



ataccama

# Kvalitní data ve státní správě jako předpoklad efektivní služby

## Konference ISSS 2008

Michal Klaus, Generální ředitel,  
Ataccama Software s.r.o.

## Je „velká kráva“, uvádí o ženě telefonní seznam

27. února 2008 8:40

velikost textu: □ ■ □

Většina lidí má ve Zlatých stránkách Mediatelu a Telefóniky O2 vedle svého jména uvedeno jen číslo, případně adresu. Jindřiška Bielová z Bochova na Karlovarsku však má smůlu. Vedle jejího jména je ještě tučně vytištěno "velká kráva".



foto: Lukáš Kuběna, MAFA

Jindřiška Bielová tvrdí, že se jí motí někdo z blízkého okolí.

Tuto ženu nejnovější vydání Zlatých stránek pořádně rozčlilo. "Nedala jsem souhlas se zveřejněním mobilu, takže jsem si telefonní seznam neprohližela. Až mě na to upozornili známí," podivovala se žena, která pracuje jako řidička z povolání.

Tuší, že se jí hanlivým výrazem mstí někdo z jejího nejbližšího okolí, ale konkrétní být nechce. Stačilo prý jen znát telefonní číslo s přístupovým heslem a přes [internet](#) požádat o zveřejnění v telefonním seznamu. "Někdo moje heslo prolomil," je si jistá Jindřiška Bielová.

Podle Filipa Hrubého z tiskového oddělení Vodafonu je vše v pořádku. "Neevidujeme žádný případ, kdy by v rámci webové samoobsluhy neoprávněná osoba zneužila zákaznický účet," sdělili mluvčí.

Bez viny se cítí i vydavatel Zlatých stránek, firma Mediatel. "Podle [zákona](#) i [smlouvy](#) tiskneme data, která dostaneme od operátora," vysvětlila mluvčí Lucie Kučerová.

# O společnosti Ataccama

- Ataccama vyvíjí a dodává specializované technologie pro
  - **Data Quality**
  - **Master Data Management**
- Mezinárodní působnost
  - Aktuálně expanze na trhy **střední a východní Evropy**
  - **Německo, Rakousko**
  - **Canada, USA**
- Nejvyšší **znalostní báze pro ČR, SR**
- Více než **20 zákazníků**
  - **Banky**
  - **Pojišťovny**
  - **Telekomunikace**
  - **Státní správa**
- Navazujeme na úspěšný produkt Purity

# Vize – jak pomáháme našim zákazníkům

## Lepší data

- Řízení kvality dat
- Data Governance
- Master Data Management

## Bez komplikací

- Bohatá funkcionálnita v základu
- Rychlé nasazení
- Bezproblémový provoz
- Nejmodernější technologie

# Příčiny nekvality

- Větší organizace využívají typicky více různých primárních systémů a nadstavbových aplikací s různými pravidly, procesy, kontrolami, omezeními a způsoby uložení
- Jednotlivé primární systémy (tzv. datová sila) obsahují roztroušené/nekonsolidované údaje o osobách a navazujících informacích (adresách, kontaktních údajích apod.)
- Neexistuje jednotná datová integrační platforma pro správu klíčových údajů
- Neexistují efektivní procesy eliminace pořizování nekvalitních dat
- Neexistují efektivní procesy údržby a aktualizace klíčových údajů
- Nejsou dostupné úplné, kvalitní a aktuální etalony (s výjimkou adres – UIR ADR)
- Existují omezení pro změny/opravy údajů (legislativa, náklady na komunikaci se zákazníkem)

# Znalostní báze dodávaná jako součást produktu v ČR, SK

## Typ osoby

- Fyzické osoby
- Právnícké osoby

## Rodné číslo – primární ID pro residential zákazníky

- vedle závazného formátu se ověřuje i vazba na další atributy (datum narození, pohlaví)

## Datum narození, Pohlaví

- kontrolují se jak nezávisle, tak v kontextu

## Jméno, příjmení

- parsing, obsáhlá báze replacementů, extrakce titulů

## Tituly

- akademické/společenské

## Občanský průkaz, příp. další doklady

- standardizace/formát

## Kontakty, Telefonní čísla

- E-mail, URL

## IČ – primární ID pro firemní zákazníky

- vedle závazného formátu se provádí kontextové vyhledávání v etalonu (nevalidní nebo chybějící IČ)

## Název firmy

- Normalizace právních forem, porovnání s etalonem

## Číslo bankovního účtu

- validace na základě závazné vyhlášky, předčíslí, č. účtu

## Čísla karet

- validace na bázi tzv. Luhnova vzorce

## Adresní informace

- Parsing, identifikace komponent
- Rozsáhlá báze replacementů

## Identifikace automobilů

# Co je to kvalitní adresa?

- Doručitelná
- Čistá a identifikovatelná
  - Identifikace adresy spočívá v určení unikátního „adresního bodu“
    - V ČR a SR = domovního vchodu
- Ve vztahu (v kontextu)
  - Existuje vazba na subjekt (občan, agenda, účet, nemovitost apod.)
- Aktuální (platná)
  - Objektivně
    - Vzhledem k etalonu
  - Subjektivně
    - Vzhledem k subjektu

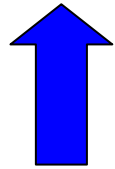
# Proč je kvalita adres důležitá?

- Konsolidace údajů o firmách a organizacích ve městech a krajích
- Konsolidace informací o občanech tamtéž
- Sběr dat a jejich konsolidace na úrovních krajů a centrálních orgánů
- Vytvoření a provozu jednotných centrálních registrů – prvotní konsolidace bude velmi náročná na ruční kontrolu a automatizované „předčištění“ může znamenat významné úspory.
- Identifikaci daného subjektu v registru i při neúplném/chybném vstupu
- Veškeré aplikace portálového typu a elektronických podání, kdy uživatel vkládá svoji identifikaci, adresu, identifikaci firmy apod.
- Vyčištění a deduplikace historických dat v rámci migrace na nové systémy ministerstev
- Obecně kdekoliv, kde je potřeba jednorázově nebo pravidelně konsolidovat více zdrojů dat a nalézt v nich záznamy týkající se jedné entity

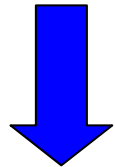


# Vztahy identifikačních úloh

- Identifikace osoby



- Identifikace adresy



- Identifikace domácnosti

- Identifikace vozidla

- Obecná identifikace

# Pohled zvenčí

- „Více než 10 % údajů o adresách je do informačních systémů zadáváno chybně“
  - EDI Group [1]
- „Podíl nevalidních adres v rámci jednoho systému dosahuje až jedné čtvrtiny z jejich celkového počtu“
  - Innovative Systems, Inc. [2]

1. Schwartz, K.: The Electronic Antidote to Slow Business – With Bucks, EDI Can Add Zip to Routine Commerce; Reengineering Resource Center; Enterprise Reengineering; 1996
2. Schardong, C.: 7 Business Benefits of Postal Coding – And What to Look for in Choosing the Right Solution; White Paper; Innovative Systems, Inc.; 2005



# Pohled zevnitř – Rožnov pod Radhoštěm

ROŽNOV	ROŽNOV P/R.	ROŽNOV POD RADHOŠTĚN
ROŽNOV OD RADH.	ROŽNOV P/RADH.	ROŽNOV POD RADHOŠTLĚM
ROŽNOV OD RADHOŠTĚM	ROŽNOV P/RADH.	ROŽNOV POD RADHOŠTM
ROŽNOV P OD RADHOŠTĚ	ROŽNOV P/RADHO	ROŽNOV POD RADHOŠTTĚM
ROŽNOV P R.	ROŽNOV P/RADHOŠTĚM	ROŽNOV POD RADHOTĚM
ROŽNOV P RADH.	ROŽNOV P:R.	ROŽNOV POD RADHOTŠĚM
ROŽNOV P RADHOŠTĚM	ROŽNOV PD RADH.	ROŽNOV POD RADNOŠTĚM
ROŽNOV P. R	ROŽNOV PDO RADHOŠTĚM	ROŽNOV POD RADOŠTĚM
ROŽNOV P. R-	ROŽNOV PDOD RADHOŠTĚM	ROŽNOV POD RADSHOŠTĚM
ROŽNOV P. R,	ROŽNOV PO RADHOŠTĚM	ROŽNOV POD RAHOŠTĚM
ROŽNOV P. R.	ROŽNOV POD	ROŽNOV POD RDAHOŠTĚM
ROŽNOV P.- R.	ROŽNOV POD ADHOŠTĚM	ROŽNOV POD RDHOŠTĚM
ROŽNOV P. R..	ROŽNOV POD EADHOŠTĚM	ROŽNOV POD. RADH.
ROŽNOV P. RAD.	ROŽNOV POD R.	ROŽNOV POD. RADHOŠTĚM
ROŽNOV P. RADH	ROŽNOV POD RA	ROŽNOV POD. RODAHOŠTĚM
ROŽNOV P. RADH-	ROŽNOV POD RAD	ROŽNOV POD.RADH.
ROŽNOV P. RADH.	ROŽNOV POD RAD.	ROŽNOV POD.RADHOŠTĚM
ROŽNOV P. RADHOS TĚM	ROŽNOV POD RADH	ROŽNOV PODRADHOŠTĚM
ROŽNOV P. RADHOŠTEM	ROŽNOV POD RADH,	ROŽNOV POPD RADH.
ROŽNOV P. RADHOŠTĚM	ROŽNOV POD RADH.	ROŽNOV POSD RADHOŠTĚM
ROŽNOV P. RADHOŠTĚM	ROŽNOV POD RADH.,	ROŽNOV PPOD RADH.
ROŽNOV P. RADHOŠTĚM0	ROŽNOV POD RADHO3T2M	ROŽNOV PPOD RADHOŠTĚM
ROŽNOV P.. R.	ROŽNOV POD RADHOŠTĚM	ROŽNOV PR.
ROŽNOV P.R	ROŽNOV POD RADHOS TĚM	ROŽNOV P-R.
ROŽNOV P.R-	ROŽNOV POD RADHOŠ TĚM	ROŽNOV.P.RADH.
ROŽNOV P.-R-	ROŽNOV POD RADHOŠEM	ROŽNOV/RADHOŠT
ROŽNOV P.R,	ROŽNOV POD RADHOŠTĚ	ROŽNOV/RADHOŠTĚM
ROŽNOV P.R.	ROŽNOV POD RADHOŠTEM	ROŽNOVA POD RADHOŠTĚM
ROŽNOV P.R.;	ROŽNOV POD RADHOŠTĚM	ROŽNOVB POD RADHOŠTĚM
ROŽNOV P.RAD	ROŽNOV POD RADHOŠTĚM	ROŽNOVP OD RADH.
ROŽNOV P.RADH	ROŽNOV POD RADHOŠTĚM	ROŽNOVP OD RADH.
ROŽNOV P.RADH.	ROŽNOV POD RADHOŠTĚM 1	ROŽNOVP. R.
ROŽNOV P.RADH..	ROŽNOV POD RADHOŠTĚM 3	ROŽNOVP.R.
ROŽNOV P.RADHO	ROŽNOV POD RADHOŠTĚM OKR. VSETÍN	ROŽNOVPOD RADH
ROŽNOV P.RADHOŠTĚM	ROŽNOV POD RADHOŠTĚM, CZ	ROŽNOVPOD RADH.
ROŽNOV P.ROŽNOV	ROŽNOV POD RADHOŠTĚM;	ROŽNOVPOD RADHOŠTĚM
ROŽNOV P/R	ROŽNOV POD RADHOŠTĚM0	ROŽNOVPOD RADHOŠTĚM

# Pohled zevnitř - Slovensko

&SLOVENSKO	SLOIVENSKO	SLOVENAKO	SLOVENSKEO	SLOVENSKU	SLOVNSKO
SLOVENSKO	SLOIVENSKO	SLOVENASKO	SLOVENSKI	SLOVENSKÝ	SLOVONSKO
S LOVENSKO	SLOLVENSKO	SLOVENCKO	SLOVENSKÍ	SLOVENSLKO	SLOVRENAKO
S.LOVENSKO	SLOVENSKO	SLOVENDSKO	SLOVENSKIO	SLOVENSLO	SLOVRENSKO
SALOVENSKO	SLONESKO	SLOVENENSKO	SLOVENSKJO	SLOVENSLP	SLOVRNSKO
SDLOVENSKO	SLOPVENSKO	SLOVENESKO	SLOVENSKKO	SLOVENSMOLENICE	SLOVSENSKO
SKLOVENSKO	SLOV	SLOVENKO	SLOVENSKKOO	SLOVENSO	SLOVSNSKO
SKOVENSKO	SLOV ENKO	SLOVENKOS	SLOVENSKLO	SLOVENSSKO	SLOVVENSKO
SKSLOVENSKO	SLOV ENSKO	SLOVENKSKO	SLOVENSKMO	SLOVENSWKO	SLOVWENSKO
SL	SLOV REPUBLIKA	SLOVENKSO	SLOVENSKNO	SLOVENWSKO	SLOVWNSKO
SL.	SLOV,.	SLOVENKÝ	SLOVENSKO--	SLOVENWSKOÍ	SLÖOVENSKO
SLAVENSKO	SLOV.	SLOVENNSKO	SLOVENSKO	SLOVERNSKO	SLPOVENSKO
SLEVENSKO	SLOV. REP	SLOVENSAKO	SLOVENSKÓ	SLOVESKO	SLSOV
SLOVENSKO	SLOV. REP.	SLOVENSEKO	SLOVENŠKO	SLOVESNKÁ REPUB	SLSOVENSKO
SLKOVENSKO	SLOV. REPUBLIKA	SLOVENSJO	SLOVEŇSKO	SLOVESNKO	SLVENSKO
SLL	SLOV.REP.	SLOVENSJP	SLÓVENSKO	SLOVESNKO6	SLVOENSKO
SLLOVENSKO	SLOV.REPUBLIKA	SLOVENSK	SLOVENSKO	SLOVESNKOI	SLVOVENSKO
SLO	SLOVAK REPUBLIC	SLOVENSKO	SLOVENSKO SK	SLOVESNKOÍ	SOLENSKO
SLO VENSKO	SLOVAKIA	SLOVENSKA	SLOVENSKO,	SLOVESNSKO	SOLVENSKO
SLO VENSKO	SLOVAKOFARMA	SLOVENSKÁ	SLOVENSKO;	SLOVEŠNSKO	SOVENSKO
SLOBENSKO	SLOVANSKO	SLOVENSKA REP	SLOVENSKOA	SLOVEVENSKO	SR
SLOBVENSKO	SLOVBENSKO	SLOVENSKÁ REP	SLOVENSKOBRATIS	SLOVEWNS	SR 195
SLOCENKO	SLOVCENSKO	SLOVENSKA REP.	SLOVENSKOÉ	SLOVEWNSKO	SR;
SLOCENSKO	SLOVE NSKO	SLOVENSKÁ REP.	SLOVENSKOI	SLOVGENSKO	SR0
SLOCVENSKO	SLOVEBNSKO	SLOVENSKA REPUB	SLOVENSKOÍ	SLOVIENKSO	SR9
SLOENSKO	SLOVEČNSKO	SLOVENSKÁ REPUB	SLOVENSKOK	SLOVIENSKO	SRA
SLOENSKOÉ	SLOVEMSKO	SLOVENSKÁ REPUK	SLOVENSKOM	SLOVLENSKO	SSLOVENSKO
SLOESNKO	SLOVEN	SLOVENSKAO	SLOVENSKOP	SLOVNEKSO	SSR
SLOEVENSKO	SLOVEN S,O	SLOVENSKÉ	SLOVENSKOSK	SLOVNENKO	SVOVENSKO
SLOÉVENSKO	SLOVEN SKO	SLOVENSKEJ	SLOVENSKPO	SLOVNESKO	SWLOVENSKO

# Dobře nebo špatně?

- Kuřim 674
- 530 02 Spojil 92
- Brno 17
- Lužná  
163 00 Praha 6
- Opraveno:
  - Vojtova 674 664 34 Kuřim
- Opraveno:
  - Pod Hrází 92 530 02 Spojil
- Automaticky neopravitelné:
  - Velké město
- Automaticky neopravitelné:
  - Chybějící č.p./č.o.

# Dobře nebo špatně?

- Pernerova 35  
186 00 Karlín
- F. Procházky 50  
509 01 Nová Paka
- U Nádraží 140  
262 03 Mokrovraty
- Dvořákovo nábr. 12  
Příbram
- Opraveno:
  - Pernerova 35 186 00 Praha 8
- Opraveno:
  - F.F. Procházky
  - **PROCHÁZKA František Faustin**  
(\* 13. 1. 1749 Nová Paka, + 2. 12. 1809)
  - ulice F. Procházky je v Sušici
- Automaticky neopravitelné:
  - Neexistující ulice v Mokrovratech
  - 262 03 Mokrovraty 140 je správně
  - ale: U nádraží 140 existuje 3x
- Automaticky neopravitelné:
  - Dvořákovo nábrží v Příbrami existuje,  
má ale pouze 2 domy (č.p. 20 a 170)

# Důsledky nekvality

- 54900 NOVE MESTO – KRCIN 471, CZ
- Krátká 471, N.Město n/M., 54901, CY
  - Jsou výše uvedené adresy shodné?
  - Která z nich je správná?

✓ Obec:	Nové Město nad Metují
✓ Část obce:	Krčín
✓ Ulice:	Krátká
✓ Číslo popisné:	471
✓ Číslo orientační:	-
✓ PSČ:	54901
✓ Kód země:	CZ



# Důsledky nekvality

- Jestliže informační systém středně velké organizace obsahuje 3 miliony adres
  - => Řešení otázek správnosti a shody adres nastává ve více než 300 000 případech**
- Ani korektně, tedy podle všech pravidel zadaná adresa nemusí být správná (platná)
  - => V případě korespondenčních adres uvedené organizace by to znamenalo téměř 750 000 potenciálně nedoručitelných zásilek**



# Výhody dosažené kvality

- **Příklad:** Sdružování korespondence pro klienta do jedné obálky:
  - 4 zdrojové systémy
  - Hromadné rozesílání výpisů s měsíční periodou
  - Před konsolidací: cca 1 mil. jednotlivých zásilek
  - Po konsolidaci: cca 600 000 sdružených zásilek
  
  - => **úspora 400 000 zásilek** á 10 Kč/kus (obyčejné psaní do 50 g)
  - => 4 mil. x 12 = **48 mil. Kč/rok**

# Problémy s kvalitou a jejich řešení

- Doručitelnost adresy



# Produktové řady



- Čištění a identifikace, **adres**, obohacování
- Čištění a identifikace dat o osobách, obohacování
  - Fyzické osoby
  - Právníkové osoby
- **Prevence** nekvalitních dat – „Data Quality Firewall“
- Kompletní **Data Quality Management** – podpora celooorganizačních programů pro řízení datové kvality
- Jednotný pohled na občana a všechny jeho související informace
  - Kontaktní informace
  - Transakce
  - Rizika
  - Události
- Centrální správa těchto dat včetně automatického rozpoznání souvisejících záznamů
- Včetně kompletního řízení kvality dat

## Ověřování adres v registru UIR-ADR verze dat 631 stav dat k 6.3.2008

Nápověda

Vstupte údaje adresy:

	základní	rozšířené	kód
okres:	<input type="text"/>	<input type="checkbox"/>	
obec:	<input type="text"/>	<input type="checkbox"/>	
část obce:	Eliščino nábř. 1., HR. KRÁLOVÉ	<input type="checkbox"/>	
ulice:	<input type="text"/>	<input type="checkbox"/>	
PSC:	<input type="text"/>	<input type="checkbox"/>	
číslo:	domovní <input type="text"/>	<input type="checkbox"/>	
číslo:	orientační <input type="text"/>	<input type="checkbox"/>	
<input type="button" value="Vymazat vše"/>			

Max. počet položek: 100

Třídít podle:

názvu

Stav ověření adresy: Počet položek: 0

Zpráva: Zadaná část obce neexistuje.

MAPY.CZ

Internet **Firmy** **Mapy** Zboží Obrázky

Hr. Králové, Eliščino nábř. 1

Hledej

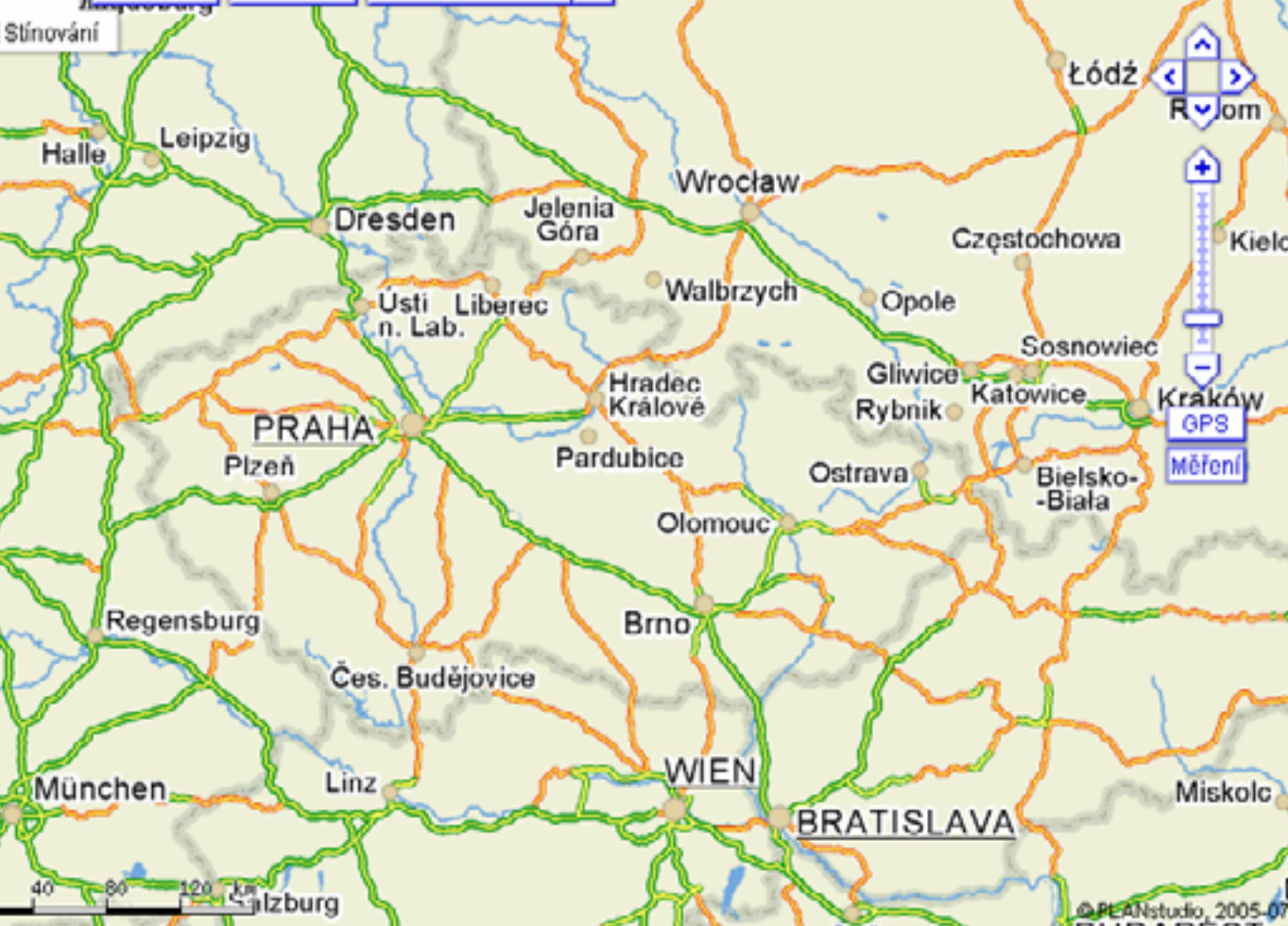
Zadejte ulici, město, adresu nebo firmu [Příklady](#) - [Nápověda](#)

[Vytiskni mapu](#) | [Sezn](#)

[Najdi v okolí](#) | [Další funkce](#)

Kladní **Turistická** **Fotomapa** **Datř mapy**

Vyhledávání **Plánovač tr**



**Vyhledávání**


Pro zadaný dotaz nebylo nic nalezeno

Copyright © 1996 - 2008 Seznam.cz, a.s.  
Zaměření adres: ČSÚ  
[Seznam](#) - [Nápověda](#) - [Připomínky](#)


**ataccama**  
Address Cleaner

Street:	<input type="text"/>
City:	<input type="text"/>
Post Office:	<input type="text"/>
Country:	<input type="text"/>
State:	<input type="text"/>
Country:	Česká republika <input type="button" value="v"/>
<input type="button" value="Reset"/> <input type="button" value="Submit"/>	
Address:	Eliščino nábř. 1., HR. KRÁLOVÉ
Country:	Česká republika <input type="button" value="v"/>
<input type="button" value="Reset"/> <input type="button" value="Submit"/>	

Address	Matching values	Post Office	UIR-ADR
Street:	Eliščino nábřeží		
LRN:	375		
SN:	1		
City:	Hradec Králové		
City district:	Věkoše		
District:	Hradec Králové		
ZIP:	50003		
Post district:	Hradec Králové 3		
Address ID:	23928085		

Map data ©2008 PPM, TeleAtlas, Terms of Use

# Znalostní báze dodávaná jako součást produktu v

## Typ osoby

- Fyzické osoby
- Právnícké osoby

## Rodné číslo – primární ID pro residential zákazníky

- vedle závazného formátu se ověřuje i vazba na další atributy (datum narození, pohlaví)

## Datum narození, Pohlaví

- kontrolují se jak nezávisle, tak v kontextu

## Jméno, příjmení

- parsing, obsáhlá báze replacementů, extrakce titulů

## Tituly

- akademické/společenské

## Občanský průkaz, příp. další doklady

- standardizace/formát

## Kontakty, Telefonní čísla

- E-mail, URL

## IČ – primární ID pro firemní zák

- vedle závazného formátu s kontextové vyhledávání v e nebo chybějící IČ)

## Název firmy

- Normalizace právních fore etalonem

## Číslo bankovního účtu

- validace na základě závazné vyhlášky, předčíslí, č. účtu

## Čísla karet

- validace na bázi tzv. Luhnova vzorce

## Adresní informace

- Parsing, identifikace komponent
- Rozsáhlá báze replacementů

## Identifikace automobilů





**ataccama**

**CZECH REPUBLIC**

**Ataccama Software, s.r.o.**

**Karolinská 654/2**

**186 00 Praha 8 - Karlín**

**david.holes@ataccama.com**

**info@ataccama.com**

**www.ataccama.com**

**CANADA**

**Ataccama Corporation**

**8500 Leslie St.**

**Markham, Ontario, L3T 7M8**

**info@ataccama.com**

**www.ataccama.com**