě a J. Červinka je Technologickým podnikatelem roku 2006 v ČR (Ernst & Young)
- Adastra obhájila certifikace „Microsoft Gold Partner“ a „Oracle Certified Partner“
- Adastra získala status Platinum mezi „50 Nejlépe řízenými společnostmi v Kanadě“
- Adastra Kanada se umístila v žebříčcích Top 25 Canadian IT Professional Services Companies a Top 250 Canadian Technology Companies (Branham Group)



