

 **ADASTRA**



# Základní registry

*Kvalita dat a jejich čištění  
v základních registrech veřejné správy*

**Připraveno pro konferenci ISSS**

Ing. Jiří Vácha

Hradec Králové, 6.4.2009

© Adastra Group

# Agenda

- Základní teze datové kvality
- Řízení datové kvality
- Datová kvalita a základní registry
  - Data a zprovoznění základních registrů
  - Procesy datové kvality
  - Plnění dat
  - Rizika
- Doporučení
- Dotazy





# Datová kvalita – základní teze

# Základní teze datové kvality

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

**Velká 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

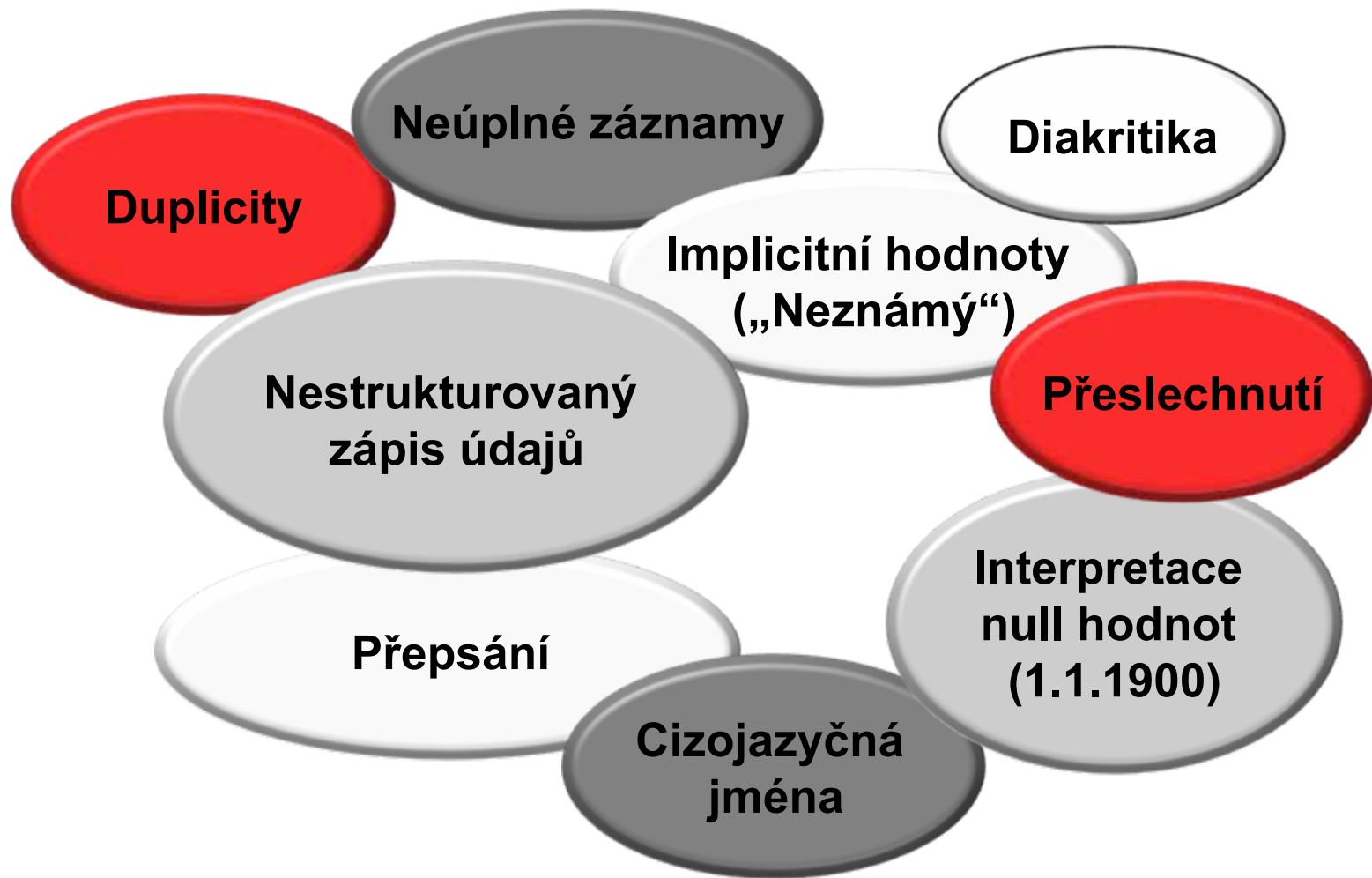
Překlepy

Zápisy do jiných polí

„Lidová tvořivost“

Změna legislativy

# Typické problémy s kvalitou dat



# Příklad: zápis titulu Ing.

Ing.,.	Engenheiro	Ing.	Ing. +	Dipl.-Ing.
ing	iIng.	Ing.,	Ing. +	Dipl.ing.
ing.	lIng.	Ing.-	Ing.	Dipl.Ing.
Ing.	img	Ing..	ing	Dipl.Ing..
dipl. ing	img.	ING	Ing. Ing.	DIPL.ING.
dipl. ing.	in.	INC	Ing.l	Dypl.Inz.
dipl. Ing.	Inf	ín	Ing.1	iNG.
Dipl. ing.	infg	ing,	Inhg	Ing
Dipl. Ing	.Ing	ing.	Inng.	Ing,
Dipl. Ing.	ing	Ing,	Inž.	Ing.,
Dipl. ING.	ing,	Ing.	iong	Ing-
dipl ing	ing-	Ing. *	Ing .	Ing-.
dipl.ing.	ing.	Ing. .	ing	Ing
dipl.Ing.	ing.,	Ing	Ing.	Ing.



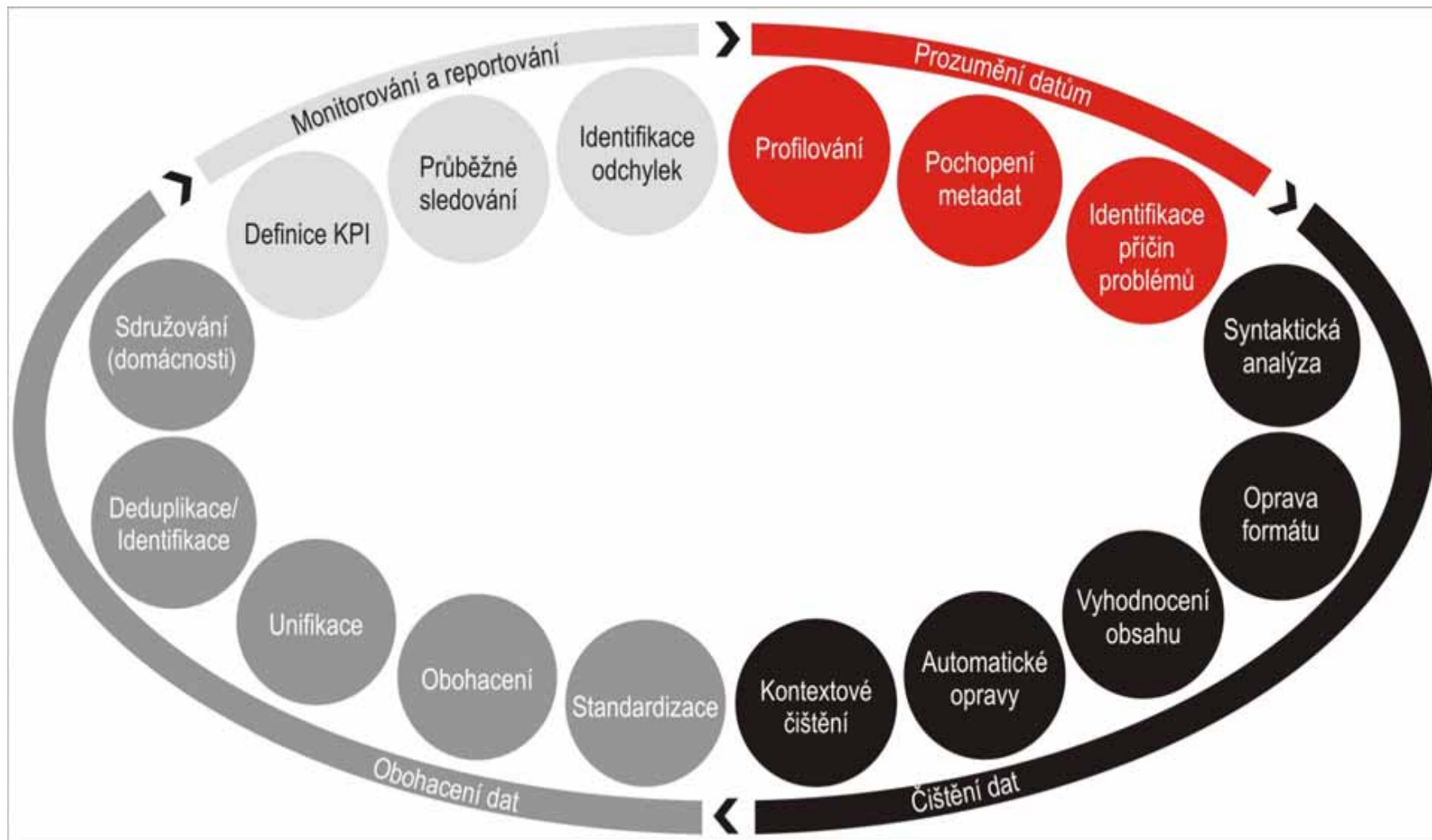
# Řízení datové kvality



# Data Governance



# Cyklus řízení datové kvality



# Měření datové 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 znamená, že data splňují kritéria doménové integrity, tzn. že odpovídají definovanému datovému typu a formátu. Pro test validity mohou být definována také další formální pravidla, například rozsah přípustných hodnot pro IČO, algoritmy pro kontrolní číslice, RČ obsahující na začátku kódované datum narození, atd.

# Správné měření datové kvality

## Anonymizované

Může být prováděno nezávisle na systému i vlastníkov 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 registrů je konsolidace

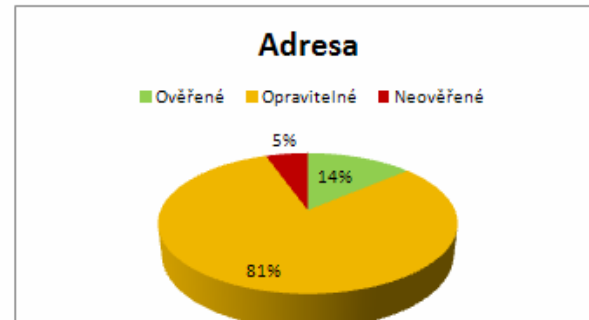
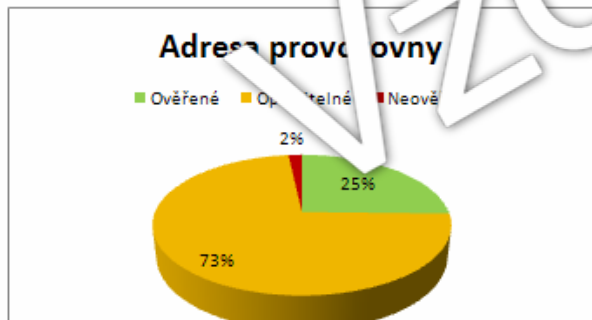
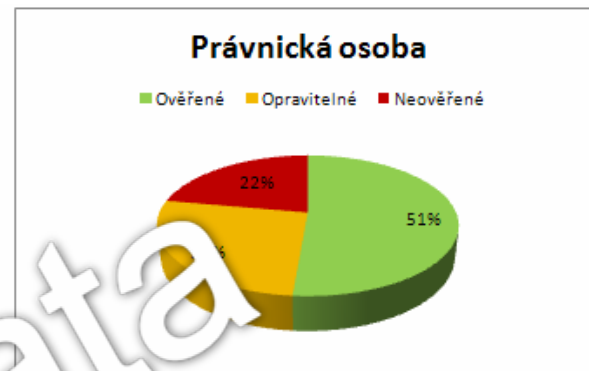
Konsolidace = čištění+unifikace

**Příklad:** Matka s dcerou ve 2 různých systémech. Dcera dostudovala (titul Ing.), provdala se (Zapletalová má **stejně 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	Zapletalová	50001	Hra Král	Úzká 2

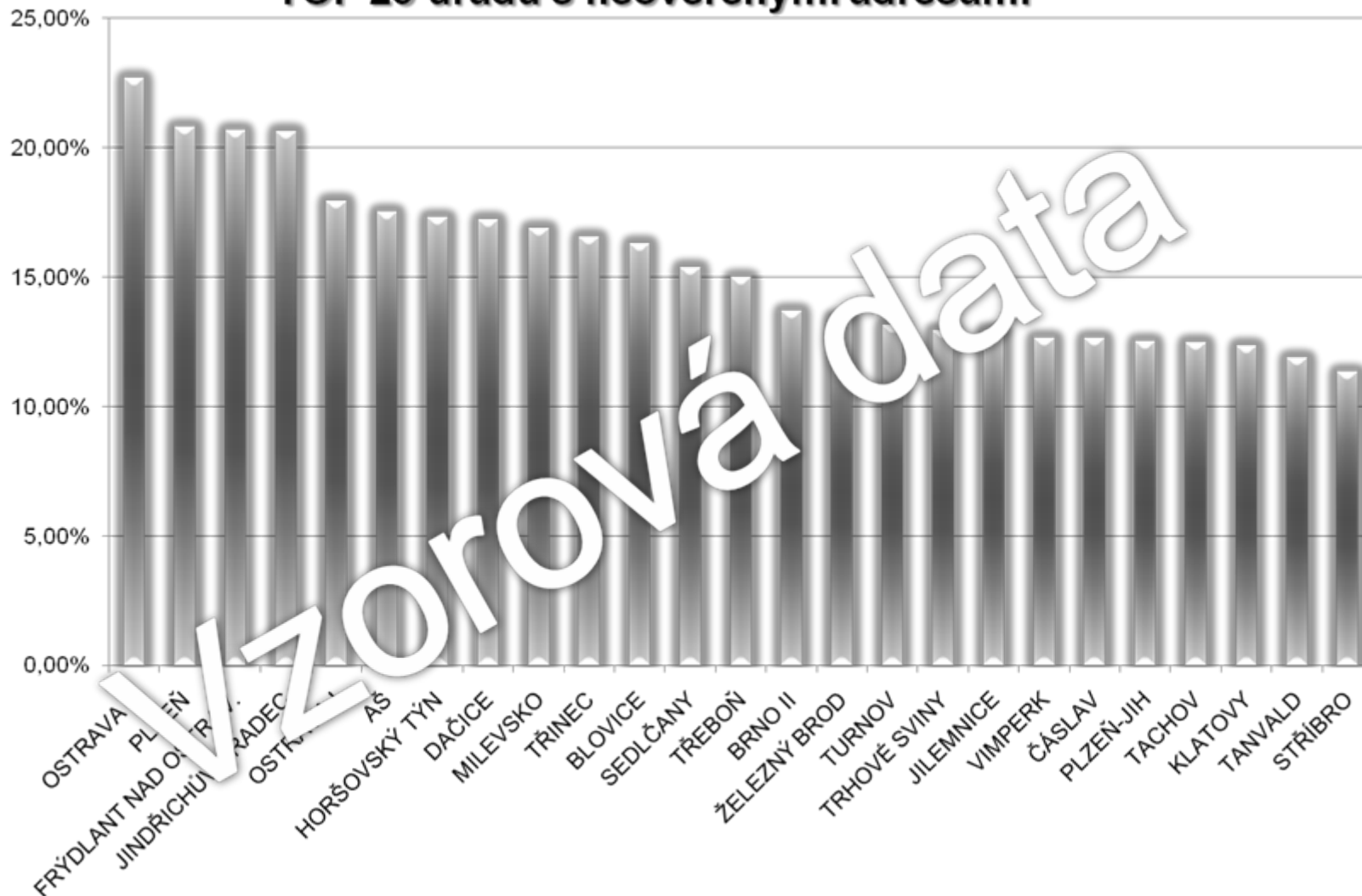
# Reporty datové kvality (příklad 1.)

Sum of hodnota	typ subjektu	entita	atribut	kategorie	Ověřené	Opravitelné	Neověřené	NewpIněné	Grand Total
PO	Právnícká osoba	Právnícká osoba	IČ		15925		4530		20455
			Název firmy		10501	5424	4523	7	20455
			<b>Právnícká osoba</b>		<b>10501</b>	<b>5424</b>	<b>4530</b>		<b>20455</b>
FO	Fyzická osoba	Fyzická osoba	Jméno		43796		98	5	43899
			Příjmení		43643	48	208		43899
			Titul		5391	194	52	38262	43899
			RČ		43892		7		43899
			<b>Fyzická osoba</b>		<b>43322</b>	<b>235</b>	<b>342</b>		<b>43899</b>
	Adresa provozovny	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
<b>Adresa provozovny</b>		<b>602</b>	<b>1730</b>	<b>41</b>		<b>1773</b>			
Fyzická osoba podnikající	Fyzická osoba podnikající	IČ		16671				17970	
		Název firmy		428	16243		1291	17970	
		<b>Fyzická osoba podnikající</b>		<b>428</b>	<b>16243</b>			<b>17970</b>	
Společné	Adresa	Adresa	Obec		11253	526	175		64354
			Část obce		2863	321	3202	273	64354
			Ulice		5177	9812	3012	463	64354
			Číslo popisné/orientační		6011	1	3047	455	64354
			PSČ		6221	1454	582	31	64354
			<b>Adresa</b>		<b>8801</b>	<b>52051</b>	<b>3502</b>		<b>64354</b>



# Reporty datové kvality (příklad 2.)

## TOP 25 úřadů s neověřenými adresami





# Datová kvalita a Základní registry



# DQ v kontextu základních registrů

Provedením opatření na vstupu se vždy zvyšuje úspěšnost

Kvalitu dat je potřeba soustavně hodnotit a provádět korekce

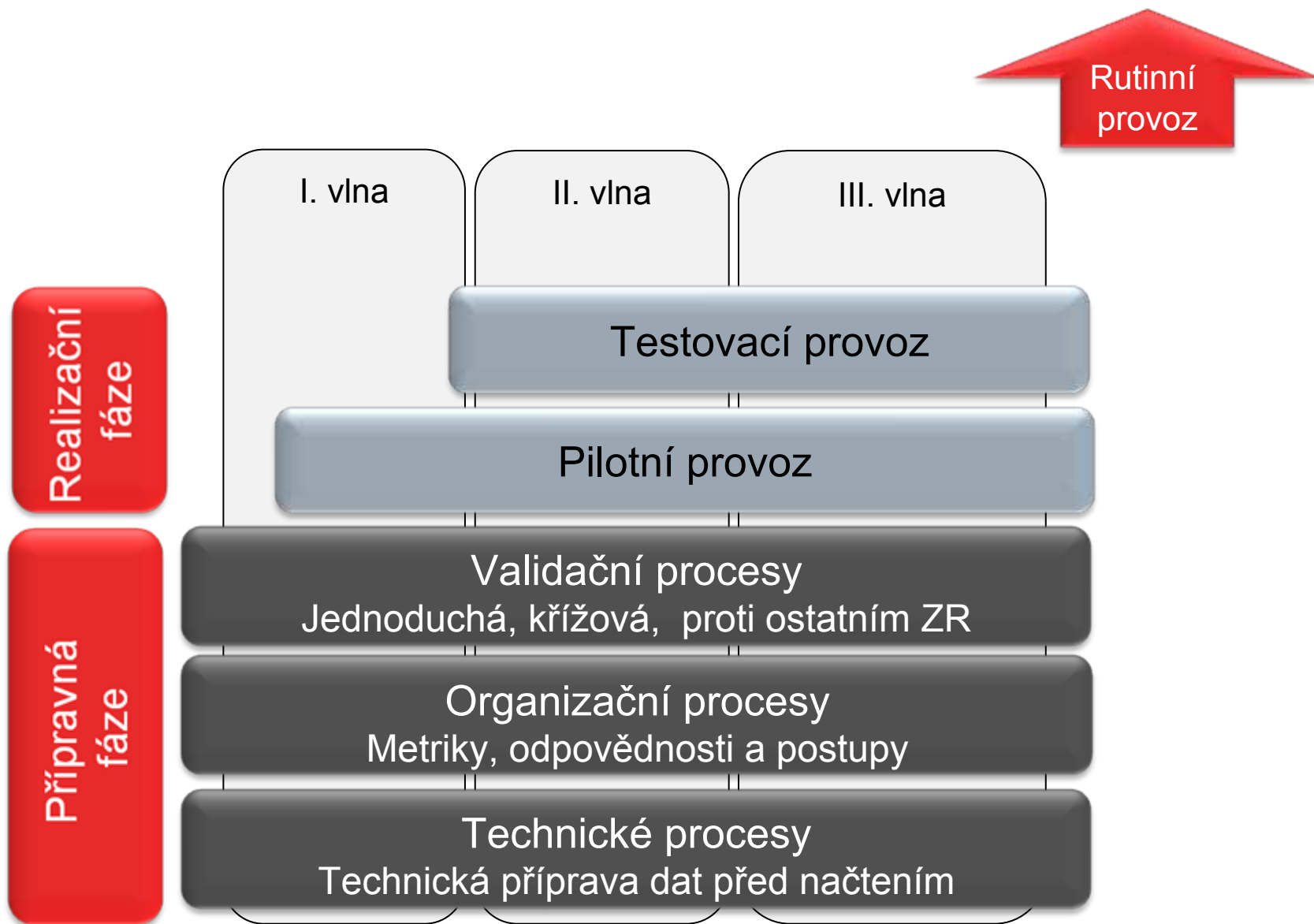
89

1/3

3

Kvalita registrů je závislá na kvalitě agendových systémů

# 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 PSČ, 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 (jednotlivé 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

**TESTOVACÍ PROVOZ**  
ověření celkové provázanosti, integrace a funkčnosti, ověření  
kvality dat a napojení agendových IS

**PILOTNÍ PROVOZ**  
úspěšné naplnění jednotlivých vln průběžně ověřovat  
funkčnost řešení v omezeném rozsahu

**ROZDĚLENÍ NA VLNY**  
z důvodu značné komplexnosti a počtu agendových systémů  
je účelné provádět plnění registrů v postupných vlnách

# Možná rizika

## Chyby v datech agendových IS

byly tam, jsou a budou, jde o to, znát výchozí stav

## Opravy dat

agendové IS 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 agendových IS

úskalí integrace (různí dodavatelé, různé konvence, různé standardy...)

## Jednoznačná 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

# Doporučení

**Datová kvalita není cíl, ale prostředek  
k úspěšnému řešení základních registrů**

**Nejde jen o systémy a data, ale o lidi a procesy**

**Základní registry jsou strategickým projektem**

**Datová kvalita je velmi specifická disciplína  
a vyžaduje specifické nástroje a zkušenosti**



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

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



[www.adastragrp.com](http://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