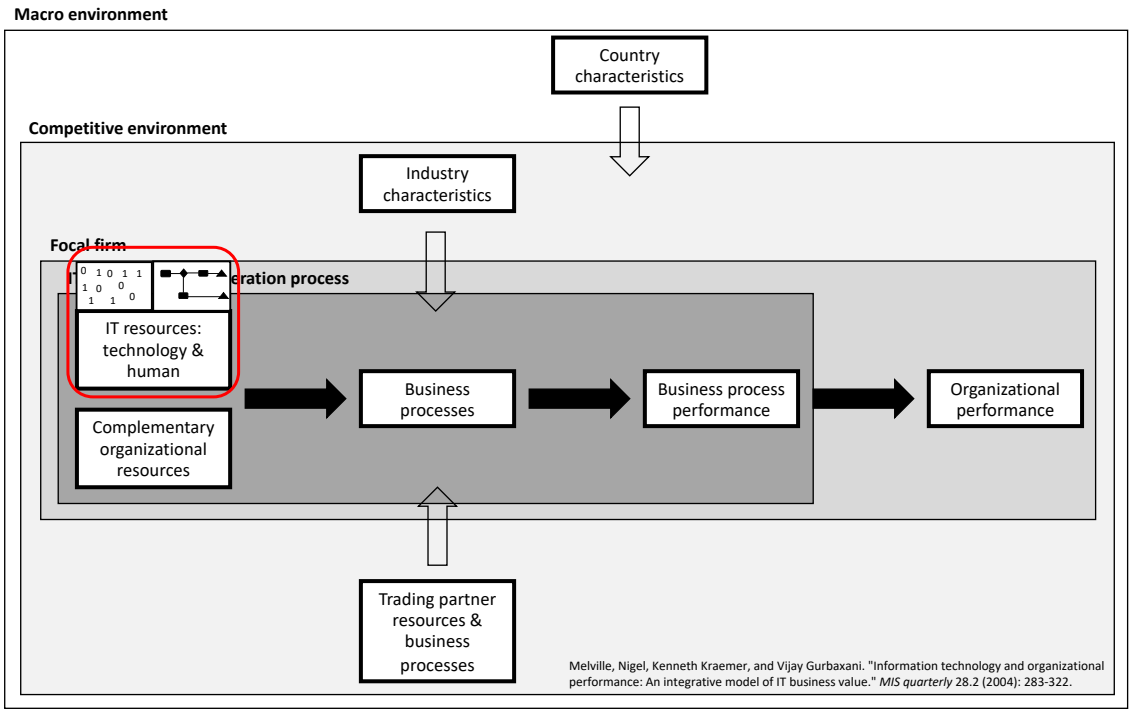
Strategic IT management - 37E00200

Information infrastructures and data driven decision making

Esko Penttinen
Associate Professor, Information Systems Science, Aalto University School of Business
Chairman, XBRL Finland
Director, Real-Time Economy Competence Center

1



2

Agenda

1. Defining data
2. Generating value from data
3. Information infrastructures and case XBRL (eXtensible Business Reporting Language)

3

1. Defining data

4

Data as the new oil

Regulating the internet giants

The world's most valuable resource is no longer oil, but data

The data economy demands a new approach to antitrust rules

The Economist 6.5.2017

APR 13, 2017 @ 02:22 PM 5,459 The Little Black Book of Billionaire Secrets

What Will We Do When The World's Data Hits 163 Zettabytes In 2025?

Forbes 13.4.2017

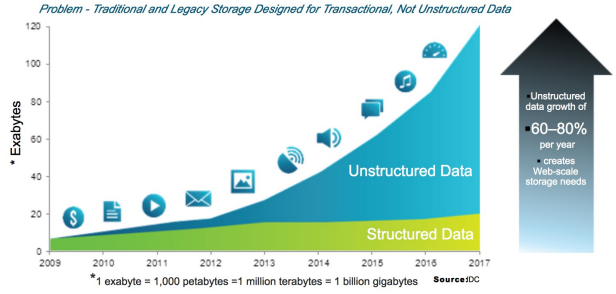
Structured vs. Unstructured Data

Structured data is far easier for Big Data programs to digest, while the myriad formats of unstructured data creates a greater challenge. Yet both types of data play a key role in effective data analysis.

Datamation 3.8.2017

Data Growth

Problem - Traditional and Legacy Storage Designed for Transactional, Not Unstructured Data



5

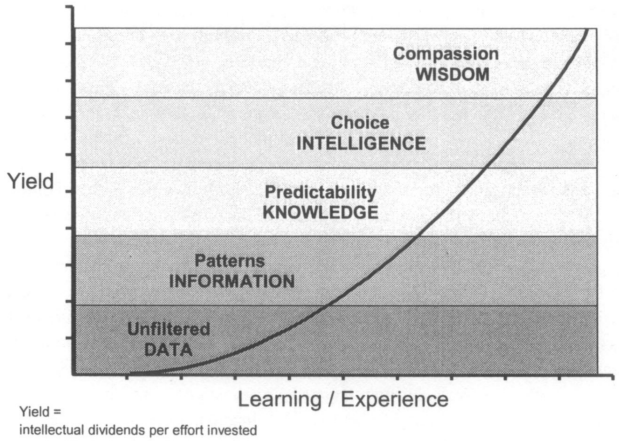
Defining data

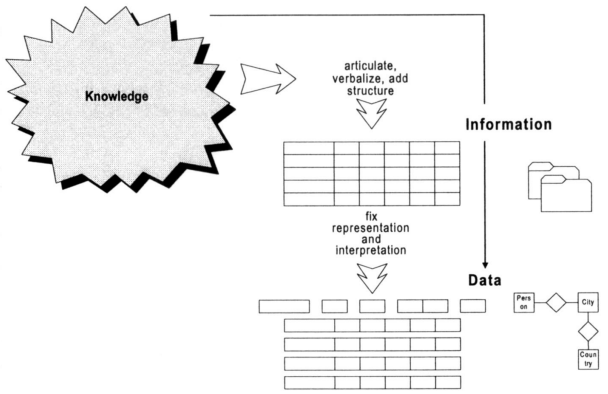
- Etymology: latin word *data* is the plural of *datum*, "(thing) given"
- In computational systems, data are the coded invariances (an expression whose value does not change during program execution)
- In human discourse, data are that which is stated
- Data → information → knowledge
  - Data often assumed to be the least abstract concept
  - Tuomi (1999) challenged this view by claiming that raw data do not exist; rather data emerge last – only after information and knowledge are available

Tuomi, Ilkka. "Data is more than knowledge: Implications of the reversed knowledge hierarchy for knowledge management and organizational memory." *Journal of Management Information Systems* 16.3 (1999): 103-103.

6

Contradicting views on data-knowledge relationship





Tuomi, Ilkka. "Data is more than knowledge: Implications of the reversed knowledge hierarchy for knowledge management and organizational memory." *Journal of Management Information Systems* 16.3 (1999): 103-103.

7

Types of data

- Unstructured data
- Semi-structured data
- Structured data

	Concept(s)	Definition	Example from accounting
Levels of structuredness	Unstructured data	A mishmash of semantic entities that can differ from an observation to another; it may not always be clear what constitutes an individual observation	Data residing in a note written with a text editor to be refined into a receipt to be booked into an accounting information system
	Semi-structured data	Data organized using an irregular or unstable data structure which hampers the usability and interoperability of the data	Data residing in an electronic sales invoice adhering to a proprietary XML-format that needs to be converted to the XML-format of electronic purchase invoices
	Structured data	Data residing, for instance, in a database under a rigid and regular structure with well-defined fields that correspond to distinct variables	Company's financial statement stored in a standardized, taxonomy-compliant XBRL instance document
Types of data	Bitstring	Series of binary distinctions encoded into a material medium	Magnetic marks on a hard disk platter
	Data token, raw unorganized facts, invariances	Data token refers to the most granular element of data; also called invariances as they remain unchanged when a specific transformation is applied	"Deferred Tax Assets, Net" in an XBRL instance document containing a company's financial statements
	Data object	Aggregated or computed entity made out of data tokens	Key financial figure computed using data tokens such as return on capital employed
	Metadata	Data that provide information about other data	Metadata in an XBRL instance document (e.g., currency, periodicity, and credit/debit status of Deferred Tax Assets, Net)
	Data model or schema	Definition of the organization of data; articulates allowed data tokens and their attributes, and specifies the possible relationships between them	XBRL taxonomy (e.g., US GAAP XBRL taxonomy for financial statements)
	Data source	A location from where the data being used originates	Relational database (e.g., the EDGAR repository for US GAAP XBRL financial statements)

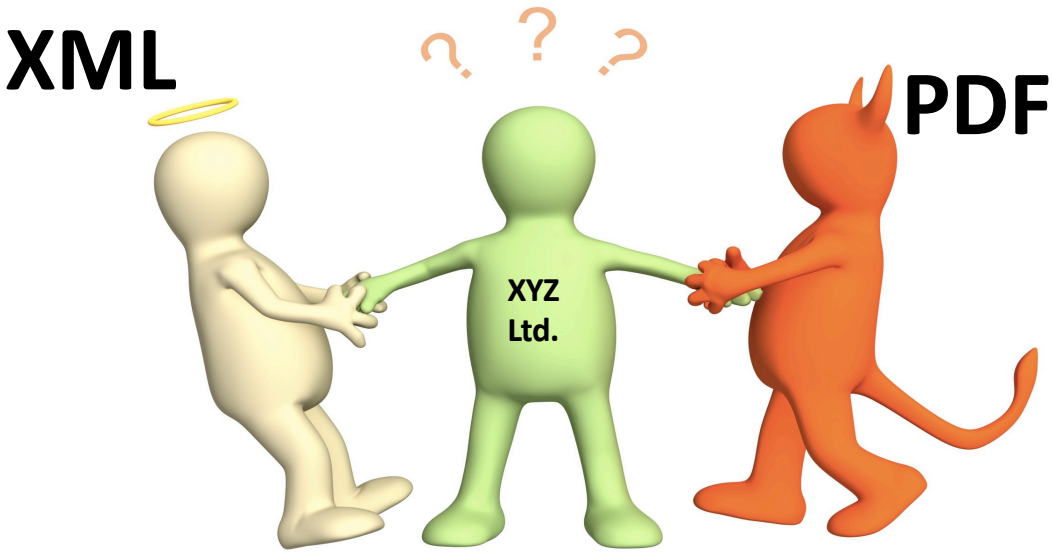
Aaltonen, A. & Penttinen, E. (2021) What Makes Data Possible? A Sociotechnical View on Structured Data Innovations, *54th Hawaii International Conference on System Sciences (HICSS)*, 2021.

8

Structured data

- Characteristics of structured data
  - Semantic meaning of data elements
    - What the data actually means/measures
    - Data type (numeric, currency, alphabetic, name, date, address)
  - Data model
    - Conceptual model (e.g. SBR/XBRL taxonomy structure) -> logical model (e.g. XBRL taxonomy with linkbases) -> physical schema (e.g. XBRL data elements)
  - Easily entered, stored, queried and analyzed
    - Machine-readability
- Examples of structuring data
  - XML-based e-invoicing formats
  - XBRL-based financial reporting
  - Dartfish-tagged football match
  - iPhone people album

9

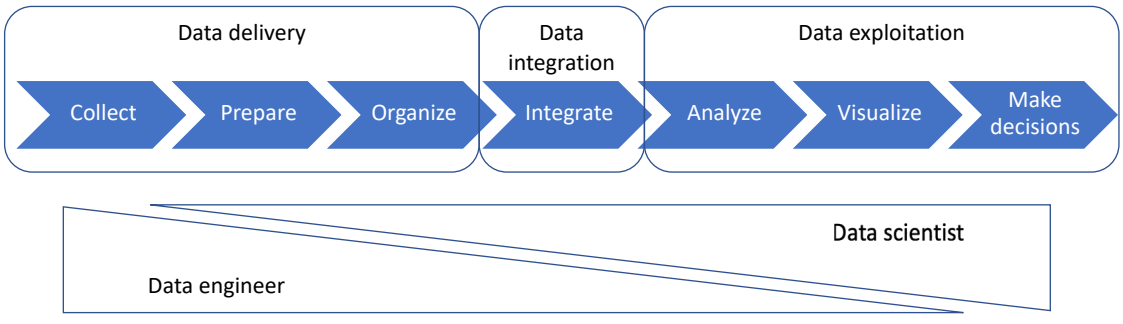


11

2. Generating value from data

12

Steps towards successful data exploitation



Source: CEMS KONE case group 2016

13

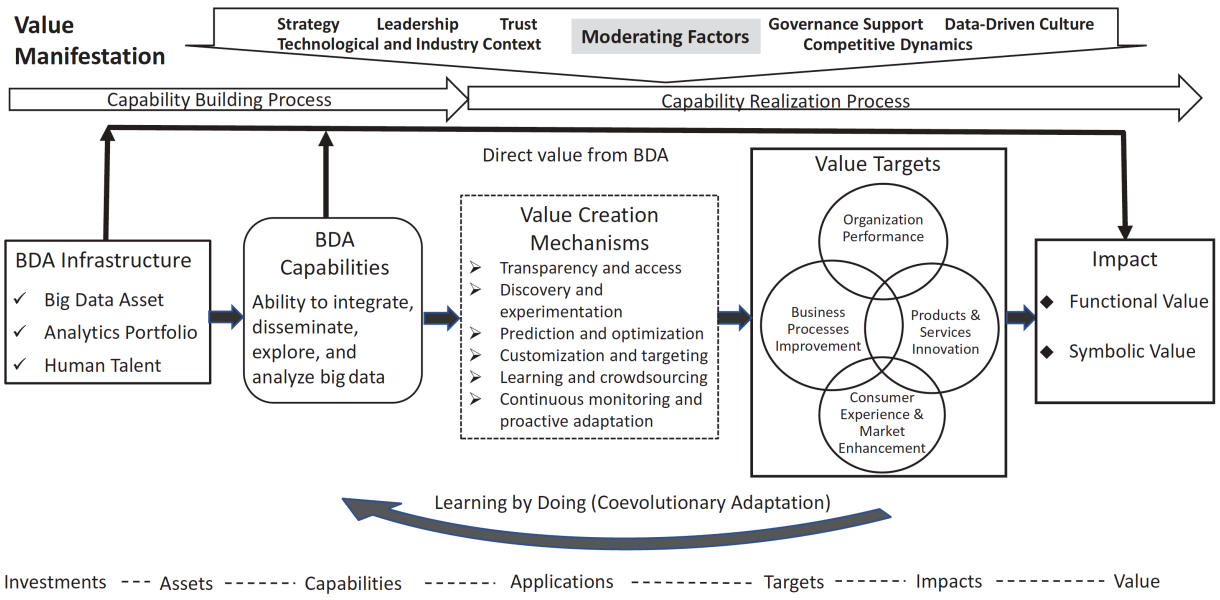


Figure 3. Value Creation by Big Data Analytics

Grover, Chiang, Liang & Zhang (2018). Creating Strategic Business Value from Big Data Analytics: A Research Framework, Journal of Management Information Systems, 35:2, 388-423

14

(cont.) Value creation mechanisms

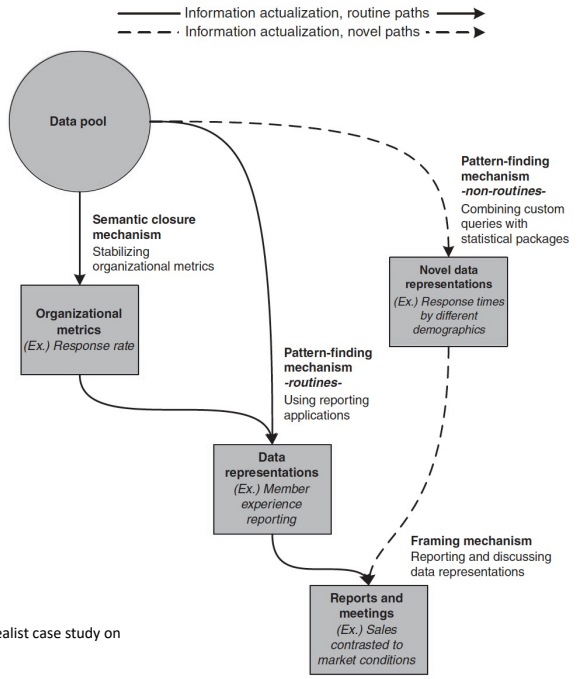
- Transparency and access
  - The ability to generate descriptive data and disseminate them widely across a firm not only allows consistency in viewing the data but also facilitates a more complete visibility of the firm's business processes and outcomes. For instance, dashboards can provide real-time access to the wellbeing of various activity systems in a firm.
- Discovery and experimentation
  - Discovery is often the most emphasized aspect of BDA—digging into data for both deep and pragmatic insights can yield important outcomes for various BDA targets. Similarly, in an increasingly digital world, big data can involve many small experiments. For instance, running longitudinal experiments can provide insights into causality that may have strong implications for the delivery of customer services, among other targets.
- Prediction and optimization
  - In the former, the value is in determining probabilistic outcomes for the future on which present-day action can be taken. In the latter, leveraging big data and powerful analytics can determine the "best path" forward—something that was largely suboptimized and dependent on managerial judgment in the recent past.
- Customization and targeting
  - BDA can facilitate customization of products and services, as well as targeting different market segments with digitally versioned products. Such mechanisms greatly bolster customer retention and other customer-related outcomes.
- Learning and crowdsourcing
  - Machine learning has been applied to many different contexts from online education to automated vehicles, while crowdsourcing is being used for predictions and leveraging innovative talent.
- Continuous monitoring and proactive adaptation
  - The ability to monitor situations and adapt rapidly allows preemption of future problems. This is particularly prevalent with the explosion of data from the Internet of things, which can be used to monitor, warn, and adjust for situational abnormalities.

Grover, Chiang, Liang & Zhang (2018). Creating Strategic Business Value from Big Data Analytics: A Research Framework, Journal of Management Information Systems, 35:2, 388-423

15

Mechanisms for creating value from data

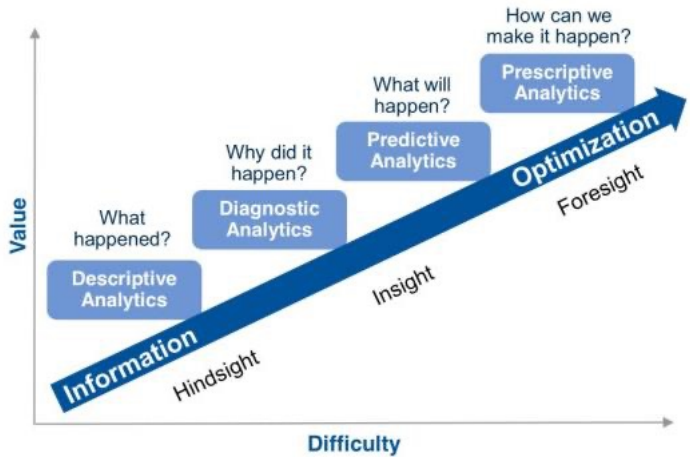
- ”Semantic closure mechanism” means a stable way to interpret the data for a specific purpose
- Human operator can query, tabulate and visualize the data, thus activating the ”pattern finding mechanism”
  - Routines vs. experiments
- ”Framing mechanism” means the way in which the metrics and patterns observed in the data are brought to bear upon daily operations, sensemaking



Aaltonen, A., & Tempini, N. (2014). Everything counts in large amounts: a critical realist case study on data-based production. *Journal of Information Technology*, *29*(1), 97-110.

16

From descriptive to prescriptive analytics



http://www.rosebt.com/blog/descriptive-diagnostic-predictive-prescriptive-analytics

17

Experimentation vs. approved business case

- Experimentation
  - Case XBRL Denmark: ”We have lots of XBRL data, let’s use it to optimize something!”
- Approved business case
  - ”We need to optimize our weekly production plan for Q3, what kinds of data do we need?”

18

Data replacing human judgment?

- Data is poor at…
  - … deciding upon choice of type of approach (e.g. decision trees, rules-based expert systems, neural networks)
  - … setting objectives and trade-offs between conflicting objectives
  - … placing penalties for falce positives and false negatives in predictive analytics
  - … conducting sanity checks of algorithm output

→ Some level of human judgement will always be required!

19

3. Information infrastructures and Case XBRL

20

Information infrastructure

- Information infrastructure is a
  - *shared* (i.e., not the sole property of one organization but is often shared among multiple organizations or stakeholders),
  - *evolving* (i.e., dynamic and continually evolving to meet changing organizational needs and technological advancements),
  - *heterogeneous* (i.e., consist of diverse components, technologies, and standards that are interconnected)
  - *installed base of IT capabilities* (i.e., represent the existing technological infrastructure upon which organizations build their IT capabilities)
  - that organizations employ to coordinate, control, and compete within and across organizational contexts (Monteiro & Hanseth, 1996).

21

Examples of information infrastructures

- Internet
- Land and property registration (e.g., Residential and Commercial Property Information System in Finland)
- XBRL-based financial reporting

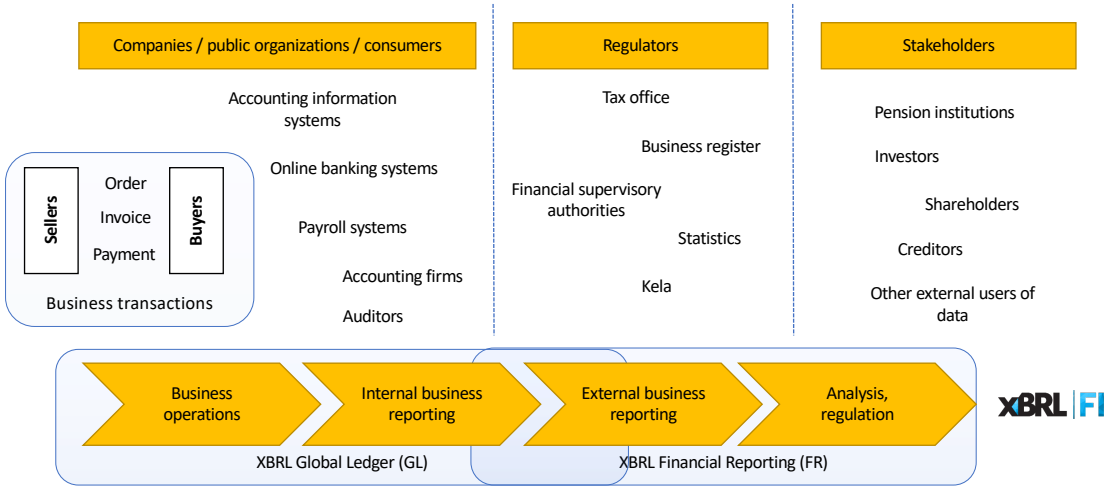
22

Paradoxical tensions in information infrastructures

- Stability vs. flexibility
  - Stability in information infrastructures is often achieved through standardization which can promote efficiency and interoperability, but, at the same time, it may stifle innovation and flexibility by imposing constraints.
- Privacy vs. openness
  - Openness in information infrastructures can promote collaboration and innovation, but it can also introduce security vulnerabilities.
- Legacy systems vs. modernization
  - Many organizations rely on legacy systems that are deeply ingrained in their information infrastructures. Balancing the need to maintain and integrate legacy systems with the desire to modernize can create tensions.
- Centralize vs. decentralize
  - Control over information infrastructures can be a source of tension. Centralization may lead to efficiency, but decentralization can foster innovation and responsiveness.
- Global vs. local
  - Information infrastructures may need to function in global contexts while respecting local regulations and cultural norms.

23

Financial value chain



24

What is XBRL?

CURRENT ASSETS	
Assets Held for Sale	100,000
Construction in Progress, Current	100,000
Inventories	100,000
Construction in Progress, Current	100,000
Hedging Instruments, Current [Asset]	100,000
Current Tax Receivables	100,000
Trade and Other Receivables, Net, Current	100,000
Prepayments, Current	100,000
Cash and Cash Equivalents	100,000
Other Assets, Current	100,000
Current assets, Total	1,000,000

```
<ifrs-gp:AssetsHeldSale contextRef="Current_AsOf" unitRef="U-Euros"
  decimals="0">100000</ifrs-gp:AssetsHeldSale>
<ifrs-gp:ConstructionProgressCurrent contextRef="Current_AsOf"
  unitRef="U-Euros" decimals="0">100000</ifrs-
  gp:ConstructionProgressCurrent>
<ifrs-gp:Inventories contextRef="Current_AsOf" unitRef="U-Euros"
  decimals="0">100000</ifrs-gp:Inventories>
<ifrs-gp:OtherFinancialAssetsCurrent contextRef="Current_AsOf"
  unitRef="U-Euros" decimals="0">100000</ifrs-
  gp:OtherFinancialAssetsCurrent>
<ifrs-gp:HedgingInstrumentsCurrentAsset contextRef="Current_AsOf"
  unitRef="U-Euros" decimals="0">100000</ifrs-
  gp:HedgingInstrumentsCurrentAsset>
<ifrs-gp:CurrentTaxReceivables contextRef="Current_AsOf" unitRef="U-
  Euros" decimals="0">100000</ifrs-gp:CurrentTaxReceivables>
<ifrs-gp:TradeOtherReceivablesNetCurrent contextRef="Current_AsOf"
  unitRef="U-Euros" decimals="0">100000</ifrs-
  gp:TradeOtherReceivablesNetCurrent>
<ifrs-gp:PrepaymentsCurrent contextRef="Current_AsOf" unitRef="U-Euros"
  decimals="0">100000</ifrs-gp:PrepaymentsCurrent>
<ifrs-gp:CashCashEquivalents contextRef="Current_AsOf" unitRef="U-
  Euros" decimals="0">100000</ifrs-gp:CashCashEquivalents>
<ifrs-gp:OtherAssetsCurrent contextRef="Current_AsOf" unitRef="U-Euros"
  decimals="0">100000</ifrs-gp:OtherAssetsCurrent>
<ifrs-gp:AssetsCurrentTotal contextRef="Current_AsOf" unitRef="U-Euros"
  decimals="0">1000000</ifrs-gp:AssetsCurrentTotal>
```

Source:
https://infocenter.informationbuilders.com/wf8005/index.jsp?topic=%2Fpubdocs%2FAdapter_Admin%2Fsource%2Ftopic504.htm

25

XBRL explained

https://www.youtube.com/watch?v=YlgTN2k375s&nohtml5=False

https://www.youtube.com/watch?v=YIjWVAh42Vk

https://www.youtube.com/watch?v=zTyhOvE79DQ

26

XBRL basics

- eXtensible Business Reporting Language provides an identifying tag for each individual item of data (instead of treating financial information as a block of text - as in a standard internet page or a printed document)
  - XBRL allows information modeling and the expression of semantic meaning commonly required in business reporting
  - For example, company net profit has its own unique tag
- XBRL tags enable automated processing of business information by computer software, cutting out laborious and costly processes of manual re-entry and comparison
- Computers can treat XBRL data “intelligently”: they can recognise the information in an XBRL document, select it, analyse it, store it, exchange it with other computers and present it automatically in a variety of ways for users
  - XBRL greatly increases the speed of handling of financial data, reduces the chance of error and permits automatic checking of information.
- XBRL (eXtensible Business Reporting Language) is a freely available, market-driven, open, and global standard for exchanging business information
- XBRL is XML-based and the XBRL Specification is developed and published by XBRL International, Inc. (XII)

27

XBRL taxonomy

- Taxonomy schema
  - Contains data elements (or concepts in XBRL terms)
- Linkbases
  - Definition linkbase
  - Label linkbase
  - Calculation linkbase
  - Reference linkbase
  - Presentation linkbase
  - Formula linkbase

28

XBRL in Finland

- Development driven by the XBRL Finland consortium
- Three large-scale XBRL-reporting programs
  - ESMA mandate on publicly listed companies starting financial statements from 2020 (the ESEF-XBRL financial reporting requirement)
    - https://www.finanssivalvonta.fi/paaomamarkkinat/liikkeeseenlaskijat-ja-sijoittajat/esef_xbrl/
  - SME voluntary reporting of financial statements to business register started in 2019
    - https://www.prh.fi/fi/uutislistaus/2019/P_18080.html
  - Financial reporting from municipalities to State Treasury (Valtiokonttori) in XBRL-format in 2021
    - https://kuntatietoluotsi.fi/mista-on-kyse/

29

XML vs. XBRL

- Both XML and XBRL attach a semantic meaning of the data being transmitted, in XBRL, however, this semantic meaning is defined in standardized format
- XBRL distinguishes between data elements and their relationships. The relationships are defined in separate linkbases, which makes software development is easier.
- XML contains only hierarchical relationships between the data elements. In XBRL, many relationships can be defined with the help of the linkbases in a more exact way than in XML.
- XBRL linkbases (formula, calculation) enable instance validation.
- Reference linkbase can be used to include references to accounting law, which, again, helps in software development.

30

Case Danish Business Authority (DBA)

- One of the pioneers in using structured data in financial reporting, DBA established its machine learning lab in 2017
- Danish Business Authority (DBA) is a Danish government unit with regulatory obligations related to supervision of Danish companies and fraud prevention
- Extensive use of structured data (as opposed to paper and PDF) in financial statements (XBRL) has paved the way for data analytics at DBA
- Numerous machine-learning projects on-going at DBA

31

ML projects at the DBA

Project name	Project description (use case within the DBA and end users)	Purpose	Input	Output	Model and tool
Auditor's Statement	The Auditor's Statement model speeds up verification that the valuations of company assets given in an auditor's statement are correct and that the statement does not feature violations. The algorithm is used by internal DBA case workers.	Prevent misreporting of company assets	Text from auditor's statements that presents asset valuations	Probability of violations in asset valuations	Random forest, bag of words
Bankruptcy	The Bankruptcy model predicts company distress and insolvency and ties in with the Early Warning Europe (EWE) initiative. The algorithm is used not at the DBA but by external consultants in the EWE community in Denmark and in the European Union. The DBA is not responsible for actions and consequences related to the tool.	Identify companies in distress, to enable timely intervention	Data from the business registry and annual financial reports	Probability of bankruptcy	Scikit-learn, gradient boosting
Company Registration	The Company Registration model is aimed at detecting fraud-indicating behavior among newly registered Danish companies. The algorithm is used by internal DBA case workers.	Prevent abusing incorporation to commit fraud	Data from the business registry, annual reports, and VAT reports	Probability of fraudulent behavior	XGBoost
Land and Buildings	The Land and Buildings model predicts violations of accounting policies related to property holdings and long-term investments. The algorithm is used by internal DBA domain experts.	Prevent violations of accounting policy	Text about accounting policies, from the auditor's statement	Probability of violations of accounting policies	Random forest, bag of words
ID Verification	The ID Verification model expedites processing of the documents submitted, by supplying a text string from the machine-readable portion of an ID and comparing it against input data from the user. The algorithm is used by internal DBA case workers.	Facilitate processing of documents	Pictures of IDs submitted to the DBA	JSON string with text from the machine-readable portion of the ID	PassportEye
Recommendation	The Recommendation model improves the user experience of the DBA's virk.dk online portal by focusing on personalized content and optimized interfaces. The algorithm improves the portal's usability for external customers (end users).	Improve usability of the online portal	Telemetry data from virk.dk	Recommendation of relevant content	TBD
Sector Code	The Sector Code model speeds up verifying a company's industry-sector code. At present, 25% of the company codes are incorrect. The algorithm is used by internal DBA case workers.	Prevent misreporting of industry-sector codes	Activity-description text from a company's annual statements	Probability distribution over the set of sector codes	Neural network
Signature	The Signature model, in combination with the associated document filter, speeds up verification of whether a company-establishment document is signed or not. The algorithm, used by internal DBA case workers, returns three probabilities: of whether the document is physically signed, whether it is digitally signed, and whether the signature is missing.	Facilitate the process of establishing a company	An image of a company-establishment document	Probability of whether a document is signed or not	Neural network (ResNet 16)

32