Information Extraction based Monitoring of Intangibles and Risk Measurement

Marcus Spies, Chair of Knowledge Management,
LMU University of Munich

OUTLINE

- PART 1 – Monitoring Intangibles – Background and Approach
- PART 2 – A Pilot Application from the EU MUSING project
- PART 3 – A brief Look at appropriate Information Management Infrastructures – Conclusion

Problem Statement – Monitoring Intangibles

- goal – monitoring based assessment of intangibles
  - "tangibles-based managerial information systems are wholly inadequate for the management of knowledge-based enterprises" (Baruch Lev, Intangibles)
- extend the "real time enterprise" to management of intangibles
  - ✓ up-to-date actionable reporting and accounting
  - ✓ risk management for intangibles and their business impact
- analysis of weakly structured or unstructured data
  - ✓ footprints of intangibles in written reports, comments, posts, chats, process logs, log files from IT infrastructure

Generic analysis approach

- EVENT
  - ✓ register internal or external events relevant to intangibles
  - ✓ capture properties of events using information extraction
- CONDITION
  - ✓ classification of events in terms of suitable taxonomies
  - ✓ firing of appropriate rules, statistical inferences
- ACTION
  - ✓ corrective or improving actions
  - ✓ measures for managing performance
- … extension of processing architecture in business rules engines

Event driven analysis of intangibles

- What are relevant events?
  - ✓ External – an invention threatening your patent portfolio
  - ✓ Internal – loss of a customer
- Events can originate in core, management or support processes
- only a limited subset of these events becomes visible in conventional EPR data
  - ✓ event subscription mechanisms will not work
- ➔ need a method for capturing events from different data sources
  - ✓ we discuss information extraction in part 2

Identifying event impact by matching conditions

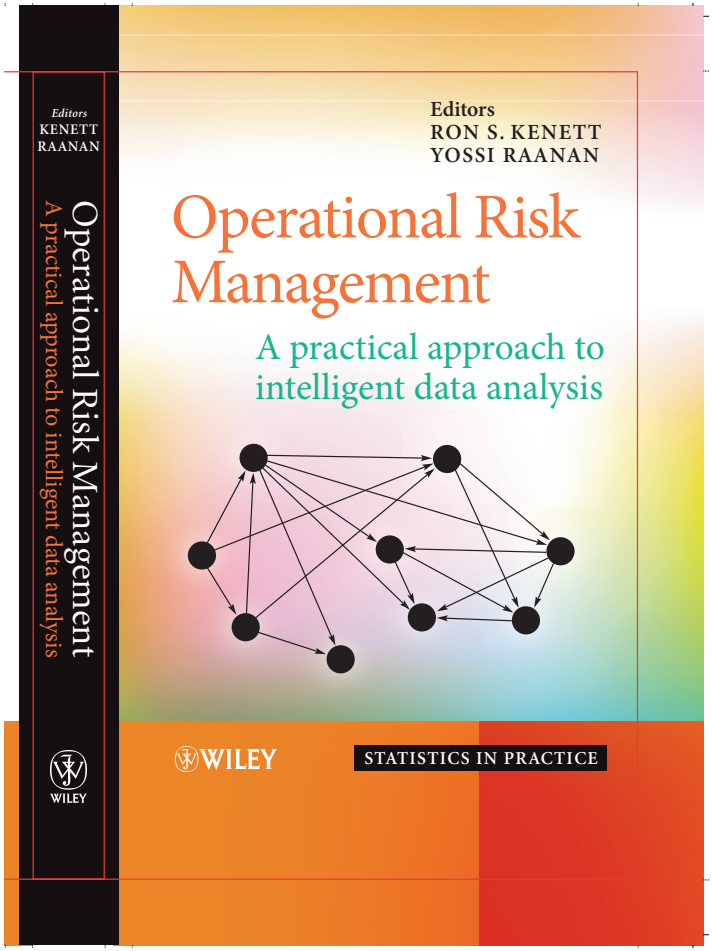
- in common business applications, an event is monitored and acted on by observing components
  - ✓ e.g., component failure → check for needed maintenance action, issue a warning etc.
  - ✓ this has given rise to *event driven architectures*
- properties of events impacting *intangibles* must be evaluated against items in suitable taxonomies
  - ✓ some degree of semantics based or rule based processing is needed
  - ✓ this is where reporting and accounting methods come in
- event properties influence action by evaluating them against a changing set of business rules
  - ✓ e.g., regulatory compliance rules / customer issues

Taking Actions on the basis of identified intangibles issues

- Basis for our approach – the eXtensible Business Reporting Language XBRL
  - ✓ recent extensions to XBRL of high relevance to monitoring intangibles
    - WICI – XBRL for intellectual capital, see contribution by Amy Pawlicki to this symposium
    - GRC-XML – XBRL for governance and regulatory compliance
    - both extensions allow to define KPIs and link control or correction actions to conditions or issues

Operational Risk Management extended to Intangibles

- ✓strong relationship of present work also to operational risk (OpR) management
- ✓focus on operational data driven analysis methods
- ✓focus on risk measurement approach to intangibles assessment
- ✓in OpR, actions are modelled as risk minimizing options, this brings in a decision making perspective



A pilot application – knowledge intensive business analytics

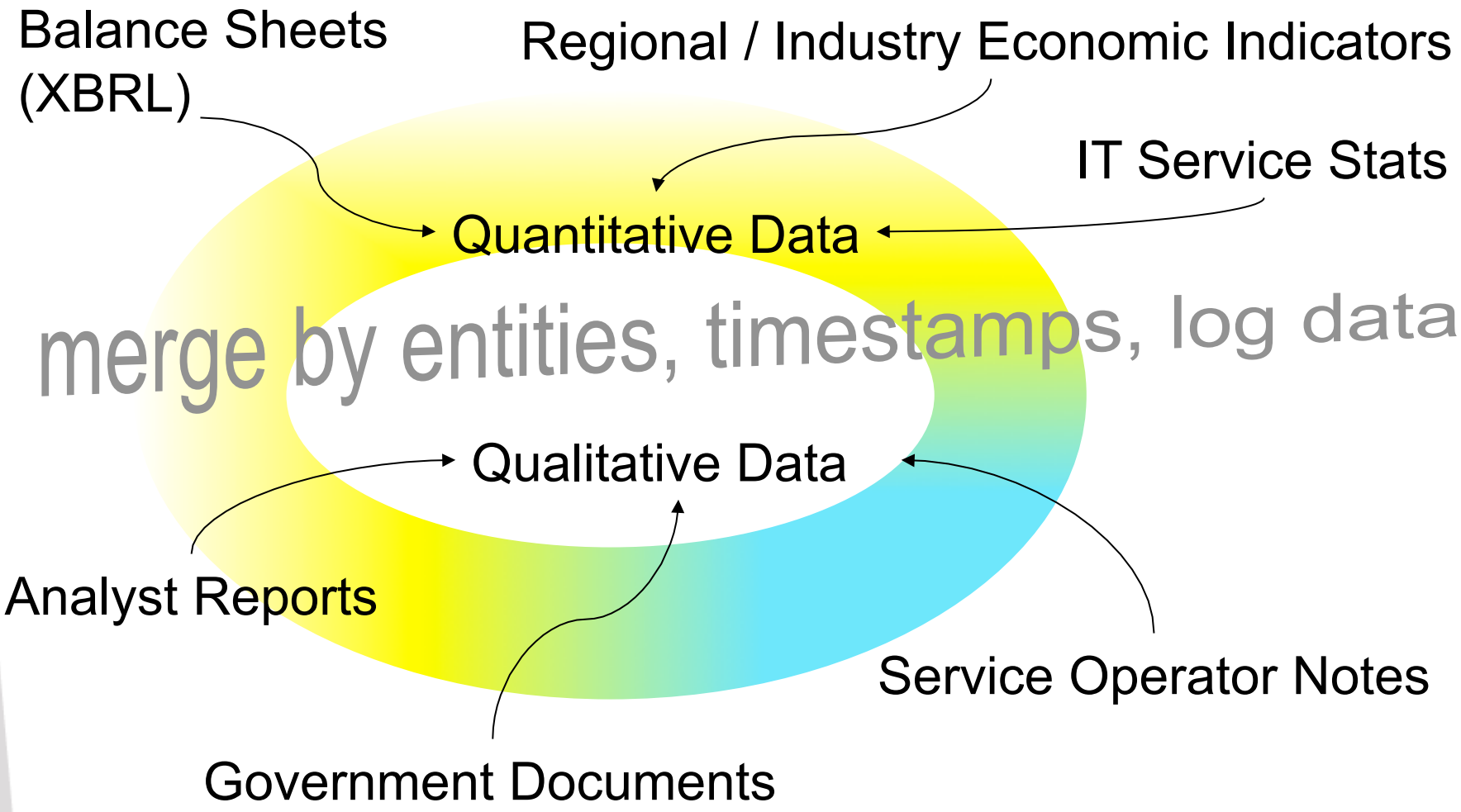
- Result from EU MUSING project
  - ✓ Multi-Industry Semantics Based Next Generation Business Intelligence
    - www.musing.eu
    - April 2006 – April 2010

Next Generation
Business Intelligence

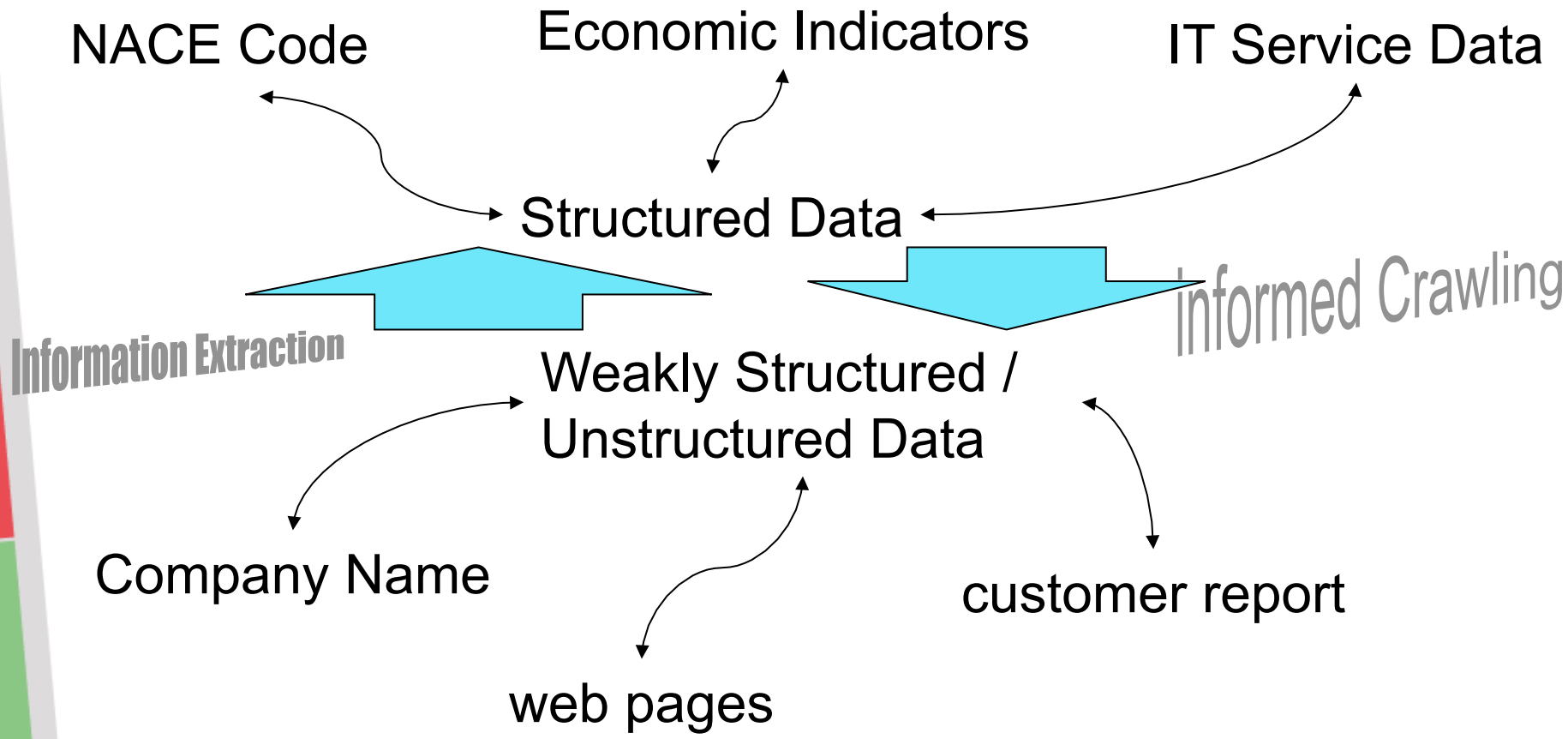
- goal – combine the strengths of artificial and business intelligence
  - ✓ integrate knowledge modelling and statistical inference
  - ✓ blueprint new generation of analytics services

MUSING Partners

Key for MUSING – Leverage the potential of combined qual / quant data



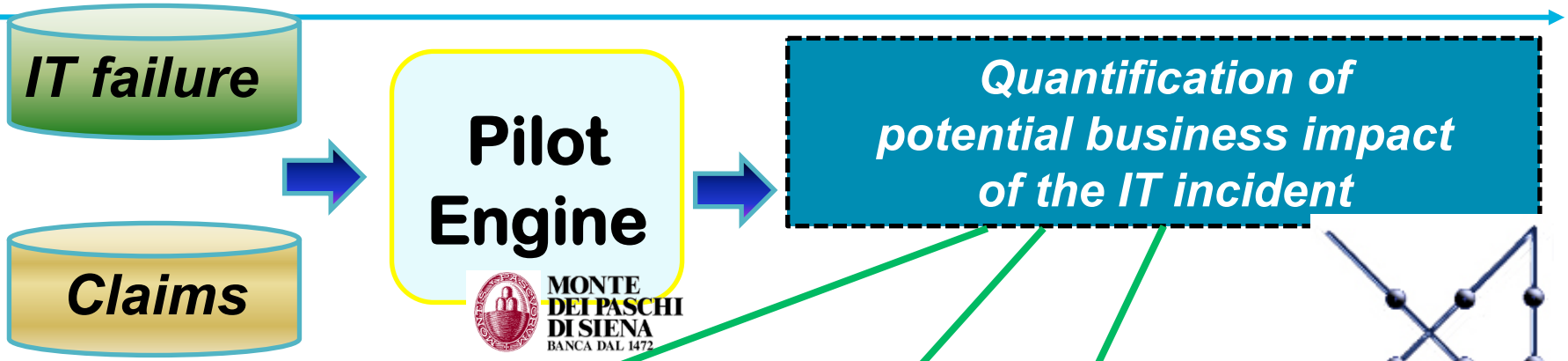
Use weakly or not structured information to extract qual / quant data



A MUSING pilot study – CRM in IT Services

- goal – define KPIs to enable high responsiveness to service performance issues
  - ✓ specific scenario in IT services –
  - ✓ business events affecting intangibles with (often tangible) consequences
    - causal event – hardware failure, network breakdown, software malfunction
    - affected intangibles – customer capital, process capital
    - visible consequences – customer claims, even lawsuits
- intangibles mediate the cause – effect relationship in a non-deterministic way
  - ✓ extending conventional operational risk analysis

Specific MUSING Pilot Objectives



"***multiple loss***" is an OpR event that produces pecuniary losses in several business units (e.g., Transaction Server Crash)

"***opportunity loss***" is an OpR event that gives rise to a potential loss in terms of missed business opportunity (e.g., local application crash)

"***near miss***" is an OpR event that produces neither a multiple nor an opportunity loss by itself but indicates increasing probability of other loss events (e.g. fraudulent resource usage, hacker attacks)

Intangibles involved in IT Services CRM

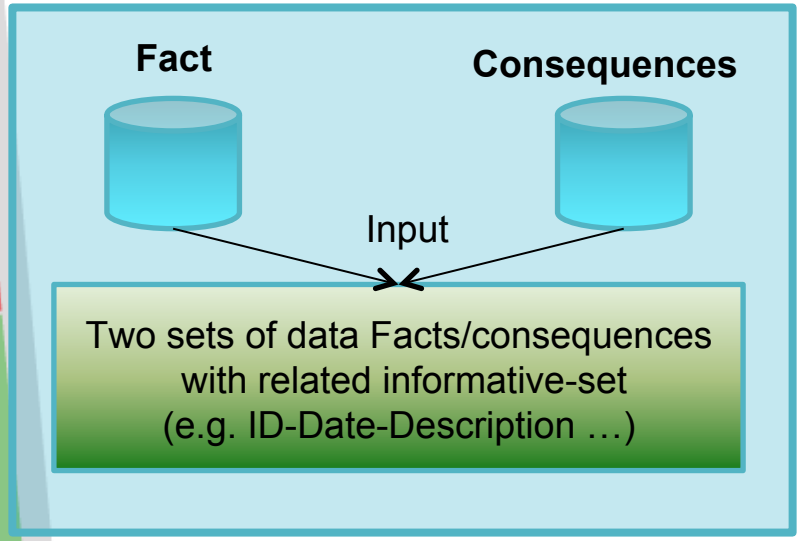
- intangibles related to process capital,
  - ✓ in terms of the WICI taxonomy, risk management and service governance
    - wicijp:InternalControlsStructure
    - wicijp:InternalWarningSystemAndResponse
- intangibles related to customer capital
    - wici-kpi:CustomerServiceCommunicationAndRelationships
    - wici-kpi:ManagingCustomerSatisfaction
    - wici-kpi:ManagingCustomerRetentionLoyaltyAndAdvocacy
    - wici-kpi:ManagingCustomerRevenueGeneration
- need assessments of these intangibles for proactive or protective action

Addressing Intangibles from operational data

- operational event data
  - log files
  - human annotations, explanations
  - ✓ … too early to allow conclusions
- CRM data
  - call center transcripts, online forms, mails
  - ✓ … first footprints of influences on intangibles, but unstructured data
- consequences data
  - claims, lawsuits
  - ✓ … too late for proactive measures, but key to overall improvement on KPIs, again mostly unstructured

Data sources

The data sources of "fact" (**F**) and "consequences" (**C**):



1. multiple losses

F: central IT system logs, operator notes (RelDB incl text)
C: claims and lawsuits against the Bank (RelDB incl text)

2. opportunity loss

F: IT Dept. (MO provider) service logs, operator notes (RelDB incl text)
C: staff reports, customer complaints (RelDB incl text)

3. near misses

F: IT Server log records (invalid login, connect attempts, attacks etc) (RelDB incl text)
C: risk profile obtained from 1. and 2., business process logs (various formats)

Information Extraction by Natural Language Processing

- Input – short texts
  - e.g., failure comments, customer free form comments
- Procedure
  - pipeline of processing steps
    - tokenization
    - stop word elimination
    - matching against domain terminology
    - stemming (lemmatization)
    - document /inverted document term frequency extraction
    - topic analysis
- Output – relevance vector of topics for text

Topic modelling and topic analysis

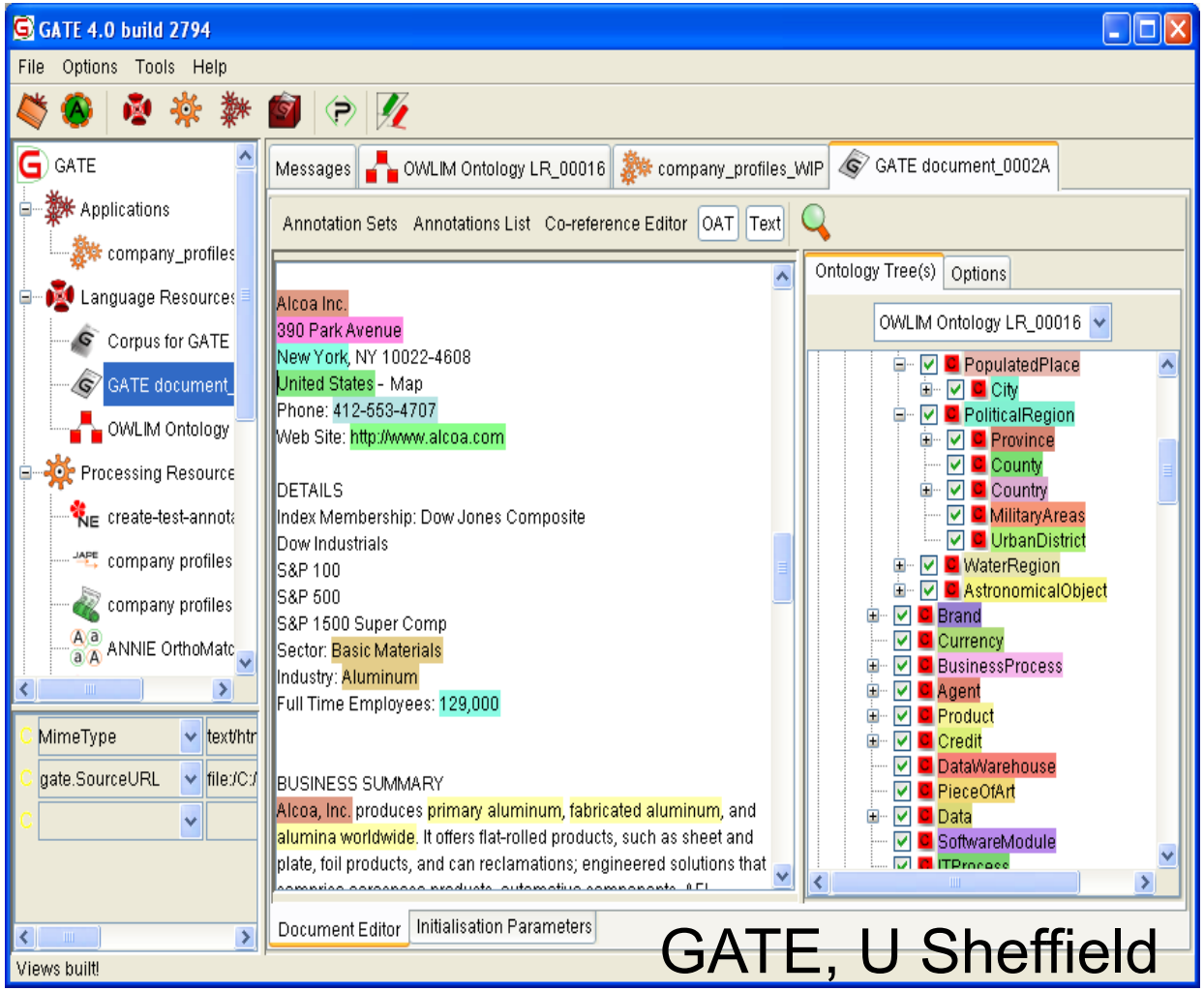
- generic topics
  - ✓ topics are identified in an unsupervised way from co-occurrences of terms
  - ✓ methods e.g. latent semantic analysis
    - based on singular value decomposition of suitable frequency / inverted document frequency matrix
    - used in Apple's SpotLight application
  - ✓ recently, this has been developed further, latent Dirichlet allocation
    - U Stanford Nat. Lang. Group Topic Modelling Tool

Ontologies – knowledge beyond topics

- An ontology is a formal representation of a conceptual system comprising
  - ✓ one or more taxonomies (concept hierarchies)
  - ✓ concept definitions by data and object (has-a) properties
  - ✓ … related to entity relationship models, but based on logic – declarative knowledge representation
  - ✓ benefit – can run inference engines to derive properties of a concept or an individual
  - ✓ can detect inconsistencies or apply rules to enforce requirements

Ontology based information extraction

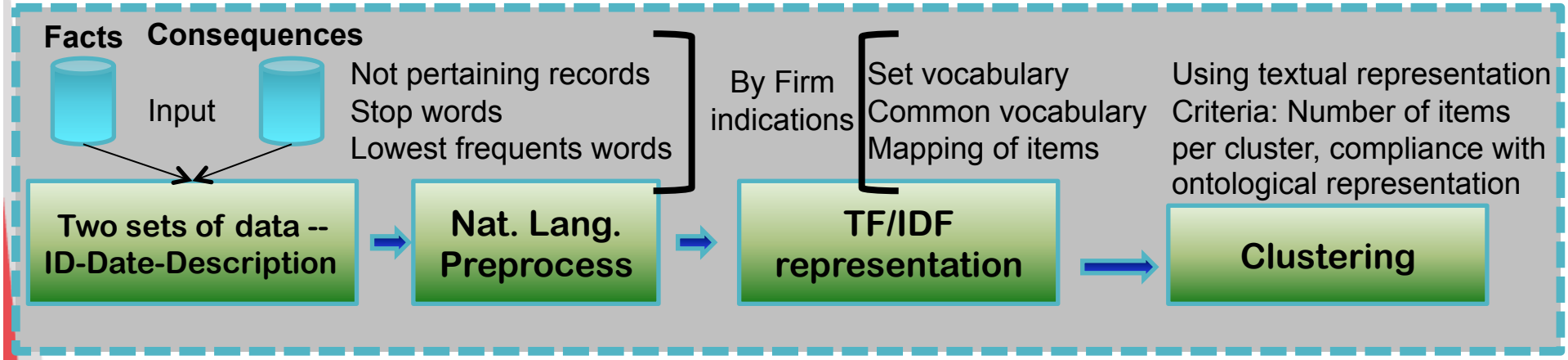
- spot items of qual / quant information in texts, web pages etc
- match information against ontology class instances and their relationships
- populate ontologies repository
  - ✓ OWL – web ontology language
- perform logical inferences
  - ✓ CROWL – developed by DFKI Nat Lang Lab
- query results from ontologies
  - ✓ JOSEKI – web service infrastructure for storing / querying ontology data



Application of Topic Analysis to Intangibles Monitoring

- Identify vocabularies
  - ✓ for loss event descriptions
  - ✓ for CRM textual data
- Use Topic Analysis to assess relevant descriptors for loss events / customer claims
  - ✓ cluster descriptors –
    - what are key loss event groupings in the domain – Bank transactional IT services
    - what are key customer complaint issues, e.g. denial of service for chips on customer cards

Topic Representation of Facts / Consequences Data in the Pilot



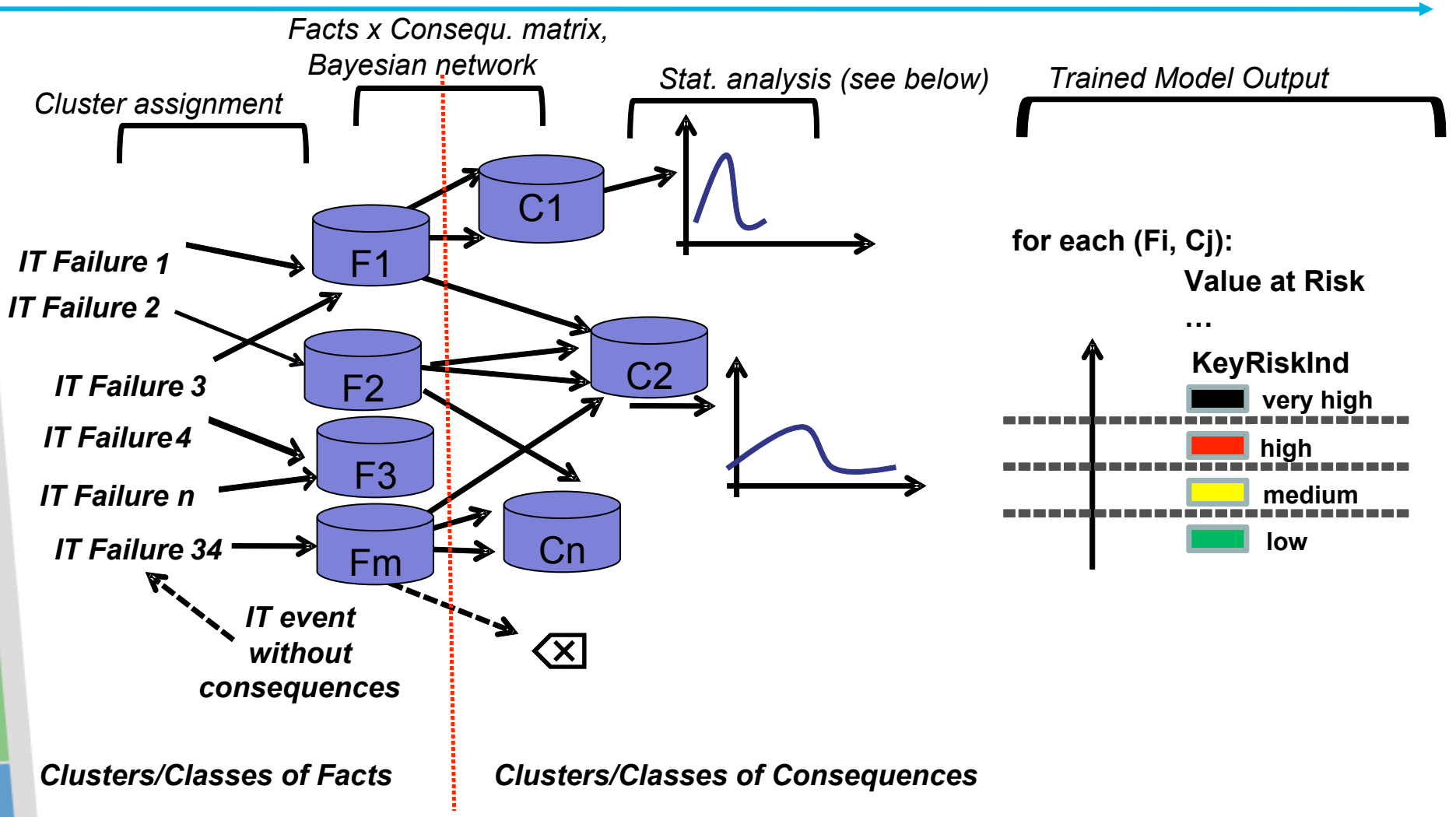
Cluster ID	Failure Topics
1	Check management
2	Contract management, printing activities
3	Hardware
4	Olap and data management
5	Data updating, Data flow, Transaction procedures
6	Corporate banking, remote banking, home banking Server Apps
7	Bancomat, POS, ATM
…	…

Cluster ID	Claim Topics
1	ATM-Bancomat withdrawal
2	Cheques management and fraud
3	Loan management
4	Bank account management
5	Bill and cash order management
6	Credit Card Usage
7	Mortgage management
…	…

Analysis of affected intangibles

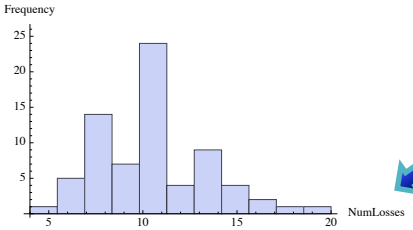
- "correlate" loss descriptions and claims by
  - ✓ temporal proximity
  - ✓ expertise on possible causal relationships
  - ✓ (can use a Bayesian network to build a full probabilistic influence model)
- add a valuation to each loss event and claims cluster
  - ✓ prob. distribution of costs incurred
- ➔the clue to affected intangibles
  - ✓ process capital – a loss event triggers many costly claims
  - ✓ customer capital – a claim can imply customer loss

Method for Risk Model Construction from Training Data Set



Statistical Processing on Training Data

Proximity (Fi ←→ Cj similarity evaluation)

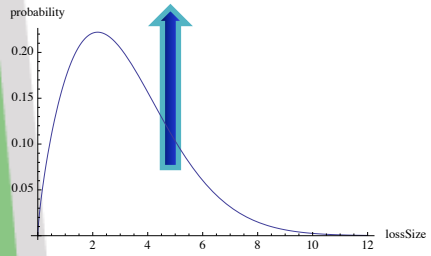


Parametric model of loss events distribution over time

Influence Matrix (Cj | Fi)

Conditional distribution of Consequences Cj for each Fi

Cj Severity Distributions



Severity Distribution Parametric Model

Montecarlo convolution

VaR Evaluation

VaR (Cj | Fi)

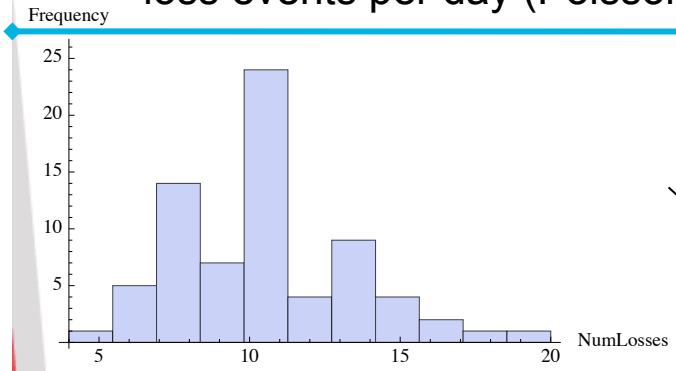
Total VaR F*i*

Applying the Classifier -- Intangibles Monitoring and Assessment

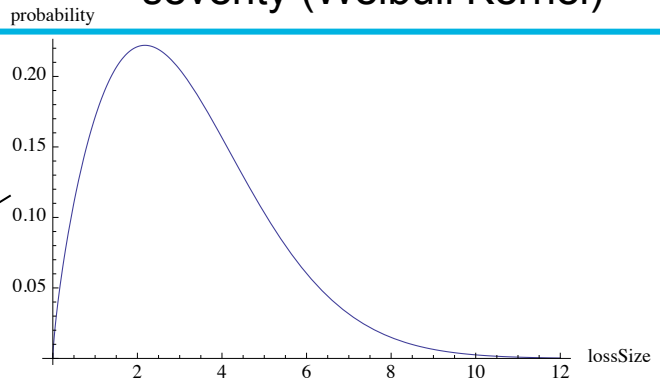
- The MUSING pilot targeted the training phase of the risk classification system
  - ✓ deployment at Bank Monte dei Paschi
- Extending this work, we come to the usage phase of the resulting risk model, in particular focussing on intangibles monitoring
  - ✓ obtain early warnings
    - WICI: building an adequate internal controls structure
  - ✓ analyze log and CRM data for patterns indicating high risks
  - ✓ update the probability distributions of the model (re-training)

An example calculation for test data (using Mathematica 7)

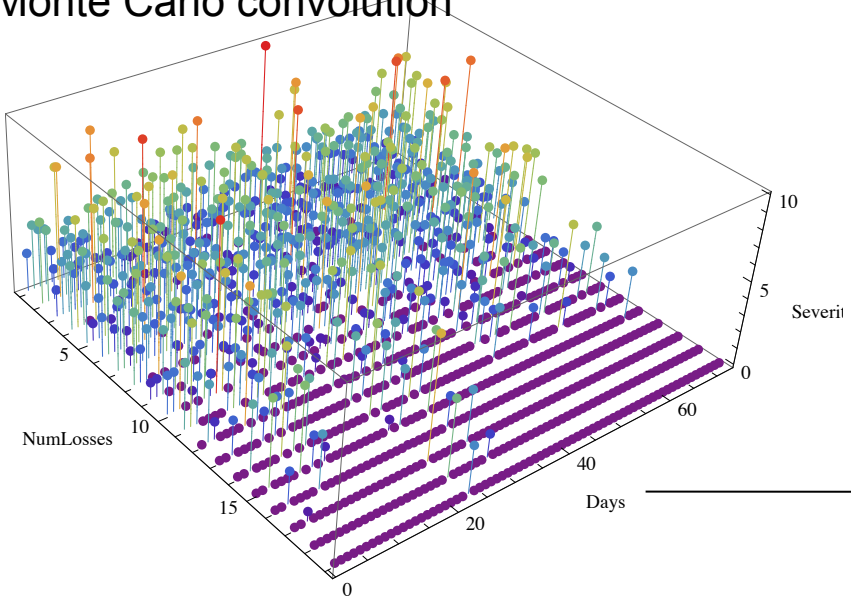
loss events per day (Poisson sample)

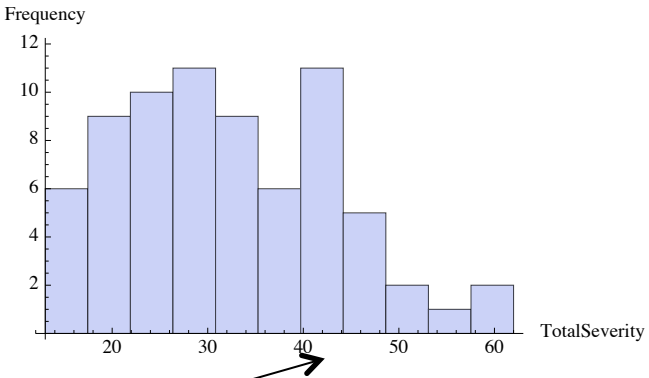


severity (Weibull Kernel)



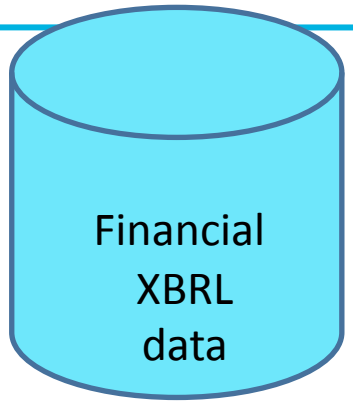
Monte Carlo convolution





resulting Total Severity distribution (for VaR)

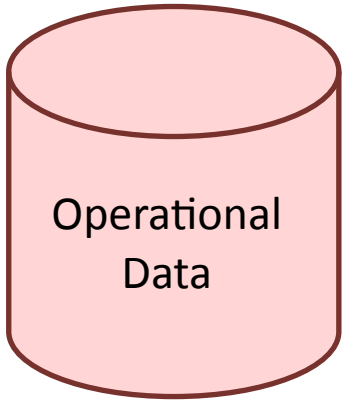
Outlook – Extending the model with financial data



FRM model

Financial risk

$$p_i^* = \delta_i \hat{p}_i + (1 - \delta_i)\tilde{p}_i$$



OPR model

Operational risk

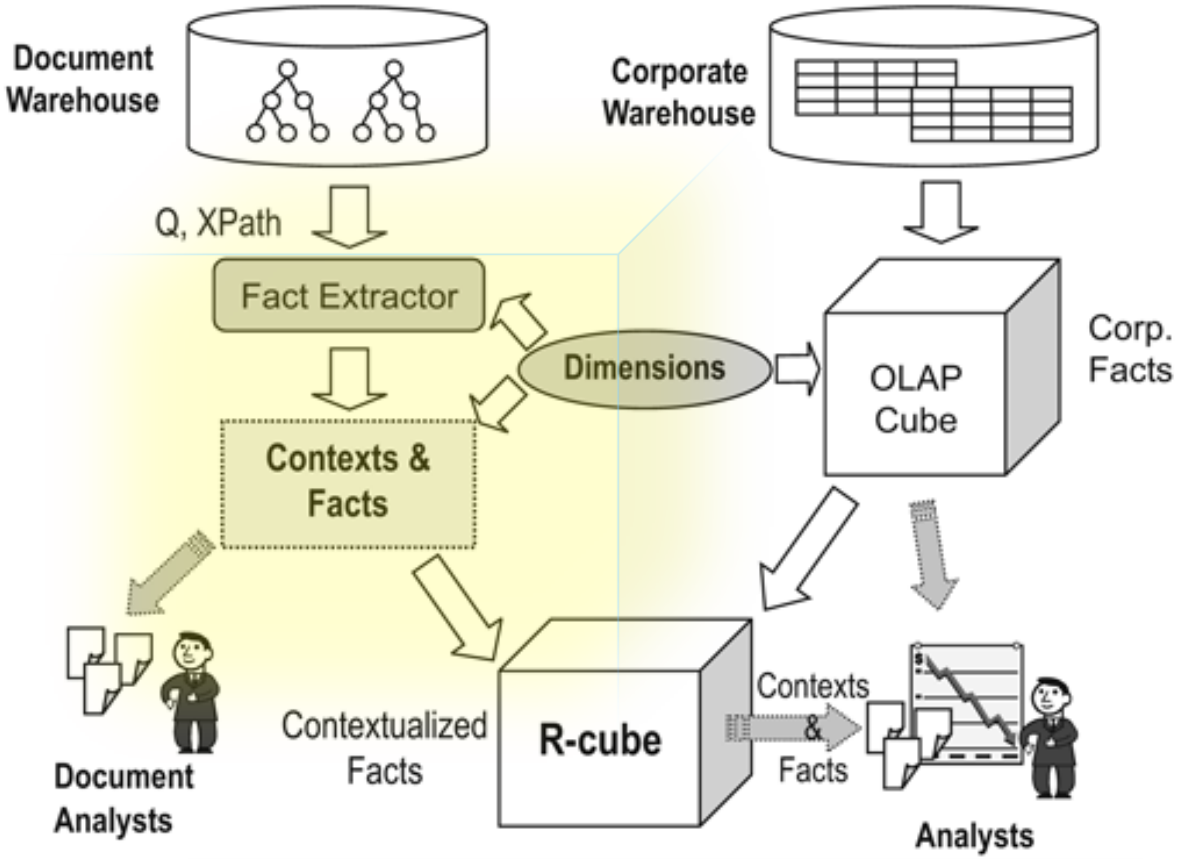
MUSING used this to predict corporate customer behaviour in IT Services, s. research by P. Giudici, S. Figini, U Pavia reported in Operational Risk Management volume

How to assemble a business service based on MUSING technologies

- Web-Apps for business / experts users
- Business layer needs to integrate
  - ✓ NLP modules (like GATE, U Sheffield)
  - ✓ statistical modules (like R scripts)
  - ✓ knowledge warehouse (ontology) querying and updating
    - provided in MUSING by DFKI and U Innsbruck
  - ✓ data access
    - dedicated services like EBR provided balance sheets in XBRL
    - crawling for company imprint pages, region documents
- Implementation
  - ✓ all services are web services w suitable partner links
  - ✓ all applications are WS-BPEL 2 processes (Glassfish 2.2 JBI Server)
  - ✓ this was used for the pilot on Multiple Losses
  - ✓ MUSING integration partner was MetaWare S.p.A. of Pisa, IT

Vision – the contextualized information warehouse architecture

- integration of weakly / un-structured info
- common dimensions language
  - ✓ interoperability with XBRL
- specific methods
  - ✓ document WH
  - ✓ corporate WH
- integrated analytics



Juan Manuel Pérez, Rafael Berlanga, María José Aramburu, Torben Bach Pedersen, "Integrating Data Warehouses with Web Data: A Survey," IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, Vol. 20, 07, pp. 940-955, JULY, 2008. 0

Conclusions (part 1)

- generic approach to intangibles focussed management information / decision support systems (MIS, DSS)
  - focus on up-to-date monitoring and assessment of intangibles needed
  - Event-Condition-Action approach
  - integration of weakly or un-structured information
  - interoperability with XBRL (esp. emerging standards WICI and GRC-XML)
  - in middle to long term perspective, an integration with business rules processing will be needed

Conclusions (part 2)

- an operational risk modelling pilot demonstrates feasibility of the approach
  - ✓ including information extraction from operational data (textual comments, notes etc)
  - ✓ intangibles mediating cause – effect relationships
    - we predict effects resulting in losses
    - reversing the signs, the method can be applied to gains, as well
      - e.g. in analysis of collaborative networks infrastructures
  - ✓ resulting KPIs and activity / risk controls assessments can be brought in line with XBRL WICI