Citizens, do we really know you?

… or how can pure data contribute to our understanding

ITAPA

15.11.2006

Adastra Corporation

- Adastra Corporation is IT consulting company with ist headquarteds in Toronto, Canada
  - Offices in Bratislava, Praha, Franfurkt N/M, Ostrava
- Many customers in public, banking, telecommunication industry in Slovak and Czech market
- **Adastra with exclusive focus on:**
  - **Data Warehousing & Business Intelligence**
  - **Citizen Intelligence**
  - **Application and data integration**
  - **Application development and testing**
  - **Industry Specific Solutions**
- 350 employees, 210 in CZ, 40 in SR
- Revenue FYI 2006 > 22 mil EUR, average growth > 40%

What does it mean?

- We help to identify frauds in customs and tax area

- We bring flexible reporting capabilities based on data warehousing approach for all managing levels in organization

- We provide data cleansing and consolidation in central public files about citizens, adresses, cars

- …..and many many more answers

  for such questions are stored

  in your data and our solutions

Agenda
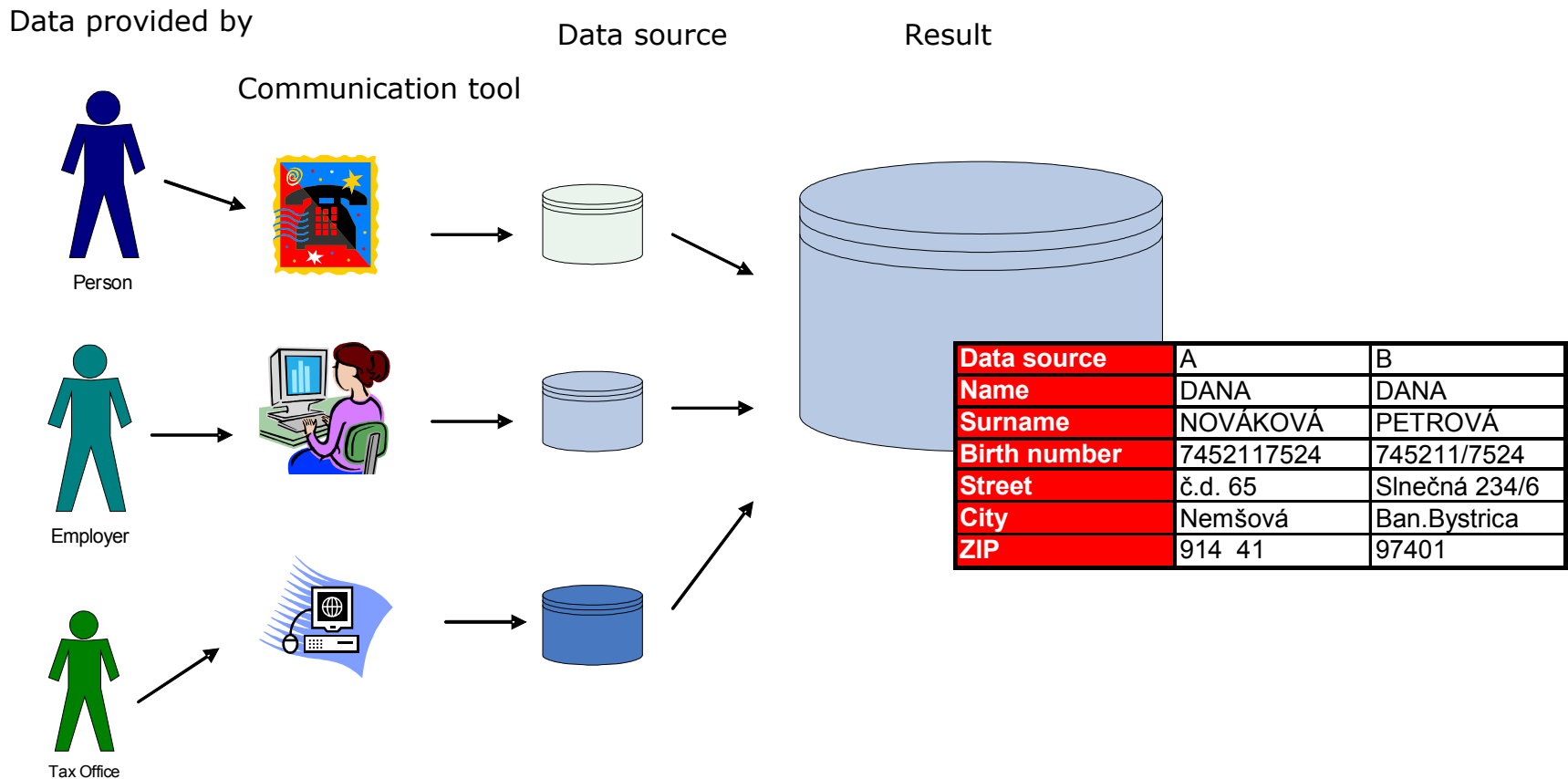
- Data quality – where is the problem?
- Purity.360 tool as an answer
- Data consolidation approach
- Solution architecture and properties
- Benefits

ADASTRA

Data Quality – cradle of problems

- **Lack of system integration** - Customer data are stored in separated systems based on different technologies, delivered by different contractors in different time – product / agenda oriented

- **Human factor** – information from many sources (directly from the citizen, employer, third party) are collected and typed by people (people make mistakes – typing errors, duplicities, inputs to incorrect data fields, incorrect interpretation, „creativity" of data entry clerks)

- **Lack of human capacity -** insufficient recourses for additional data cleansing, when it comes to data cleansing, where are always more important tasks to be accomplished ☺

ADASTRA

Data Quality

Data provided by

Data source

Result

Communication tool

Person

Employer

Tax Office

Data source	A	B
Name	DANA	DANA
Surname	NOVÁKOVÁ	PETROVÁ
Birth number	7452117524	745211/7524
Street	č.d. 65	Slnečná 234/6
City	Nemšová	Ban.Bystrica
ZIP	914 41	97401

ADASTRA

Data Quality

Data Accuracy - influenced by other „natural" changes in customer data

- Name changes 26 149 marriages and 11 553 divorces

- Change in address - citizens migration

- 51 657 new born babies and 50 951 deaths

- Immigrants – 5 273

(all data: Slovakia, 2005)

ADASTRA

Data Quality

Why is data quality hot issue now?

- **Integration** – effort to unify separate parts of IS and their interconnection to the business processes => effectiveness, time & cost saving, effective support of business processes

- Efficient **data exchange** by electronic communication

ADASTRA

Data Quality

..can influence answers to the questions..

- How many people do we have?

- What kind of social security benefits person receives?

- What can we offer to the person / what are his/her claims ?

- What is the customer value?

- What are the risks connected with the person?

Purity.360 – answer is already known

- Independent and open data quality management system fully compatible with any platform easily installed with into the existing IT infrastructure :

  - **Batch processing** – apply to all data already stored in the system to achieve single data quality improvement
  - **Online** – for all new data entered into the system with the aim to identify and to avoid existing and potential problems at the time of data entry

- And which ensures following activities :

ADASTRA

Purity.360 − Analysis

Identification, quantification, classification and data error analysis

- To know the current status of data quality

- To propose procedures, rules and standards for data quality improvement and sustaining improved status

At the same time according to agreed rules identifies records and groups of records which:

- – Can be **automatically corrected** by the system
- – Is needed to be cleansed manually because of very low data quality, low data completeness or level of incorrectness is so high, that it is impossible to automate data cleansing process

ADASTRA

Purity.360 – Measurement

Data quality measurement, monitoring and reporting

Data quality current status reports and statistics expressed by numerical and quality indicators

To be able:

- Analyze data

- Monitor data quality improvement in time when solution is implemented

ACCURACY　　　　**CORRECTENESS**　　　　**COMPLETNESS**

RELEVANCE　　　　**INTEGRITY**

Purity.360 – Processing

Data processing by cleansing and unification algorithms which automate:

- **Cleansing** –data parsed into components to help identify problematic data and its correction– typing errors, incorrect formats, etc.

- **Standardization** – data to unified format to be able to compare content with different registers and referential databases (or between different IT system)

- **Enrichment** – missing data completion

- **Unification** – one subject records identification – e.g. to find and explicit identification of all records about one person, address, vehicle

- **Deduplication** – master record selection (master is record with the highest data quality and it represents the subject)

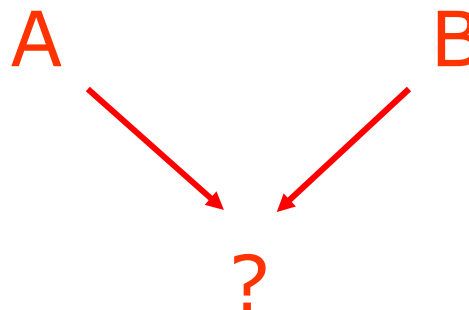- **Identification** – new data entries – to identify subject to which is the new record connected

Purity.360 – typing error example

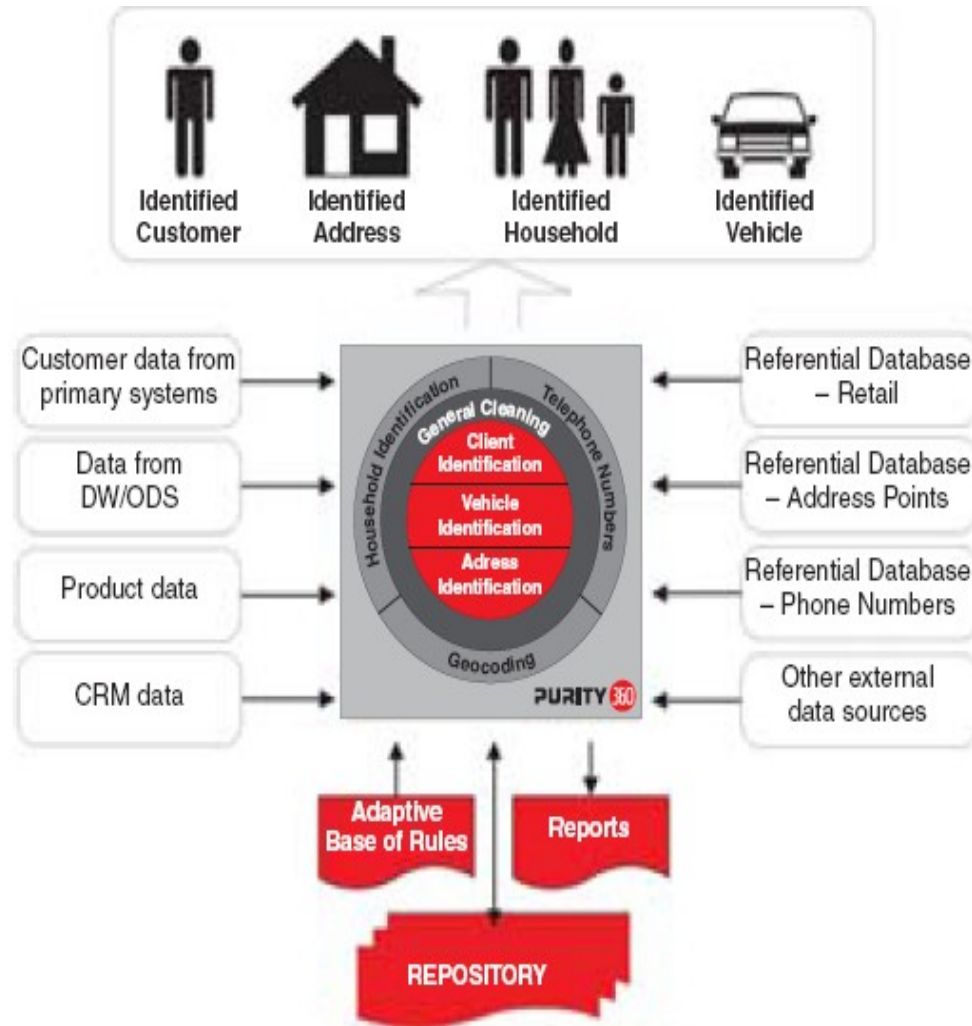|	
DUBNAD VAHOM	DUBNICA NAD VÁHOH
DUBNBICA NAD VÁHOM	DUBNICA NAD VAHON
DUBNBICA NAD VAHOM	DUBNICA NAD VAHOOM
DUBNCA NAD VÁHOM	DUBNICA NAD VAOM
DUBNCA NAD VAHOM	DUBNICA NAD VHOM
DUBNIA NAD VAHOM	DUBNICA NADV VAHOM
DUBNIC NAD VÁHOM	DUBNICA NADVAH
DUBNICA NV	DUBNICA ND VAHOM
DUBNICA VAHOM	DUBNICA NV
DUBNICA AND VAHOM	DUBNICA NV V
DUBNICA MAD VAHOM	DUBNICA NV A
DUBNICA N V	DUBNICA NVAH
DUBNICA N VAHOM	DUBNICA NVAHOM
DUBNICA N VÁHOM	DUBNICA NVÁHOM
DUBNICA NA DVAHOM	DUBNICA NVM
DUBNICA NA VAHOM	DUBNICA NVPREJTE
DUBNICA NA VÁHOM	DUBNICA NVTRICA
DUBNICA NAAAD VAHOM	DUBNICA V
DUBNICA NAD VAH	DUBNICA VAHOM
DUBNICA NAD VÁH	DUBNICANAD VAHOM
DUBNICA NAD VAHO	DUBNICANADVÁHOM
DUBNICA NAD VAHO M	DUBNICANVAHOM

Purity.360 –unification exercise example

- Cleansed and verified records from various sources are paired (identified)

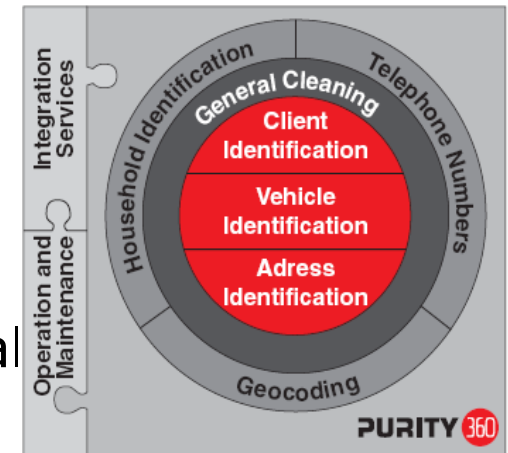| Data source	A	B
Name	DANA	DANA
Surname	NOVÁKOVÁ	PETROVÁ
Birth number	7452117524	745211/7524
Street	č.d. 65	Slnečná 234/6
City	Nemšová	Ban.Bystrica
ZIP	914 41	97401

A   B

?

Purity.360 - Solution architecture

Purity.360 - Solution properties

- Modular – **PurePerson** and **PureAddress** extending modules
  - Household Identification module (**PureHouse**) uses identified clients, addresses and additional information to define household
  - Geocoding (**PureGeo**) supplements the identified address with geographical information
  - Telephone numbers (**PurePhone**)

- **Designed for Slovakia** – developed on the bases of detailed knowledge of the specifics of clients identification within Slovakia

- Uses **external sources** – referential databases of addresses and economic subjects

- Both **batch** (incremental) and interactive regime for **on-line** verification of data

ADASTRA

Consolidation goals and benefits

- Quality of customer faced processes

- Reporting – internal – management reporting, external – to the other institution (e.g. supervisory body)

- Data exchange – electronic communication – between institutions

- Fraud detection

- Basic condition for mail delivery and mailing cost decrease, elimination of multiple mail to the same person

- Address data accuracy

- Decrease of financial cost for management, processing and archiving of incorrect, meaningless and redundant data (direct economic loss)

ADASTRA

Adastra, s.r.o.

Francisciho 4, Bratislava

info@adastracorp.sk

www.adastracorp.sk

Adastra Corporation

8500 Leslie St., Markham

Ontario, L3T 7M8 Canada

info@adastracorp.com

www.adastracorp.com

Adastra, s.r.o.

Karolinská 654/2, Praha 8

info@adastra.cz

www.adastra.cz

ADASTRA