Business Intelligence and Analytics applied to Public Housing

Doctoral Consortium @ ADBIS 2019

September 8th, 2019 in Bled, Slovenia

E. Scholly[1,2], C. Favre[1], E. Ferey[2], S. Loudcher[1]

[1]University of Lyon, Lyon 2, ERIC EA 3083
[2]BIAL-X

Introduction

Context

A business issue

Context

A business issue

- **Public Housing** : dwellings, occupants, overdue, patrimony, ...

Context

A business issue

- **Public Housing** : dwellings, occupants, overdue, patrimony, ...

Context

A business issue

- **Public Housing** : dwellings, occupants, overdue, patrimony, …

Three main thematics

Context

A business issue

- **Public Housing** : dwellings, occupants, overdue, patrimony, ...

Three main thematics

- **Business Intelligence** (BI) : ETLs, data warehouses, OLAP, ...

Context

A business issue

- **Public Housing** : dwellings, occupants, overdue, patrimony, ...

Three main thematics

- **Business Intelligence** (BI) : ETLs, data warehouses, OLAP, ...
- **Data Science** (DS) : knowledge extraction, Machine Learning, ...

Context

A business issue

- **Public Housing** : dwellings, occupants, overdue, patrimony, ...

Three main thematics

- **Business Intelligence** (BI) : ETLs, data warehouses, OLAP, ...
- **Data Science** (DS) : knowledge extraction, Machine Learning, ...
- **Big Data** : Volume, Variety, Velocity, ...

Context

A business issue

- **Public Housing** : dwellings, occupants, overdue, patrimony, ...

Three main thematics

- **Business Intelligence** (BI) : ETLs, data warehouses, OLAP, ...
- **Data Science** (DS) : knowledge extraction, Machine Learning, ...
- **Big Data** : Volume, Variety, Velocity, ...

Context

A business issue

- **Public Housing** : dwellings, occupants, overdue, patrimony, ...

Three main thematics

- **Business Intelligence** (BI) : ETLs, data warehouses, OLAP, ...
- **Data Science** (DS) : knowledge extraction, Machine Learning, ...
- **Big Data** : Volume, Variety, Velocity, ...

$\rightarrow$ How does all this blend ?

What data ?

Several data sources

What data ?

Several data sources

1. Internal data

What data ?

Several data sources

1. Internal data
   - Landlord's data
   - Dwellings, occupants, overdue, ...
   - Mostly relational data
   - BI analyses, simple DS analyses

What data ?

Several data sources

1. **Internal data**
    - Landlord's data
    - Dwellings, occupants, overdue, ...
    - Mostly relational data
    - BI analyses, simple DS analyses
2. **External data**

What data ?

Several data sources

1. **Internal data**
    - Landlord's data
    - Dwellings, occupants, overdue, ...
    - Mostly relational data
    - BI analyses, simple DS analyses
2. **External data**
    - Open data (+ social networks)
    - Environment
    - (possibly) Big Data
    - Advanced DS analyses

Table of contents

Data storage and management

Business Intelligence and Analytics

Business Intelligence (BI)

Methods and tools for collecting, storing, organizing and analyzing data to support decision-making

[Chen et al., 2012, Larson and Chang, 2016, Mortenson et al., 2015, Baars and Ereth, 2016, Gröger, 2018]

Business Intelligence and Analytics

Business Intelligence (BI)

Methods and tools for collecting, storing, organizing and analyzing data to support decision-making

Business Analytics (BA)

The use of Data Science methods on a company's data

[Chen et al., 2012, Larson and Chang, 2016, Mortenson et al., 2015, Baars and Ereth, 2016, Gröger, 2018]

Business Intelligence and Analytics

Business Intelligence (BI)

Methods and tools for collecting, storing, organizing and analyzing data to support decision-making

Business Analytics (BA)

The use of Data Science methods on a company's data

What about BI ?

- BI&A

[Chen et al., 2012, Larson and Chang, 2016, Mortenson et al., 2015, Baars and Ereth, 2016, Gröger, 2018]

Business Intelligence and Analytics

Business Intelligence (BI)

Methods and tools for collecting, storing, organizing and analyzing data to support decision-making

Business Analytics (BA)

The use of Data Science methods on a company's data

What about BI ?

- BI&A
- BI & BA

[Chen et al., 2012, Larson and Chang, 2016, Mortenson et al., 2015, Baars and Ereth, 2016, Gröger, 2018]

Business Intelligence and Analytics

Business Intelligence (BI)

Methods and tools for collecting, storing, organizing and analyzing data to support decision-making

Business Analytics (BA)

The use of Data Science methods on a company's data

What about BI ?

- BI&A
- BI & BA
- BI $\rightarrow$ BA

[Chen et al., 2012, Larson and Chang, 2016, Mortenson et al., 2015, Baars and Ereth, 2016, Gröger, 2018]

Run BI and BA analyses...

Data Intelligence

Run BI and BA analyses...

- Separately

Run BI and BA analyses...

- Separately
- Together

Data Intelligence

Run BI and BA analyses...

- Separately
- Together
- (possibly) on Big Data

Run BI and BA analyses...

- Separately
- Together
- (possibly) on Big Data

Data Intelligence

Run BI and BA analyses...

- Separately
- Together
- (possibly) on Big Data

Data Intelligence

Perform analyses, simple or advanced, on all types of data

Data Intelligence

Run BI and BA analyses...

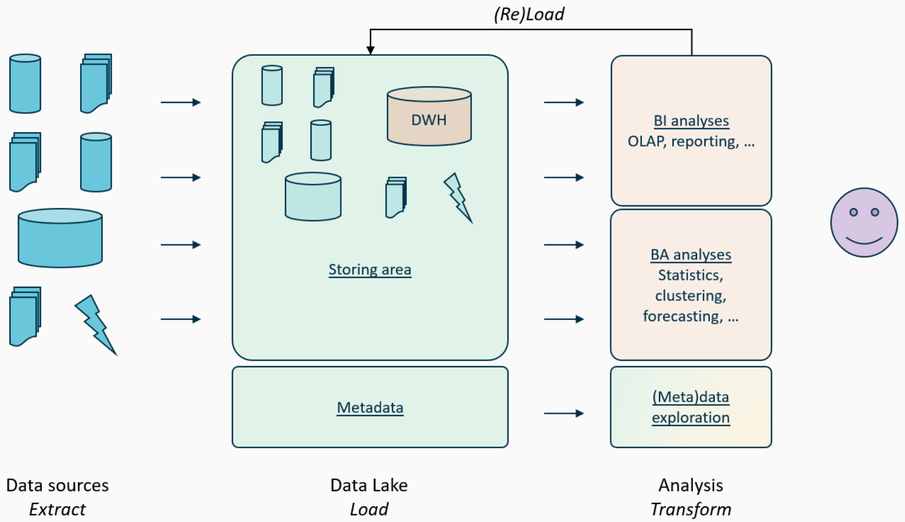
- Separately
- Together
- (possibly) on Big Data

Data Intelligence

Perform analyses, simple or advanced, on all types of data

$\rightarrow$ How ?

Data Intelligence in practice



Data Lakes

Data Lake [Dixon, 2010]

A data lake is a large repository of heterogeneous raw data, supplied by external data sources and from which various analyses can be performed.

Data Lakes

Data Lake [Dixon, 2010]

A data lake is a large repository of heterogeneous raw data, supplied by external data sources and from which various analyses can be performed.

Two main characteristics

- *Schema-on-read*

[Miloslavskaya and Tolstoy, 2016]

Data Lake [Dixon, 2010]

A data lake is a large repository of heterogeneous raw data, supplied by external data sources and from which various analyses can be performed.

Two main characteristics

- *Schema-on-read*
- Data variety

[Miloslavskaya and Tolstoy, 2016]

Data Lakes

Data Lake [Dixon, 2010]

A data lake is a large repository of heterogeneous raw data, supplied by external data sources and from which various analyses can be performed.

Two main characteristics

- *Schema-on-read*
- Data variety

[Miloslavskaya and Tolstoy, 2016]

Data Lakes

Data Lake [Dixon, 2010]

A data lake is a large repository of heterogeneous raw data, supplied by external data sources and from which various analyses can be performed.

Two main characteristics

- *Schema-on-read*
- Data variety

→ Need for a metadata system

[Miloslavskaya and Tolstoy, 2016]

Data Lakes

Data Lake [Dixon, 2010]

A data lake is a large repository of heterogeneous raw data, supplied by external data sources and from which various analyses can be performed.

Two main characteristics

- *Schema-on-read*
- Data variety

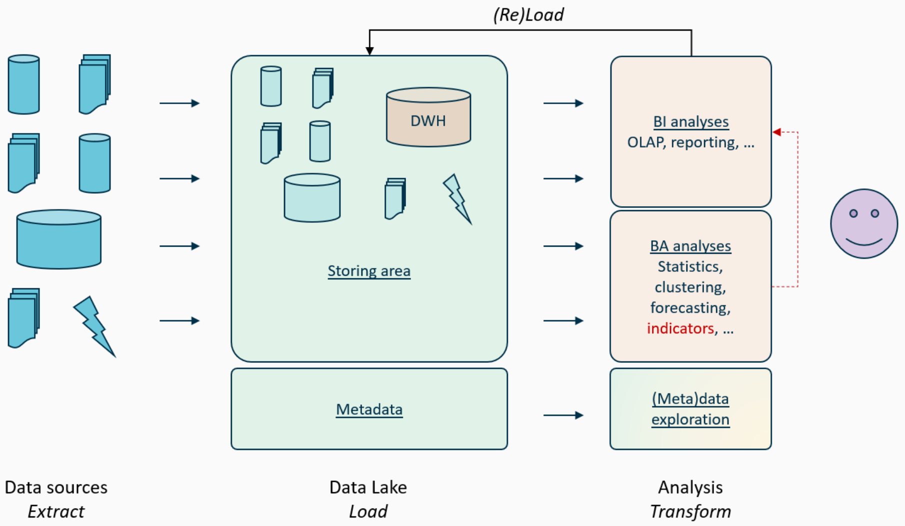
$\rightarrow$ Need for a metadata system

Big research field

[Miloslavskaya and Tolstoy, 2016]

Attractiveness

Data Intelligence in practice



Defining attractiveness

Attractiveness of what ?

Defining attractiveness

Attractiveness of what ?

1. Dwelling

Defining attractiveness

Attractiveness of what ?

1. Dwelling
2. Residency

Defining attractiveness

Attractiveness of what ?

1. Dwelling
2. Residency
3. Neighborhood

Defining attractiveness

Attractiveness of what ?

1. Dwelling (internal)
2. Residency
3. Neighborhood

Defining attractiveness

Attractiveness of what ?

1. Dwelling (internal)
2. Residency (internal - external)
3. Neighborhood

Defining attractiveness

Attractiveness of what ?

1. Dwelling (internal)
2. Residency (internal - external)
3. Neighborhood (external)

Defining attractiveness

Attractiveness of what ?

1. Dwelling (internal)
2. Residency (internal - external)
3. Neighborhood (external)

$\rightarrow$ Strategic Patrimony Plan

Defining attractiveness

Attractiveness of what ?

1. Dwelling (internal)
2. Residency (internal - external)
3. Neighborhood (external)

$\rightarrow$ Strategic Patrimony Plan

Advanced indicators

Defining attractiveness

Attractiveness of what ?

1. Dwelling (internal)
2. Residency (internal - external)
3. Neighborhood (external)

→ Strategic Patrimony Plan

Advanced indicators

- Machine Learning algorithms

Defining attractiveness

Attractiveness of what ?

1. Dwelling (internal)
2. Residency (internal - external)
3. Neighborhood (external)

$\rightarrow$ Strategic Patrimony Plan

Advanced indicators

- Machine Learning algorithms
- Back-feeding the lake

Defining attractiveness

Attractiveness of what ?

1. Dwelling (internal)
2. Residency (internal - external)
3. Neighborhood (external)

→ Strategic Patrimony Plan

Advanced indicators

- Machine Learning algorithms
- Back-feeding the lake
- Enrich BI analyses

First results and future outcomes

First contribution

Work done with P. N. Sawadogo [Sawadogo et al., 2019, Scholly et al., 2019]

First contribution

Work done with P. N. Sawadogo [Sawadogo et al., 2019, Scholly et al., 2019]

- Our definition of a Data Lake
- Key features for metadata systems
- Metadata typology in three categories
- MEtadata model for DAta Lakes (MEDAL)

Presented at 4 PM in this room !

What's next ?

Work in progress

What's next ?

Work in progress

- Implementation(s) of MEDAL

What's next ?

Work in progress

- Implementation(s) of MEDAL
- Retrieve all data

What's next ?

Work in progress

- Implementation(s) of MEDAL
- Retrieve all data
- Development of a complete data lake

What's next ?

Work in progress

- Implementation(s) of MEDAL
- Retrieve all data
- Development of a complete data lake
- Tests and comparisons

Thank you for your attention!

Questions?

Baars, H. and Ereth, J. (2016).
From data warehouses to analytical atoms-the internet of things as a centrifugal force in business intelligence and analytics.
In *24th European Conference on Information Systems (ECIS), Istanbul, Turkey*, page ResearchPaper3.

Chen, H., Chiang, R. H., and Storey, V. C. (2012).
Business intelligence and analytics: from big data to big impact.
MIS quarterly, pages 1165–1188.

Dixon, J. (2010).
Pentaho, Hadoop, and Data Lakes.
https://jamesdixon.wordpress.com/2010/10/14/pentaho-hadoop-and-data-lakes/.

Gröger, C. (2018).
Building an industry 4.0 analytics platform.
Datenbank-Spektrum, 18(1):5–14.

Larson, D. and Chang, V. (2016).
A review and future direction of agile, business intelligence, analytics and data science.
International Journal of Information Management, 36(5):700–710.

Miloslavskaya, N. and Tolstoy, A. (2016).
Big Data, Fast Data and Data Lake Concepts.
In *7th Annual International Conference on Biologically Inspired Cognitive Architectures (BICA 2016), NY, USA*, volume 88 of *Procedia Computer Science*, pages 1–6.

Mortenson, M. J., Doherty, N. F., and Robinson, S. (2015).
Operational research from taylorism to terabytes: A research agenda for the analytics age.
European Journal of Operational Research, 241(3):583–595.

Sawadogo, P., Scholly, É., Favre, C., Ferey, É., Loudcher, S., and Darmont, J. (2019).
Metadata systems for data lakes: Models and features.

Scholly, E., Sawadogo, P., Favre, C., Ferey, E., Loudcher, S., and Darmont, J. (2019).
Systèmes de métadonnées d'un lac de données: modélisation et fonctionnalités.