Some data warehouse procedures in banking

Prof. dr. sc. Damir Kalpić
University of Zagreb Faculty of electrical engineering and computing

Dr. sc. Vedran Vrbanić, dipl. ing.
Koios Consulting, ltd.

August 2016

The issues to be presented (1)

- **In university environment, acquire a real-life SWE project with a major bank**
  - The Bank opted for outsourcing instead of hiring new staff
  - Alumni influential in the Bank
  - Addressed their Alma Mater to propose a joint project
  - Negotiated and signed a contract for consultancy and programming help in data cleansing and warehousing

The issues to be presented (2)

- **Employ high quality young professionals**
  - SWE professionals in high demand on the labour market, in spite of (permanent) crisis in Croatia
  - Employ our own “products”
  - Employ them fully on the project for the Bank and provide compensation for the departmental teaching staff

The issues to be presented (3)

- **Establish a well motivated project team in university environment**
  - Transparent management (all the data accessible to the team members)
  - Charging the Bank according to each individual's level of competence (Profiles 0 to 4) and the monthly hourly workload
  - Half of the monthly individual gross income goes to the employees' individual portfolios:
    - Modest fixed salary + generous variable bonus
    - Additional HW (Faculty property, but for individual use), conference fees, travelling costs, etc. on individual choice

The issues to be presented (4)

- **Benefits for education**
  - Existing teaching staff additionally paid for performing educational activities instead of the project staff
  - Presenting to students real-life case studies
- **Benefits for scientific research**
  - Contact with real-life professional challenges triggers the motivation for research
- **General benefit for the Faculty**
  - Additional income
  - New equipment
  - Increased relevance and image,...

The issues to be presented (5)

- **Reasons to establish s start-up company**
  - Senior (i.e. working already for 3 years) project members were receiving income comparable to full professors' salary
  - What if the Bank closes the project?
  - A company would have some market value even if that happens
  - More appropriate environment for commercialisation of knowledge

The issues to be presented (6)

- **Principles of management in the new start-up Koios Consulting**
  - The ownership was split:
    - among project members (according to their accumulated contribution to the project)
    - the project leader + 2 closest collaborators of the project leader (all 3 professors at FER)
  - The project leader waived from his relatively high proposed proportion in ownership in order to increase the employees' share
    - Increase overall motivation

The issues to be presented (7)

- **Analysis after 8 years**
  - In spite of the global crisis created in 2008, Koios has weathered until now
  - In most of the years, the profit was distributed among the owners
  - In a few years, the profit was withhold to diminish the future risk
- **Scientific contribution in the PhD thesis of a Koios employee**
  - To be discussed further on...

The principal author

Vedran Vrbanić, PhD

Thesis:

- Incorporation of ontology model into banking data warehouse system, Mentor: D. Kalpić, FER, Zagreb, 2015

- Graduated at University of Zagreb
  - Faculty of electrical engineering and computing
- Employed on the Faculty's Department of applied computing
  - Professional collaborator since 2007
  - Paid from a **consulting, data-cleansing and data warehousing** project for the Privredna banka Zagreb (Intesa Sanpaolo)
  - Project leader D. Kalpić
  - A team of up to 8 graduated engineers

The Company KOIOS

- **In 2008 a start-up company Koios Consulting was established**
- **Ceding of the contract with PBZ from FER to KOIOS**
- **Owners:**
  - Former students (86%) + professors (Kalpić, Mornar, Fertalj)
  - Slowly but steadily growing and profitable
  - In 2016 transferred to UK (to diminish paperwork, endemic in Croatia), just in the wake of Brexit ☹

Wikipedia:

- **Koios** (also known as **Coeus**) is a son of Ouranos and Gaea. He is the Titan of farsight, intellect, and knowledge, and is also the Titan Lord of the North. His Roman form is **Polus**.
- **Old Greek god of business intelligence** ☺

Thesis contributions

- How to make the programmers understand what they are doing?

➢ **Banking ontology**

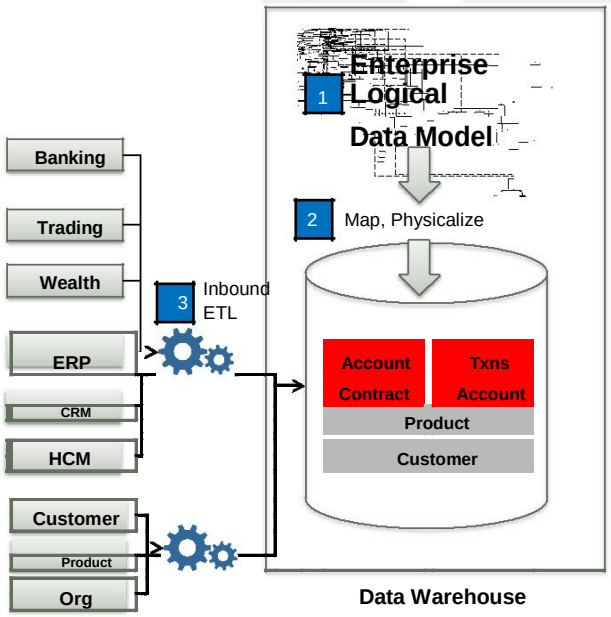
- How to schedule the ETL activities?

➢ **Priority based optimisation**

Introduction

- Data warehousing in banking
  - Large quantity of data
  - Focus on technical problems of data extraction and retrieval
  - Business rules “hidden” within the program code
  - Transparency diminished
  - High complexity
  - Problems in integration and unification
  - Obligatory reporting to regulatory authorities
  - Indispensable for decision making
  - ETL scheduling to provide timeliness
- Demystification of the business domain
- Improvement of flexibility and transparency
- Declarative approach

Introduction – How to create a unique source of facts?



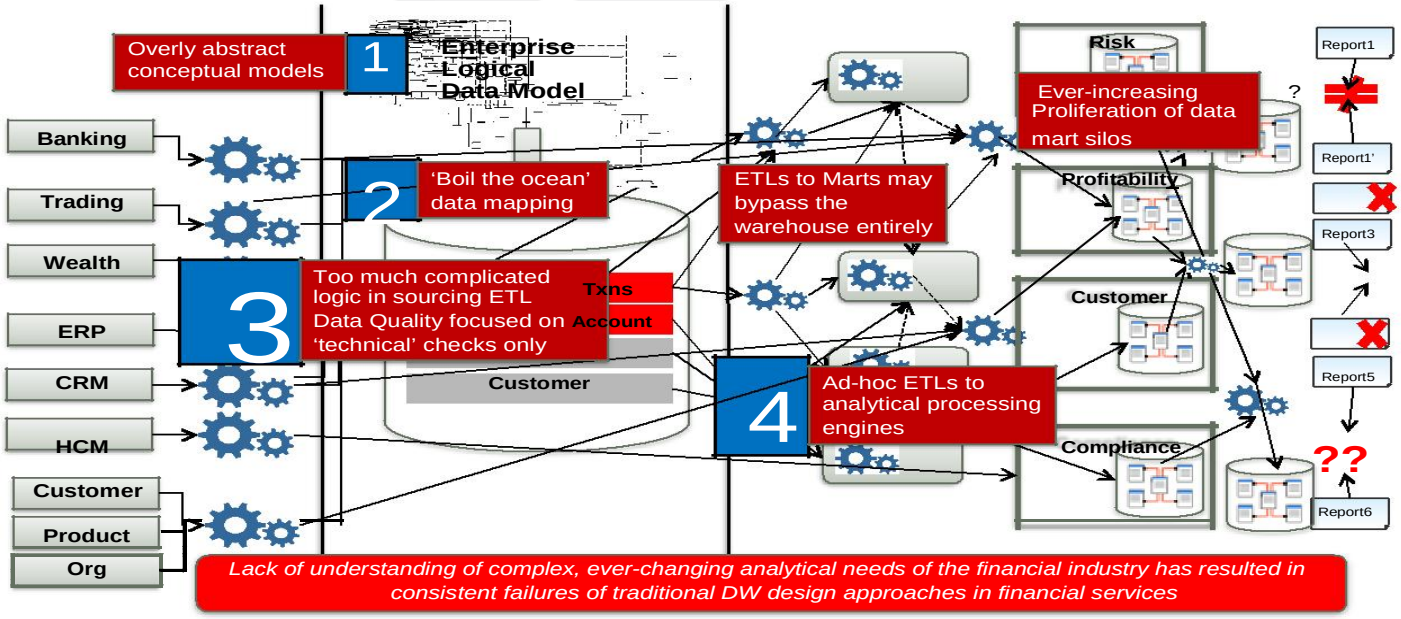
Goal: 'Single Source of Truth'

Common Approaches

- *Start with Enterprise Logical Data Model*
- *Map source data*
- *Physicalize 'selected' portions*
- *Build ETL to populate from sources*
- *Repeat process for each data source*

Source: Evolving the Data Warehouse: The Next Generation for Financial Services Institutions, An Oracle White Paper, May 2011

Introduction – Complexity



Source: Evolving the Data Warehouse: The Next Generation for Financial Services Institutions, An Oracle White Paper, May 2011

Motivation for introduction of ontology

- Regardless to
  - The level of system arrangement
  - Development methodologies
  - Technologies
  - Security policies
- Complexity of relationships among data in warehouse is unquestionable
- Idea
  - Identify the critical parts of the system, from the perspective of connectivity among different business entities
  - Model them more efficiently

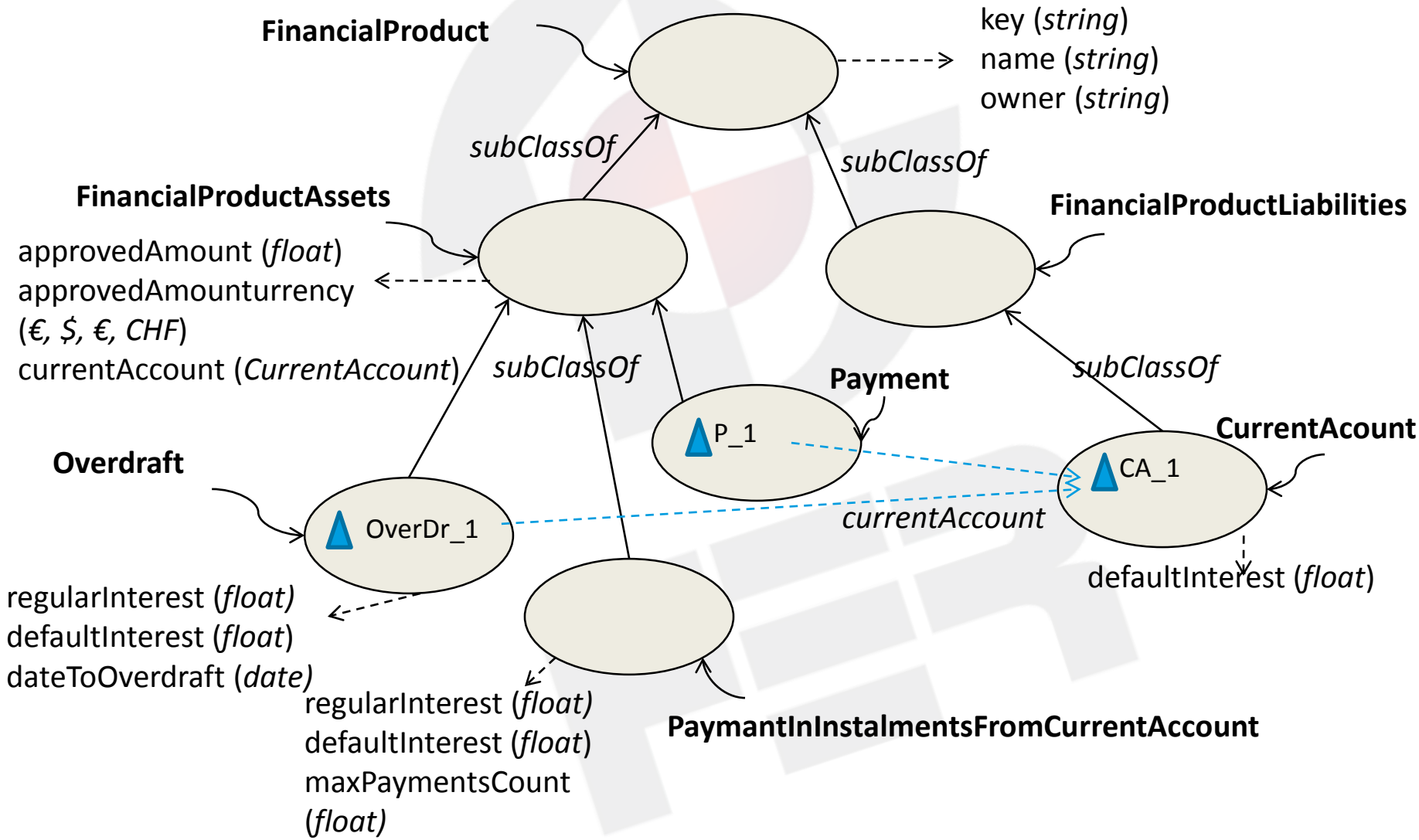
Semantic technologies

- Resource description framework (RDF)
  - Subject, predicate, object
- Resource description framework scheme (RDFS)
  - Classes, inheritance, domains, and predicate co-domains
- OWL
  - Web ontology language
- Ontology
  - Formal, explicit specification of conceptualisation
  - Set of precisely defined concepts for a framework for information exchange
  - Intrinsic problem with ontology:
    - Ambiguity, depending on the point of view

Example for ambiguity: *Food Ontology*

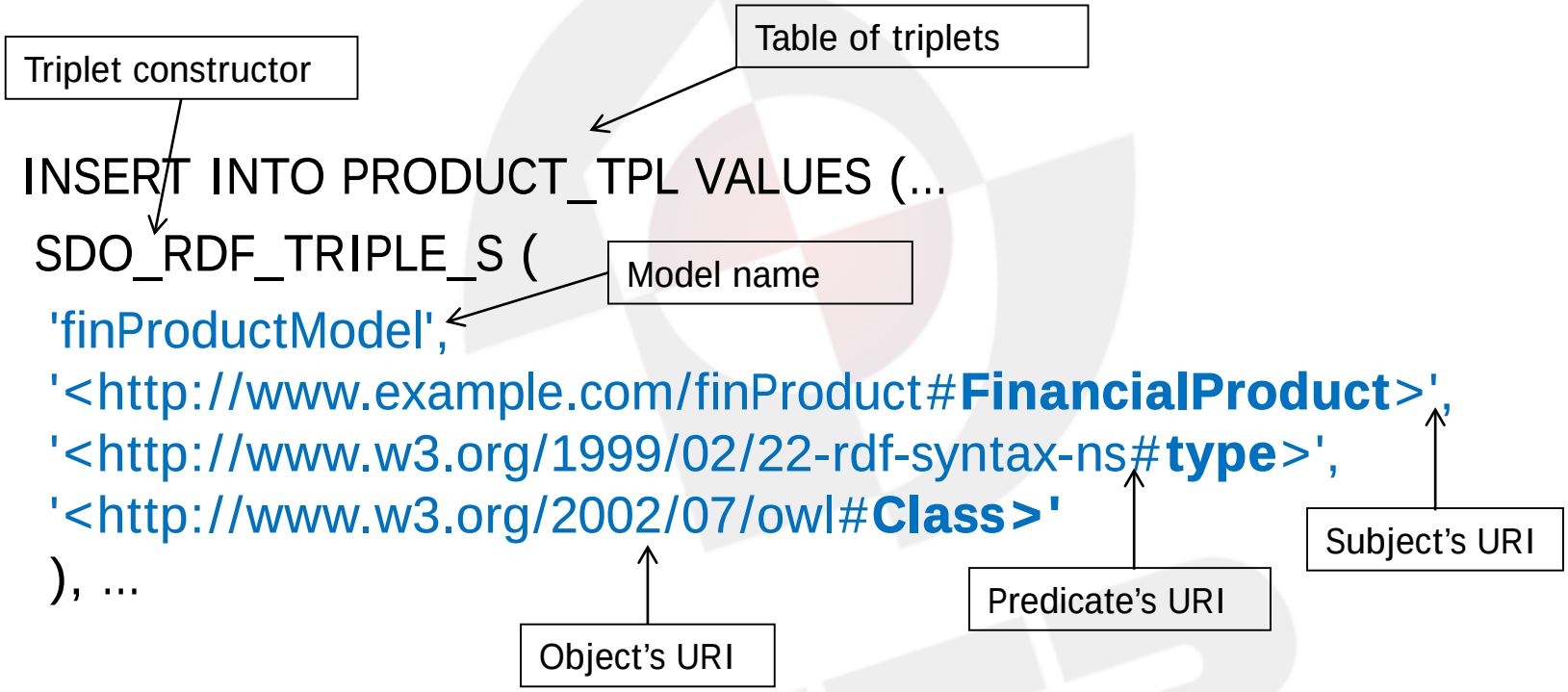
- **Gourmet's view on *recommended food*:**
- Oysters
  - Ask a Jew
- Smoked ham
  - Ask a Muslim
- Beefsteak
  - Ask a Hindu
- Hennessy Cognac and Cohiba cigar
  - Ask a sportsman
- Sacher cake
  - Ask a diabetic

Illustrative example in banking



Model coding

- Class definition



Model coding

- Inheritance

```
INSERT INTO PRODUCT_TPL VALUES (...
 SDO_RDF_TRIPLE_S (
  'finProductModel',
  '<http://www.example.com/finProduct#CurrentAccount>',
  '<http://www.w3.org/2000/01/rdf-schema#subClassOf>',
  '<http://www.example.com/finProduct#FinancialProductLiability>'
  ), ...
```

Model coding

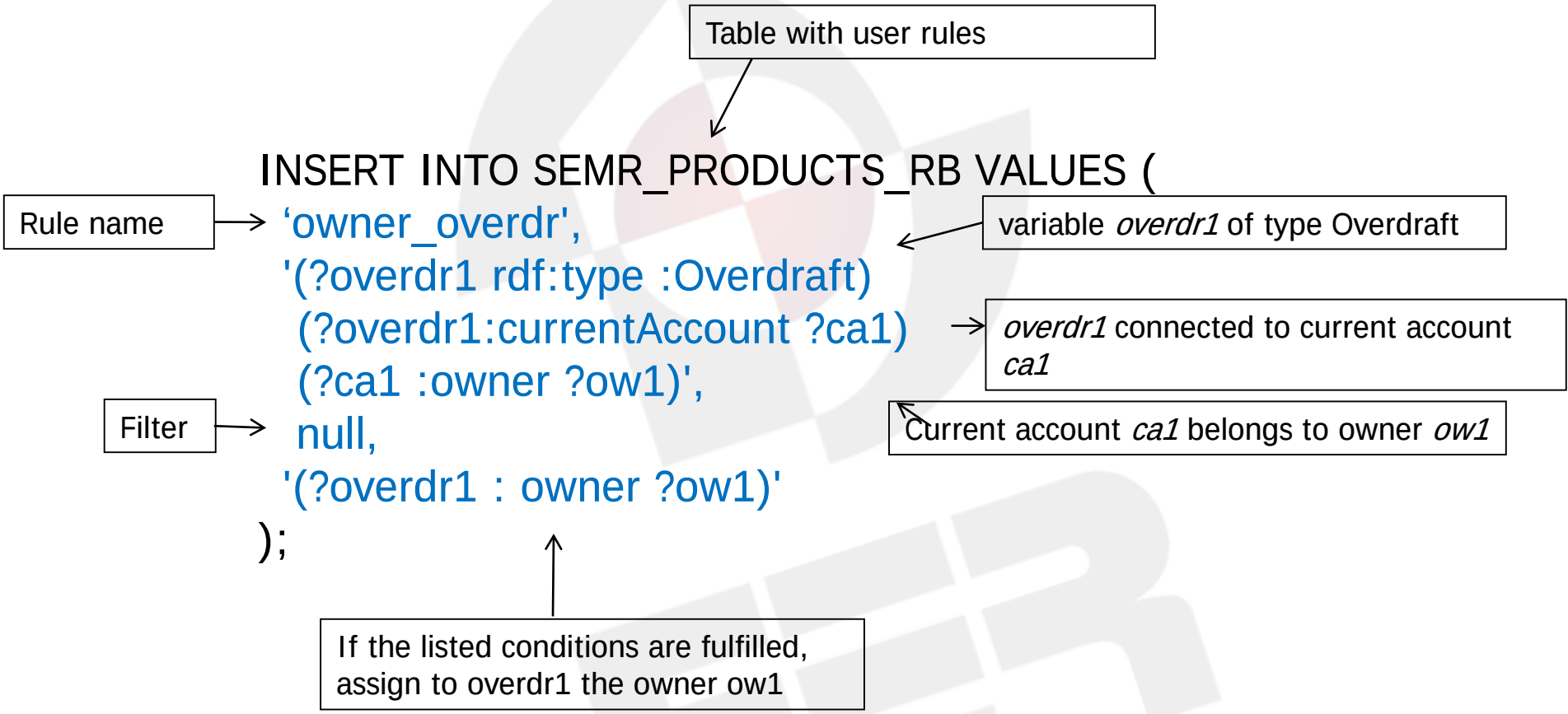
- Object's attributes

Subject	Predicate	Object
Example:currentAccount	rdf:type	owl:ObjectProperty
Example:currentAccount	rdf:type	owl:FunctionalProperty
example: currentAccount	rdfs:range	_:dummyNode
_:dummyNode	rdf:type	owl:Restriction
_:dummyNode	owl:allValuesFrom	example: CurrentAccount
example:CA_1	rdf:type	example: CurrentAccount
Example:OverDr_1	rdf:type	example:Overdraft
Example:OverDr_1	example: currentAccount	example:CA_1

SPARQL

- Query language for RDF
- Analogy to query language SQL
- Example of rules
  - If an overdraft is connected to a certain current account, the same owner is assigned to the overdraft.

Entering of business rules



Application of the defined rule

Subject	Predicate	Object
example:CA_1	rdf:type	example:CurrentAccount
example:CA_1	example:owner	John Smith
example:OverDr_1	rdf:type	example:Overdraft
example:OverDr_1	example:currentAccount	example:CA_1
example:OverDr_1	example:owner	John Smith

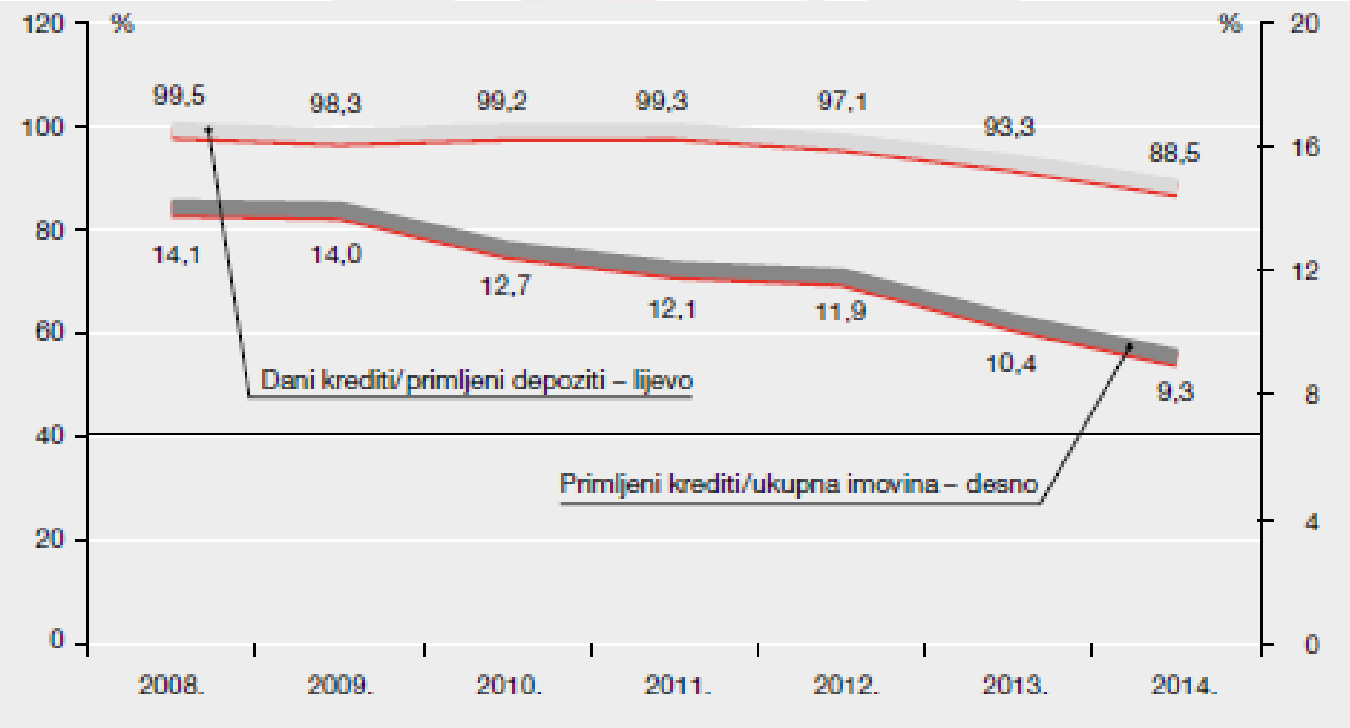
- Explicitely defined hierarchical and associative relationships

Research

- As a rule, the technology is not a problem
- Understanding of data
- Volatile system parts
  - Flexibility
  - Agility
- Ontology
  - Compulsory understanding of the topic before modelling of relationships
  - Triplets – flexibility
  - Domain description is not separated to the data dictionary
- A complex business problem – Report of maturity of assets and liabilities
  - Applicability of the described technology
  - Justification for introduction into the system
  - Into which parts of the system

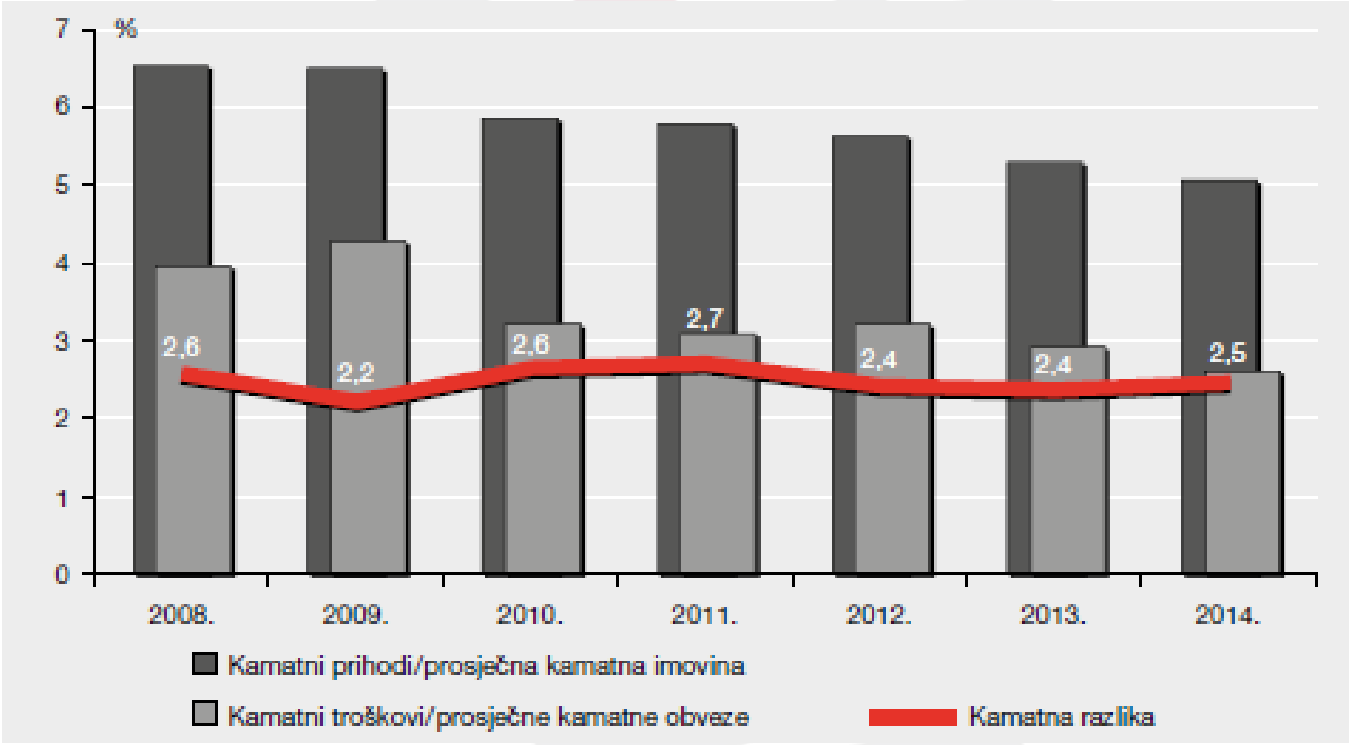
Assets and liabilities

- Collecting and placing of funds
  - Liabilities - deposits and received loans
  - Credit lines



Interest incomes and costs

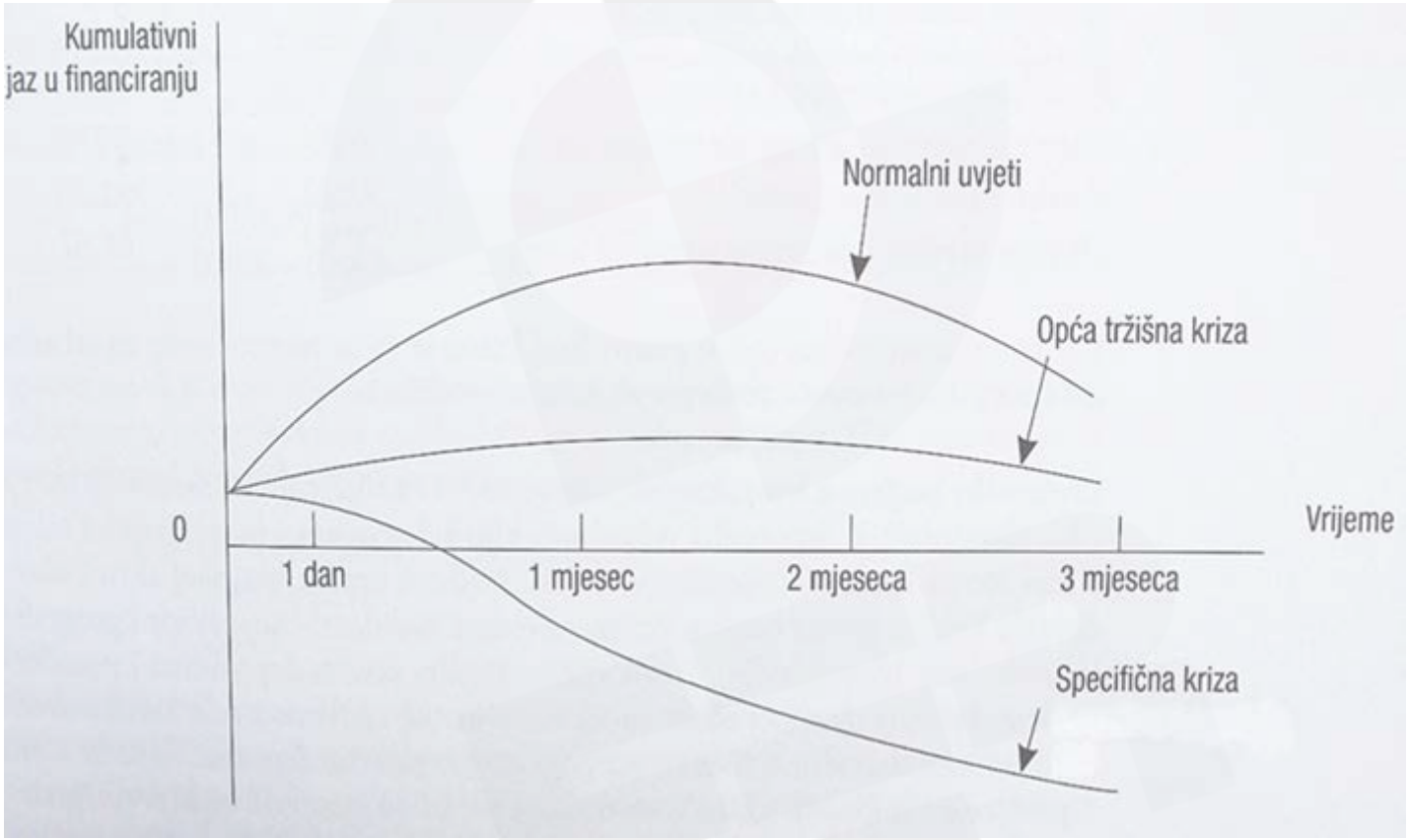
- Asymmetric forming of interest rates
  - Interest margin – about 2.5%



Maturity mismatch

- Maturity
  - Liabilities – short
  - Assets – long
  - Transformation of short term liabilities into long term assets
- Liquidity risk
  - Lack of assets within a certain perion of time
  - Market, operational and credit risks
  - Endangers the reputation
  - Overcomes single institutions

Example for projected liquidity dynamics



Analytics of residual maturity

Date	Product	Contract	Client	Account	Maturity	Amount
31.08.2016	Consumer loan	Contract 1	Client1	The principal of consumer loan maturity of up to two years	10.09.2016	630 €
31.08.2016	Consumer loan	Contract 1	Client1	Accrued interest on consumer loan	10.09.2016	58 €
31.08.2016	Consumer loan	Contract 1	Client1	Specific reserves for potential cumulative losses	01.09.2016	15 €
31.08.2016	Consumer loan	Contract 1	Client1	The principal of consumer loan maturity of up to two years	10.10.2016	625 €
31.08.2016	Consumer loan	Contract 1	Client1	Specific reserves for potential cumulative losses	01.10.2016	14 €
...	...	...	...	...	...	...
31.08.2016	Consumer loan	Contract 1	Client1	The principal of consumer loan maturity of up to two years	10.06.2017	620 €
31.08.2016	Consumer loan	Contract 1	Client1	Specific reserves for potential cumulative losses	01.06.2017	14 €
31.08.2016	Consumer loan	Contract 1	Client1	Interest income on long-term domestic currency cash loans		480 €

Business rules

- Retrieval of amount
  - BALANCE_OFF-BALANCE
  - MATURITY
  - DEBIT_CREDIT
  - PRINCIPAL_INTEREST_....
- Retrieval of products and contracts
  - N catalogues of products
  - Different granularity
- Parameterisation of accounts, products, contracts

Business rules

- Unification
  - Tens of thousands of accounts
  - Heterogeneous catalogues of products
- Regulatory reporting
  - Hundreds of amount types
  - Several hundred instrument types
- Maturity ladder
  - Manifold hierarchical and associative relationships (accounts, amount types, products, instrument types, contracts, grouped bookkeeping items, algorithm paths, ...)

Development methodologies

- Steps and methods
- Iterative development
- Definition of the business field and terms; design of classes, hierarchies, associative relationships, instances, ontology
- Five methodologies
  - Methontology
  - Methodology by Gruninger and Fox
  - Methodology by Uschold and King
  - Ordnance Survey
  - On-To-Knowledge

Construction of ontology

- Activities preceding conceptualisation
- Non-formal conceptualisation
- Formal conceptualisation

Activities preceding conceptualisation

- Motivating scenario
  - Defined by third parties (management, regulator, owners group)
  - Different levels of detail
    - General set of specifications
    - Non-formal concept descriptions
    - Possible solutions
- Competence questions
  - Which articles enter into the computation of the maturity ladder?
  - Which accounts create the bank's liabilities?
  - Which amount types and instrument types form the bank's claims?
  - Into which time intervals the analytical bank data should be aggregated?
  - What interest amount matures for payment within a week?
  - How many clients are dealing only with liability products?
  - What is the liquidity coefficient on date 31.12.2015?

Formal conceptualisation

- Coding of dictionary and relationships into a formal language
- On-To-Knowledge
  - Middle-out approach
- Database Oracle 12c
  - Supports OWL
- The same instance
  - Ontology
  - Fact and dimension tables
  - Retrieval from both modules within the same query

Dimensional model

- Size
  - Clients dimension – millions of rows
  - Contracts dimension – 5 million rows
  - Statuses fact table – 50 million rows monthly
  - Accounts dimension – 30 thousand rows
  - Destination table of accrued maturity – 100 million rows daily
- Optimisation
  - Partitioning of tables
  - Different merging techniques
  - Decomposing of queries into smaller units
  - Optimisation of ETL procedures scheduling
    - Assignment and scheduling of *n* jobs to *m* processors
      - Flow shop problem of considerable size
    - Non-standard goal function
      - Application of genetic algorithm

Application of the genetic algorithm

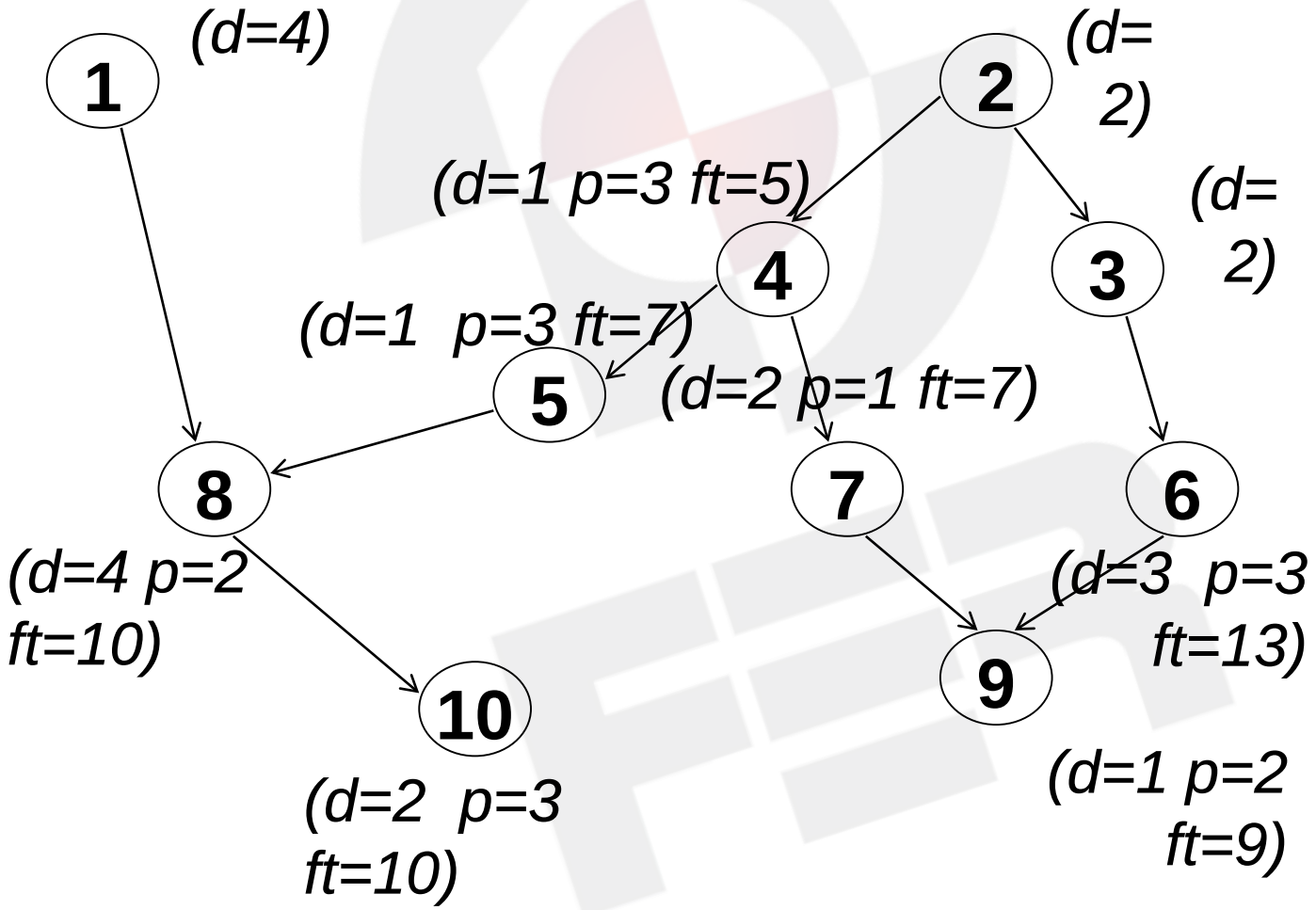
- What is the objective function?
  - Minimum all-over processing time?
  - MiniMax – Minimisation of the longest procedure?
- The objective is to provide maximum reliability for timely indispensable reports to
  - regulatory authorities
  - operational services
- Incomparable in importance to other in-house requirements

Priorities

Priority	Weighting Factor	Description
1	∞	Regulatory authorities; Operational service provisions
2	10	Data used for trend analysis across departments
3	1	Other data

Scheduling of Extract, Transform, and Load (ETL) Procedures

Procedures graph (d is Duration, p is Priority, ft is Required completion time)
Default $ft_i = \infty$ (cannot be delayed; omitted from the fitness calculation)



Conclusion

- Ontology integration into data warehouse
  - increase clarity
  - simplify the programmers' job
- Optimisation of ETL scheduling
  - increase the timeliness of reporting

Reference list reduced to authors' publications

- Vrbanić, Vedran; Kalpić, Damir.
  Scheduling of Extract, Transform, and Load (ETL) Procedures with Genetic Algorithm. // *International Journal of Business Analytics*. **2** (2015) , 3; 33-46
- Andrić, Kristina; Vrbanić, Vedran; Kalpić, Damir.
  BI in practice – developing credit risk reporting module // *32th International Convention Proceedings, Conference: MIPRO 2009*
- Kalpić, Damir; Pavlek, Stjepan; Ujević, Ivana; Vrbanić, Vedran; Ilić, Ivana; Andrić, Kristina; Botički, Ivica; Milašinović, Boris; Pukljak Zoković, Dubravka; Nižetić, Ivana; Katić, Marija; Malović, Vlatko; Budišćak, Ivan.
  Informacijski sustavi za poslovnu inteligenciju u bankarstvu (Information systems for business intelligence in banking), FER, 2008
- Vrbanić, Vedran.
  Ugradnja ontološkog modela u bankarski sustav skladišta podataka (Incorporation of ontology model into banking data warehouse system) / PhD thesis. Zagreb : Faculty of electrical engineering and computing, 17.12. 2015, 129 pages. Mentor: Kalpić, Damir.

- **FOR POSSIBLE QUESTIONS: Vedran Vrbanic [vedran.vrbanic@gmail.com]**