



Ontology-based Multilingual Access to Financial Reports for Sharing Business Knowledge across Europe

Presented for the
Monnet Consortium by
Thierry Declerck, DFKI GmbH



Partners

•Industrial Organizations

- SAP, Germany (<http://www.sap.com/index.epx>)
- BelInformed, The Netherlands (<http://www.beinformed.nl/>)
- XBRL-EUROPE, Belgium (<http://www.xbrl.org/eu/>)

•Research Institutes

- DERI, Ireland (<http://www.deri.ie/>), coordinator
- DFKI, Germany (www.dfki.de)
- UPM, Spain (<http://www.upm.es/institucional>)
- UniBi, Germany (<http://www.uni-bielefeld.de>)



Main Objectives

• **Multilingual Ontologies for Networked Knowledge**

- Linguistically enriched knowledge representation
- Multilingual access to structured/networked knowledge: ontologies, knowledge bases, linked data
=> enabling the *Multilingual Web*

- See the Multilingual Web project,
<http://www.multilingualweb.eu/>:

„The MultilingualWeb project is exploring standards and best practices that support the creation, localization and use of multilingual web-based information.“



Monnet and the Multilingual Web

- As we will see, the Monnet project investigates a specific solution to this issue. We first give examples of web pages that convey multilingual information, but that are not at the end instances of the Multilingual Web, in our sense.



Budapesti Értéktőzsde

BÉT - Kibocsátói adatok - Mozilla Firefox

http://www.bet.hu/magyar_egyeb/dinportl/companyprofile?security=608

BÉT - Kibocsátói adatok

BUDAPESTI ÉRTÉKTŐZSDE
BUDAPEST STOCK EXCHANGE

Főoldal English

KERESKEDÉSI ADATOK | ADATSZOLGÁLTATÁS | PIACOK ÉS TERMÉKEK | BEFETETŐK | TŐZSDEI KERESKEDŐK | KIBOCSÁTÓK | A TŐZSDE

reskedés tőzsdenapokon 8:30-kor folytatódik | A kereskedés tőzsdenapokon 8:30-kor folytatódik | A kereskedés tőzsdenapokon 8:30-kor folytatódik | A kereskedés tőzsdenapokon 8:30-kor folytatódik

Naptár

H	K	Sze	Cs	P	Szo	V
		1	2	3	4	5
6	7	8	9	10	11	12
13	14	15	16	17	18	19
20	21	22	23	24	25	26
27	28	29	30			

Kereskedési szünnap
Tőzsdei esemény

Kereskedési naptárak

2010. szeptember

További részletek

Bejelentkezés

Kibocsátói adatok

Cégbemutató | Kereskedési adatok | Pénzügyi adatok | Részvényesek, cégvezetés | Közzétételek | Termékjellemezők

Richter Gedeon részvény

Éves | Napi

Utolsó ár: 47 600 HUF / 2010.09.28

Historikus és interaktív grafikonok

Alapvető információk	Éves	Napi
Ticker: RICHTER	RICHTER	Utolsó ár: 47 600 HUF / 2010.09.28
ISIN: HU0000067624	48 000 HUF	
Bevezetés időpontja: 1994. nov. 09.	44 000 HUF	
Névérték: 1 000 HUF	40 000 HUF	
Bevezetett mennyiség: 18 637 486		
Kapitalizáció (millió Ft): 887 144		

Kereskedési adatok	Időszak	Napi átlagforgalom (Ft)	Átlagár
A kereskedés pénzneme: HUF	5 napos	1 022 035 714	47 533,0075
Utolsó ár: 46 500	30 napos	1 389 920 197	47 558,6055

https://mail.google.com/mail/?shva=1#inbox



Budapest Stock Exchange

BSE - Company Profile - Mozilla Firefox

http://www.bse.hu/menun_kivuli/portlets/companyprofile?security=608

BUDAPESTI ÉRTÉKTŐZSDE
BUDAPEST STOCK EXCHANGE

Home Magyar

TRADING DATA | INFORMATION SERVICES | MARKETS & PRODUCTS | MEMBERS | ISSUERS | ABOUT US

Trading will resume on trading days at 8:30 a.m.

Company profiles

Introduction | Trading data | Financials | Shareholders, management | Publications

Richter Gedeon share

Yearly

RICHTER Last price 46,500 HUF / 29/09/2010

Basic information	
Ticker	RICHTER
ISIN	HU0000067624
Date of listing	09 Nov 1994
Face value	1,000 HUF
Number of securities listed	18,637,486
Market capitalization /HUF	866,643



Comments on the localized pages of BUX

- Both the Hungarian and the English pages are parallelized and give good localized information (including on the values of certain descriptors), which can be extracted and transformed in machine-processable code (see next slide).
- But there is no possibility to address semantic queries to those pages or to consult those if the user does not understand one of both languages
 - This calls for the applications of semantic technologies that extract information and store these in language-independent way, as a base for generation in another language



Partial Terminology and Information Extraction from BUX

A screenshot of a Windows Internet Explorer browser window. The address bar shows the file path 'C:\home\Andreas\HU0000067624_data.xml'. The browser's interface includes a search bar with 'Suche' and various navigation buttons. The main content area displays XML data for a security listing, with elements and values color-coded in red and blue. The data includes fields for date of listing, face value, number of securities listed, market capitalisation, and trading data.

```
- <Date_of_listing>
  <Name xml:lang="en">Date of listing</Name>
  <Name xml:lang="hu">Bevezetés időpontja</Name>
  <Value xml:lang="en">09 Nov 1994</Value>
  <Value xml:lang="hu">1994. nov. 09.</Value>
</Date_of_listing>
- <Face_value>
  <Name xml:lang="en">Face value</Name>
  <Name xml:lang="hu">Névérték</Name>
  <Value xml:lang="en">1,000 HUF</Value>
  <Value xml:lang="hu">1 000 HUF</Value>
</Face_value>
- <Number_of_securities_listed>
  <Name xml:lang="en">Number of securities listed</Name>
  <Name xml:lang="hu">Bevezetett mennyiség</Name>
  <Value xml:lang="en">18,637,486</Value>
  <Value xml:lang="hu">18 637 486</Value>
</Number_of_securities_listed>
- <Market_capitalisation>
  <Name xml:lang="en">Market capitalisation (HUF million)</Name>
  <Name xml:lang="hu">Kapitalizáció (millió Ft)</Name>
  <Value xml:lang="en">836,823</Value>
  <Value xml:lang="hu">836 823</Value>
</Market_capitalisation>
</Basic_Information>
- <Trading_Data>
  - <Currency_of_trade>
    <Name xml:lang="en">Currency of trade</Name>
    <Name xml:lang="hu">A kereskedés pénzneme</Name>
    <Value xml:lang="en">HUF</Value>
    <Value xml:lang="hu">HUF</Value>
  </Currency_of_trade>
```



Using English as a pivot language?

- Looking at similar pages in German, Italian and Spanish, one can observe that English is used in all the cases as the second language, and that similar bi-lingual terminology and information can be extracted
- But they make use of many variations in the terms used, so that we can not use English as the intermediate language for localizing from Hungarian to, for example, Spanish
- The Monnet approach is proposing a combination of Semantic Web and Translation technologies for solving this problem



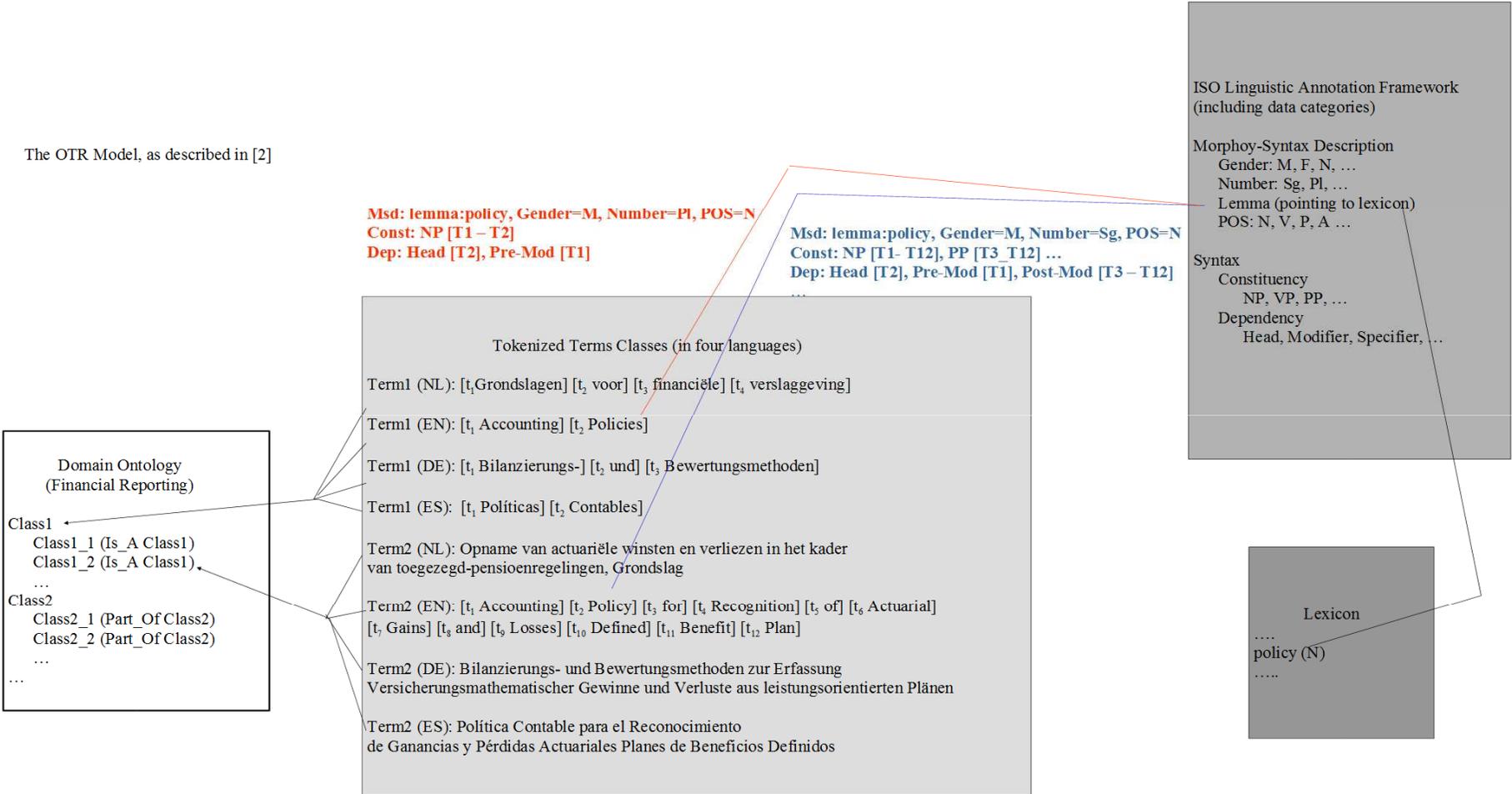
Toward a multilingual Semantic Web

- Monnet is handling information at a semantic level
 - Abstracting from language and form
 - Cross-lingual Ontology-based Information Extraction and Knowledge
 - integration,
 - aggregation,
 - querying,
 - presentation
- The Semantic Web as such is not proposing ways to deal with natural language and multilingualism, and a first step in Monnet consists in proposing a model for the representation of natural language expressions used in ontologies in the Semantic Web



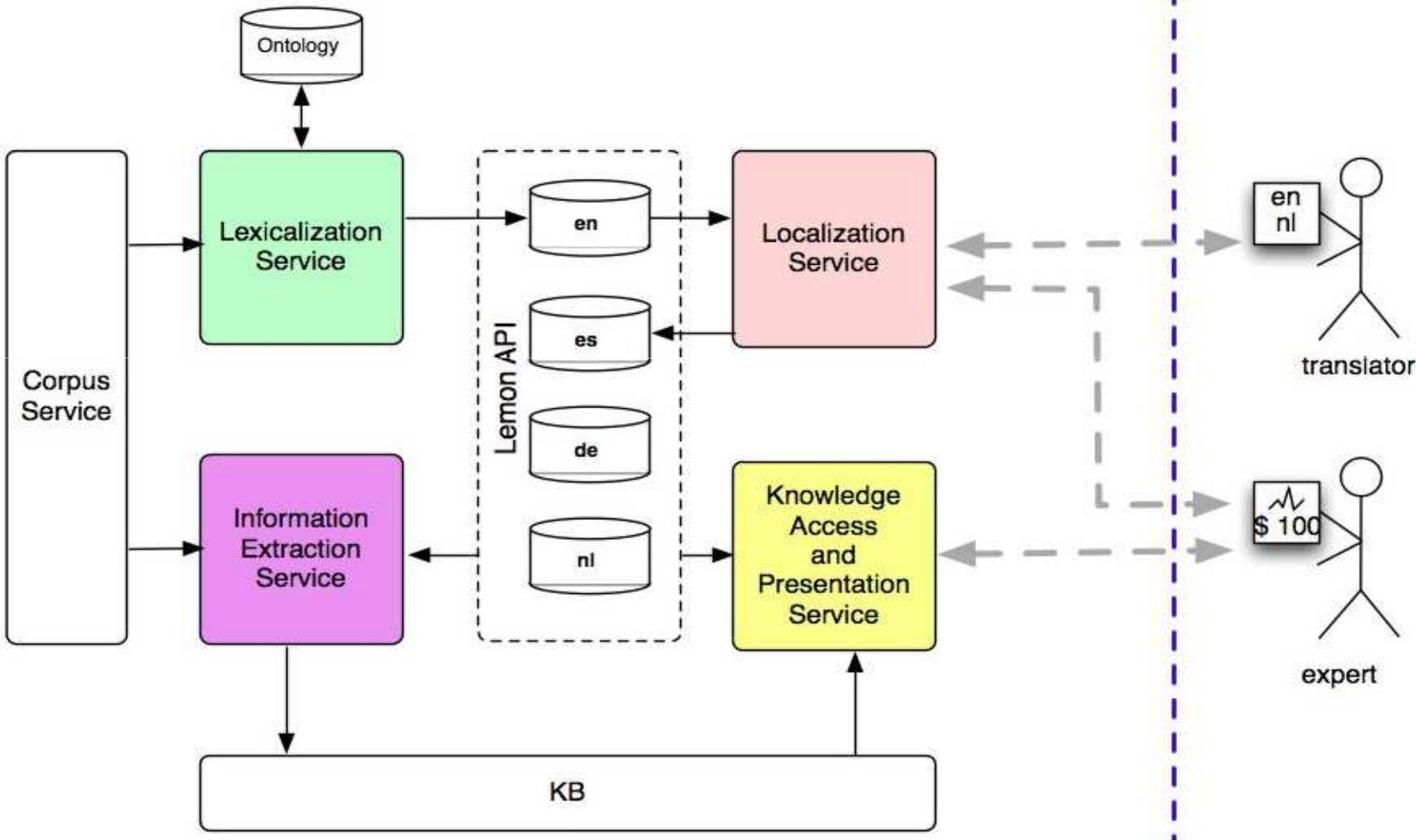
Make use of linguistically described multilingual terms in Ontologies

The OTR Model, as described in [2]



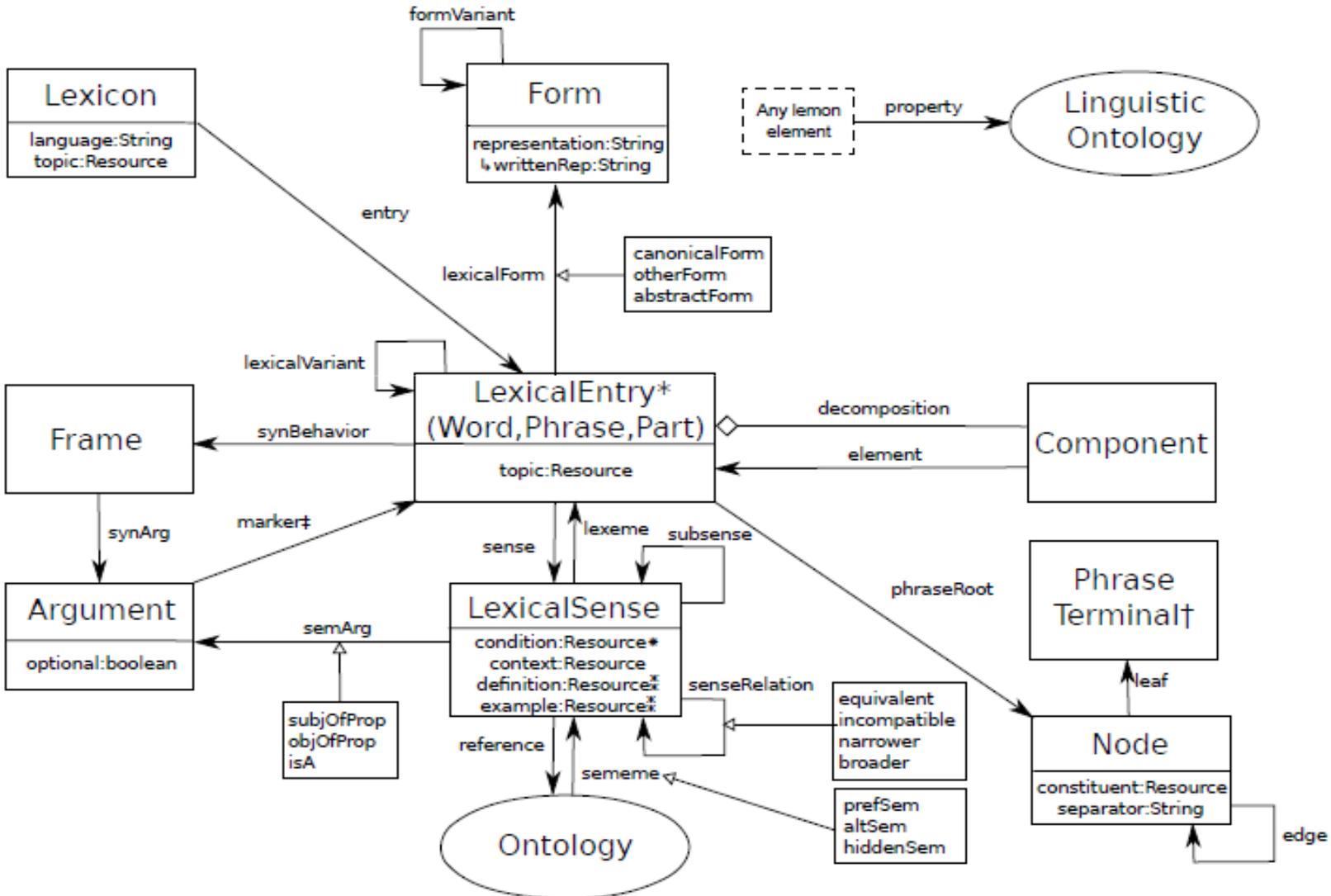


Monnet Architecture and Components





The lemon model, for encoding lexicalized ontologies





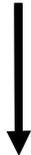
Term Analysis for Machine Translation

minimum finance lease payments receivable ...

Google-Translate: minimale Finanzierung uneinbringliche

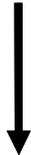
IFRS: im Rahmen von Finanzierungs-Leasingverhältnissen zu erhaltende Mindestleasingzahlungen

*term analysis – using:
IFRS, SAPterm, IATE, DBpedia, ...*



minimum finance lease payments receivable

*sub-term translation – using:
IATE, DBpedia, Leo, ...*



mindest Finanzierungsleasing Zahlungen erträge

*term synthesis – using:
rules? statistical models?*



Mindestfinanzierungsleasingzahlungserträge

Google-Translate: minimum finance lease payment income



Term Analysis for Information Extraction

minimum finance lease payments receivable ...

*term analysis – using:
IFRS, SAPterm, IATE, DBpedia, ...*



minimum finance lease payments receivable

sentence classification and instance detection



IBM Annual Report 2007

Scheduled maturities of **minimum lease payments** outstanding at **December 31, 2007**, expressed as a percentage of the total, are approximately: **2008**, 50 percent; **2009**, 30 percent; **2010**, 15 percent; **2011**, 5 percent; and **2012** and beyond, 1 percent.

Barclay's Annual Report 2009

The allowance for uncollectable **finance lease receivables** included in the allowance for impairment amounted to **£321m** at **31st December 2009** (2008: £189m).

The total of future **minimum sublease payments** to be **received** under non-cancellable **subleases** at the balance sheet date is **£147m** (2008: **£158m**).



Outcomes

- An *Ontology Lexicalization Service* to extract lexical expressions used in labels of ontologies, and to create an ontology lexicon automatically enriched with linguistic information
- An *Ontology Localization Tool* to create an ontology lexicon in a target language from an ontology lexicon in a source language, semi-automatically
- A *Cross-Lingual Ontology-based Information Extraction System* (CLOBIE) to leverage multilingual ontology lexicons to extract information from text and populate ontologies
- A *Cross-Lingual Query & Presentation System*, which uses multilingual ontology lexicons to enable quick customization of knowledge access systems to many natural languages. The knowledge being here stored in populated ontologies



Use Cases

- The project is validating its approach to enabling the multilingual web in the context of two use cases:
 - in the field of e-Government
 - in the field of financial and business information
- Focus in this talk is on 2., called IFRS-XBRL use case, which aims at enabling the search and the report creation of financial information and business service descriptions in the language of choice of the users.



IFRS-XBRL Use case

- Objective: Enable search and report creation in language of choice of financial information
 - Business intelligence on European companies, involving semantic-level analysis of business reports
- Impact: An integrated solution to providing semantic-level access to financial information across languages
 - The prototype will allow a financial analyst to search for information by filling in structured search forms localized to his/her own language. The results will be presented in terms of charts, diagrams, results lists etc. localized to the preferred language of the user

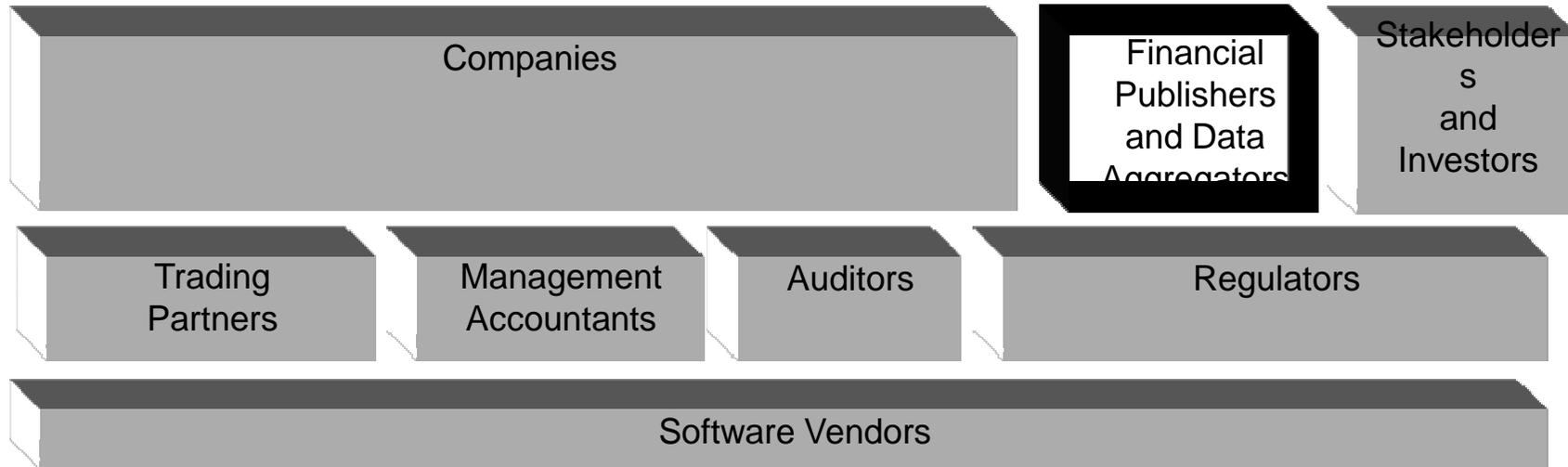


Data Sources

- Structured sources:
 - publicly available balance sheets in structured format (for example in German at <http://www.bundesanzeiger.de/>);
 - Short company profiles (e.g from Business Registers, Stock Exchange web pages, etc.);
 - Wikipedia Infoboxes; and
 - XBRL instance documents (the Belgian National Bank has published on-line all the XBRL reports of Belgian companies).
- Semi-structured:
 - Longer company profiles;
 - Imprint information on company web pages;
 - Running tickers on company information.
- Unstructured:
 - Annexes to balance sheets in annual reports of companies;
 - Newspapers;
 - Specialized web pages etc.



Potential Users



Main beneficiaries:

1. The translators of the XBRL taxonomies (who will profit from the effort reduction yielded by using our automatic ontology localization functionality)
2. The users of XBRL taxonomies (e.g. financial analysts), who will be able to see the information formalized by XBRL instance documents or extracted from unstructured resources in accordance to the XBRL taxonomies in their preferred language.



XBRL and IFRS

- XBRL (eXtensible Business Reporting Language) is an XML-based mark-up language for the exchange of business information, including financial reporting. Its use is being nowadays mandated by a growing number of regulatory bodies and stock exchanges around the world.
- IFRS (International Financial Reporting Standards) is a standard developed by the International Accounting Standards Board (IASB). IFRS is also available as a taxonomy encoded in XBRL.

<http://www.xbrl-ifs.org/ITMM/>

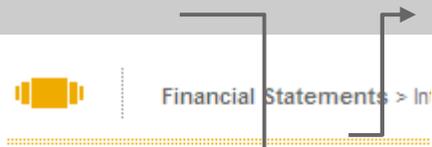
Available IFRS Modules

- ▷ IFRS 1 - First-time Adoption of International Financial Reporting Standards
- ▷ IFRS 2 - Share-based Payment
- ▷ IFRS 3 - Business Combinations
- ▷ IFRS 4 - Insurance Contracts
- ▷ IFRS 5 - Non-current Assets Held for Sale and Discontinued Operations
- ▷ IFRS 6 - Exploration for and Evaluation of Mineral Resources
- ▷ IFRS 7 - Financial Instruments: Disclosures
- ▷ IFRS 8 - Operating Segments
- ▷ IAS 1 - Presentation of Financial Statements
- ▷ IAS 2 - Inventories
- ▷ IAS 7 - Statement of Cash Flows
- ▷ IAS 8 - Accounting Policies, Changes in Accounting Estimates and Errors
- ▷ IAS 10 - Events after the Reporting Period
- ▷ IAS 11 - Construction Contracts
- ▷ IAS 12 - Income Taxes
- ▷ IAS 16 - Property, Plant and Equipment
- ▷ IAS 17 - Leases
- ▷ IAS 18 - Revenue
- ▷ IAS 19 - Employee Benefits
- ▷ IAS 20 - Accounting for Government Grants and Disclosure of Government Assistance
- ▷ IAS 21 - The Effects of Changes in Foreign Exchange Rates
- ▷ IAS 23 - Borrowing Costs (2007-03-01)
- ▷ IAS 24 - Related Party Disclosures
- ▷ IAS 26 - Statement of changes in net assets available for benefits
- ▷ IAS 27 - Consolidated and Separate Financial Statements
- ▷ IAS 28 - Investments in Associates
- ▷ IAS 29 - Financial Reporting in Hyperinflationary Economies
- ▷ IAS 31 - Interests in Joint Ventures

Available IFRS Modules

- ▷ IFRS 7 - Financial Instruments: Disclosures
- ▷ IFRS 8 - Operating Segments
- ▶ IAS 1 - Presentation of Financial Statements
 - General information about financial statements
 - Statement of financial position, current/non-current - Consolidated financial statements
 - Statement of financial position, current/non-current - Separate financial statements
 - Statement of financial position, order of liquidity - Consolidated financial statements
 - Statement of financial position, order of liquidity - Separate financial statements
 - Income statement, by function of expense - Consolidated financial statements
 - Income statement, by function of expense - Separate financial statements
 - Income statement, by nature of expense - Consolidated financial statements
 - Income statement, by nature of expense - Separate financial statements
 - Statement of comprehensive income - Consolidated financial statements
 - Statement of comprehensive income - Separate financial statements
 - Statement of comprehensive income [alternative] - Consolidated financial statements
 - Statement of comprehensive income [alternative] - Separate financial statements
 - Statement of changes in equity - Consolidated financial statements
 - Statement of changes in equity - Separate financial statements
 - Notes - Corporate information and statement of IFRS compliance
 - Notes - Analysis of income and expense
 - Notes - Subclassifications of assets, liabilities and equities
 - Notes - Share capital, reserves and other equity interest
- ▷ IAS 2 - Inventories

A Financial Reports: duration vs. Instant



- Introduction
- Consolidated Income Statements
- Consolidated Statements of Comprehensive Income
- Consolidated Statements of Financial Position
- Consolidated Statements of Changes in Equity
- Consolidated Statements of Cash Flows
- Notes to the Consolidated Financial Statements
- Financial Statement of SAP AG (HGB) - Short Version

FINANCIAL STATEMENTS IFRS

Values over a period, duration

Consolidated Income Statements of SAP Group for the years ended December 31,

€ millions, unless otherwise stated	Note	2009	2008	2007
Software revenue		2,607	3,606	3,407
Support revenue		5,285	4,602	3,852
Subscription and other software-related service revenue		306	258	182
Software and software-related service revenue		8,198	8,466	7,441
Consulting revenue		2,074	2,498	2,221
Training revenue		273	434	410
Other service revenue		85	107	113
Professional services and other service revenue		2,432	3,039	2,744
Other revenue		42	70	71
Total revenue	(5)	10,672	11,575	10,256

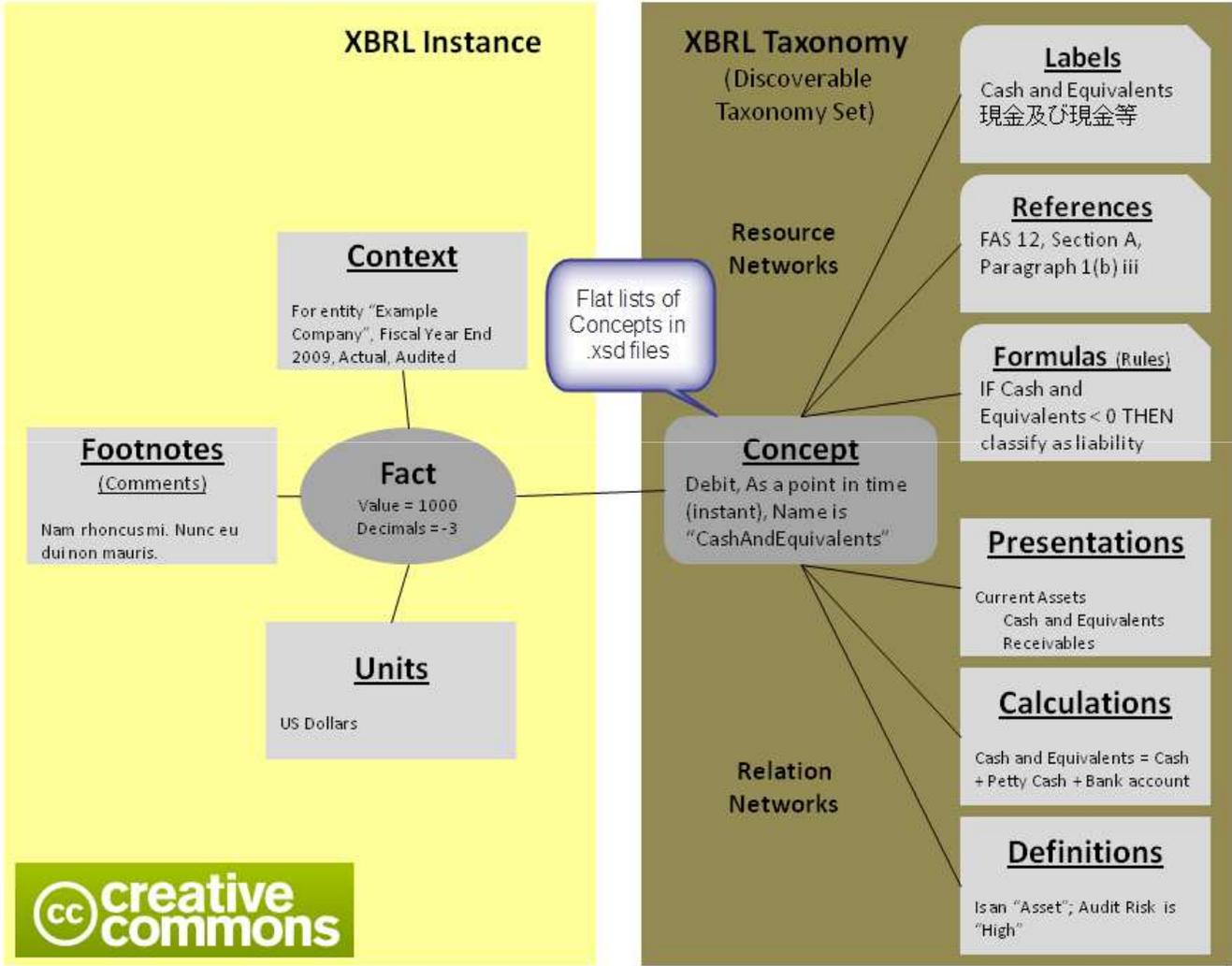
Consolidated Statements of Financial Position of SAP Group as at December 31,

Values at an instant, position

Assets	€ millions	Note	2009	2008
Cash and cash equivalents			1,884	1,280
Other financial assets		(13)	486	588
Trade and other receivables		(14)	2,546	3,178
Other non-financial assets		(15)	147	126
Tax assets		(11)	192	399
Total current assets			5,255	5,571

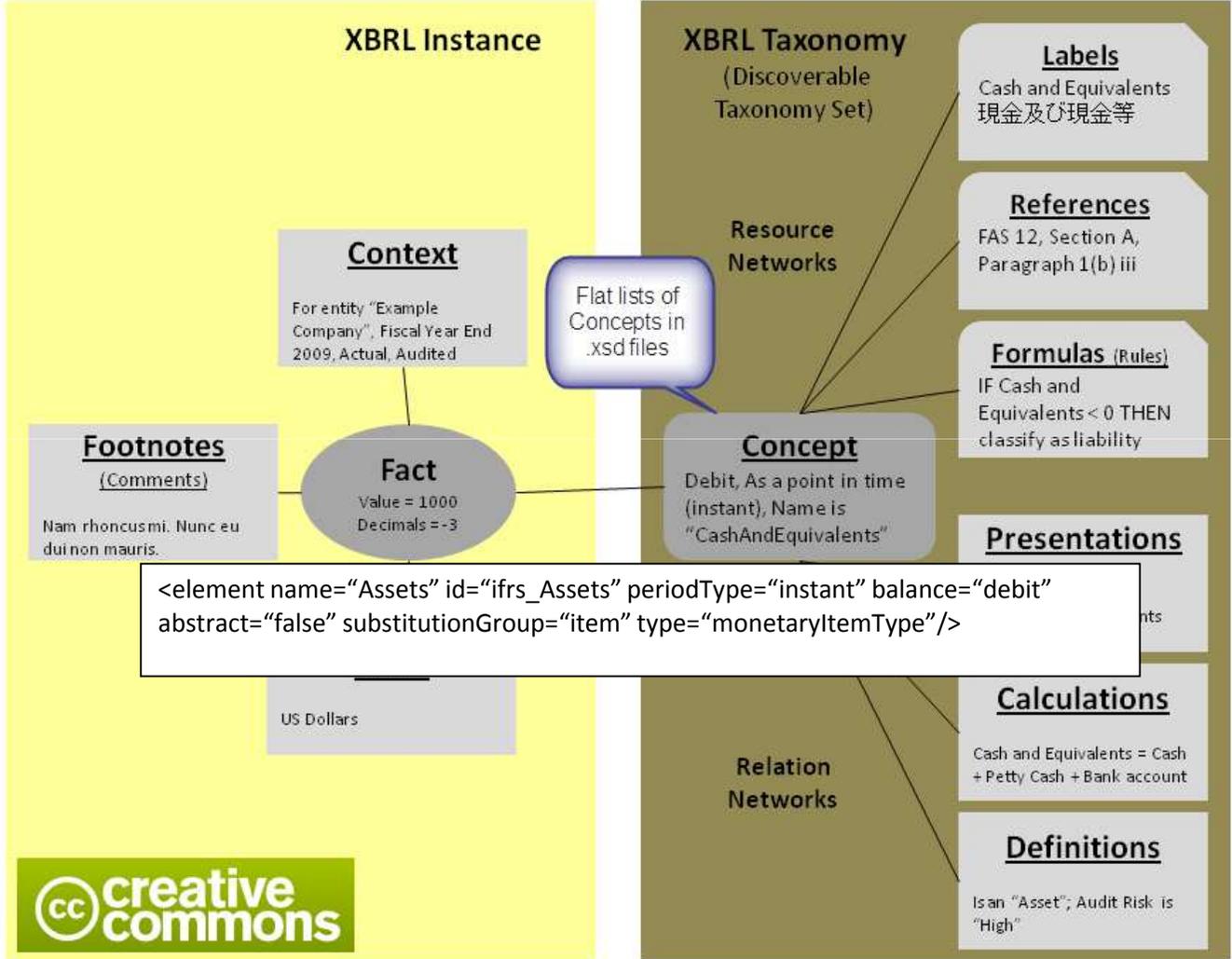
Such Instances of financial reports exists at most in 2 languages!

High Level Model of XBRL



XBRL - Instance

High Level Model of XBRL





Searching through the IFRS Taxonomy in XBRL

ABRA Search | A magic XBRL Taxonomy Search - Mozilla Firefox

http://www.abra-search.com/ABRASearch.html?locale=en&taxonomy=ifrs_2009-04-01

Meistbesuchte Seiten Erste Schritte Aktuelle Nachrichten Google Mail - Inbox - t... LEO Deutsch-Englisch...

ABRA Search | A magic XBRL Taxono...

ABRA XBRL SEARCH

profit

IFRS Taxonomy (2009-04-01) en

Submit Taxonomy Taxonomy Hosting / Licensing FAQ Imprint

Hits (total): 197 Pages (total): 8

Rank Concept

- 1 Profit (loss) of acquiree
- 2 Profit (loss) of combined entity
- 3 Adjustments for reconcile profit (loss)
- 4 Profit (loss)
- 5 Profit (loss), attributable to
- 6 Actuarial gains (losses) recognised in profit or loss, defined benefit plan
- 7 Adjustments for reconcile profit (loss)
- 8 Disclosure of effect of share-based payments on profit or loss
- 9 Expenses on financial assets reclassified out of financial assets at fair value through profit or loss recognised in profit or loss
- 10 Fair value gains (losses) on financial assets reclassified out of financial assets at fair value through profit or loss recognised in profit or loss
- 11 Gains (losses) on financial assets reclassified out of financial assets at fair value through profit or loss recognised in profit or loss
- 12 Impairment loss recognised in profit or loss
- 13 Impairment loss recognised in profit or loss, computer software
- 14 Impairment loss recognised in profit or loss, exploration and evaluation assets
- 15 Impairment loss recognised in profit or loss, licences and franchises
- 16 Impairment loss recognised in profit or loss, other intangible

- [110000] General information about financial statements
- [210000] Statement of financial position, current/non-current - Consolidated financial statements
- [210005] Statement of financial position, current/non-current - Separate financial statements
- [220000] Statement of financial position, order of liquidity - Consolidated financial statements
- [220005] Statement of financial position, order of liquidity - Separate financial statements
- [310000] Income statement, by function of expense - Consolidated financial statements
 - Income statement
 - Profit (loss)
 - Profit (loss), attributable to
 - Profit (loss), attributable to owners of parent
 - Profit (loss), attributable to non-controlling interests
 - Earnings per share
- [310005] Income statement, by function of expense - Separate financial statements
- [320000] Income statement, by nature of expense - Consolidated financial statements
- [320005] Income statement, by nature of expense - Separate financial statements

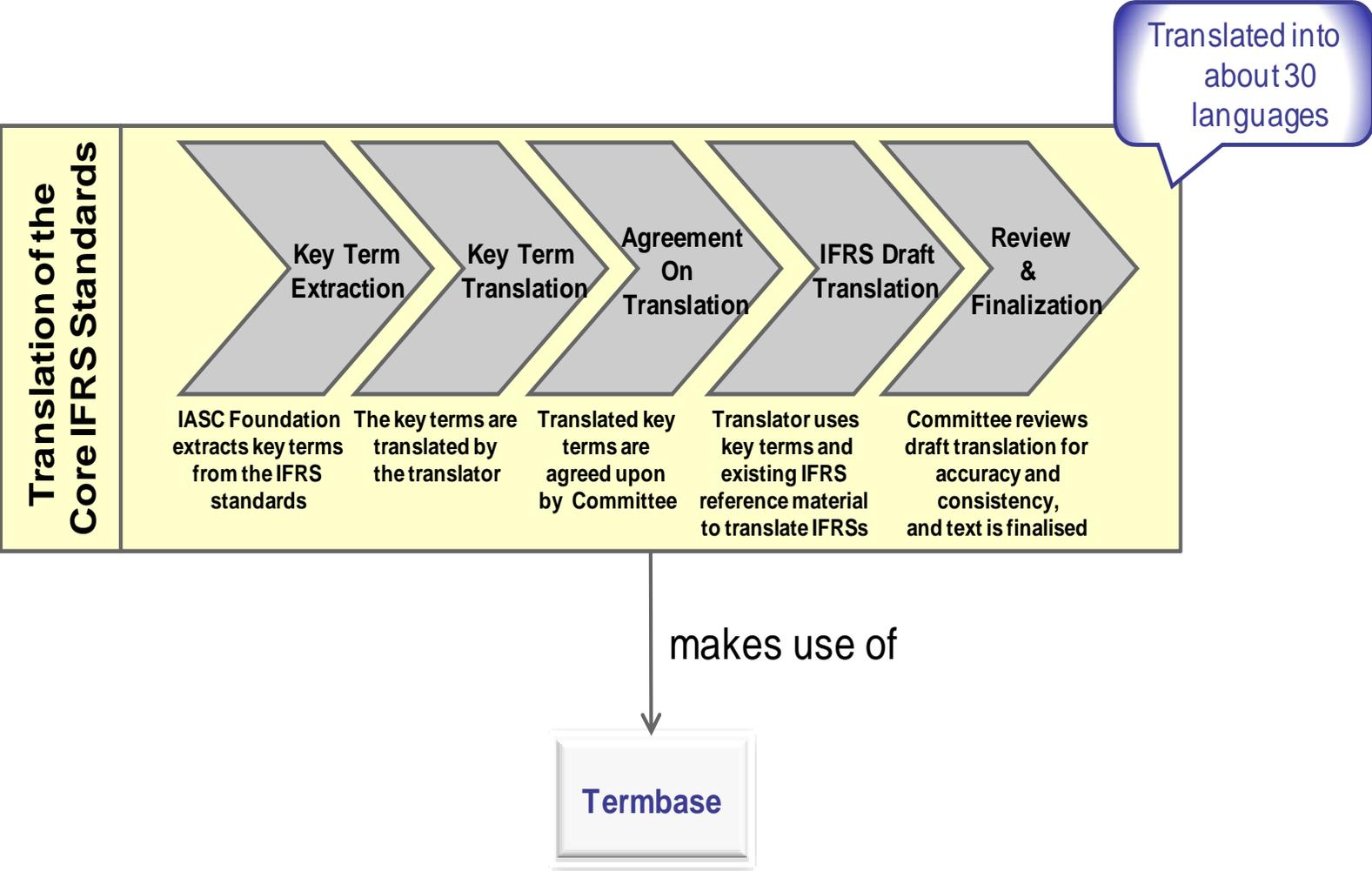
Languages References Details Calculations Bookmark

Label	Language	Purpose
Profit (loss), attributable to owners of parent	en	label
الربح (الخسارة)، المنسوب إلى مالكي الشركة الأم	ar	label
Gewinn (Verlust), der den Eigentümern des Mutterunternehmens zuzurechnen ist	de	label
Ganancia (pérdida), atribuible a los propietarios de la controladora	es	label
Résultat, attribuable aux propriétaires de l'entité mère	fr	label
Utile (perdita), attribuibile ai soci della controllante	it	label
Winst (verlies), toerekenbaar aan eigenaars van moedermaatschappij	nl	label
利润(亏损), 归属于母公司所有者	zh	label

Fertig



Translation of the Core IFRS Standards, as described by IFRS





Comparing Translation of Text to Translation of XBRL Taxonomy

	Model in Workbench	Metadata about elements	Context seen by translator	Context used for suggestions
Text	Sequence of text/segments	Text type: paragraph, header, caption, other?	Entire text	Preceding and following sentences; Subject Field?
XBRL Taxonomy	Graphs of labels (sequentialized somehow)	Type of label; Concept attributes	Locations in graphs; reference to standard	Metadata; Context seen; Subject Field of referenced standard?

Hypothesis: Given the taxonomy, English labels, and the translation of the IFRS standards into a language, the XBRL labels for the language can be automatically deduced with a very high accuracy. The translator would, however, need to check them. The IFRS termbase might aid this task.

Example IFRS Taxonomy Label and its Translation

- Minimum finance lease payments receivable, at present value, end of period later than one year and not later than five years
- Im Rahmen von Finanzierungs-Leasingverhältnissen zu erhaltende Mindestleasingzahlungen, zum Barwert, länger als ein Jahr und bis zu fünf Jahren bis zum Ende der Periode

Example of inconsistency in IFRS 2009 Taxonomy Labels:

- *'end of period later than one year and not later than five years'*
 - occurs as a segment in 8 labels.
- It is translated into German 6 times as:
 - *'länger als ein Jahr und bis zu fünf Jahren bis zum Ende der Periode'*
- 2 times as:
 - *'mehr als ein Jahr und bis zu fünf Jahren bis zum Ende der Periode'*



Information Extraction (IE) can find values in text and relate them to concepts, so that the values can be checked.

Similarly, IE can help with detailed tagging.

Example text about derivatives

- The fair value hedge has a notional amount of \$250 million, and hedges approximately 86% of the \$292 million of outstanding senior notes maturing in September 2011.

IE equates the three underlined values with XBRL tags to create three different facts:

- NotionalAmountOfInterestRateFairValueHedgeDerivatives has the value \$250 million.
- PercentageOfDebtHedgedByInterestRateDerivatives has the value 86%.
- SeniorNotes has the value \$292 million.

The facts might also have the context September 2011

Information about the taxonomy location (etc.) of the concepts may help confirm the interpretation.



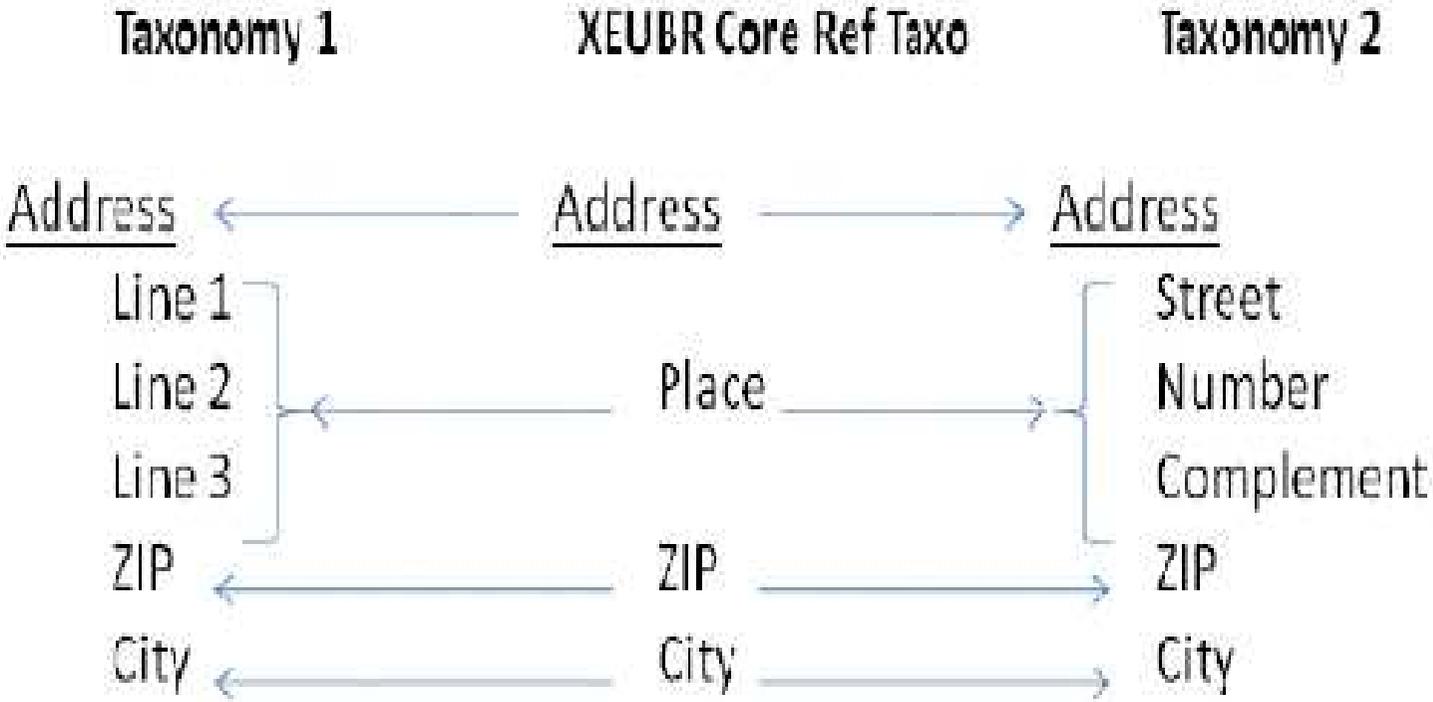
Monnet involved in **xEBR** **European Business Registers** **Mapping between taxonomies**

Registers - Information Providers **(XEU BR WG Members)**

- *National Bank of Belgium (Belgium)*
- *Eogs / DCCA (Denmark)*
- *Registrite ja infosüsteemide Keskus eRik (Estonia)*
- *Bilans Service - Infogreffe (France)*
- *Bundesanzeiger (Germany)*
- *Infocamere (Italy)*
- *RSCL (Luxembourg)*
- *Kamer van Koophandel (Netherlands)*
- *Informa DB – Colegio de Registradores (Spain)*
- *Bolagsverket (Sweden)*
- *Companies House (United Kingdom)*
- *EBR (Europe)*
- *GBR (Global)*
- *IASCF*
- *Bank of Spain*
- *Software – Audit – Consulting*

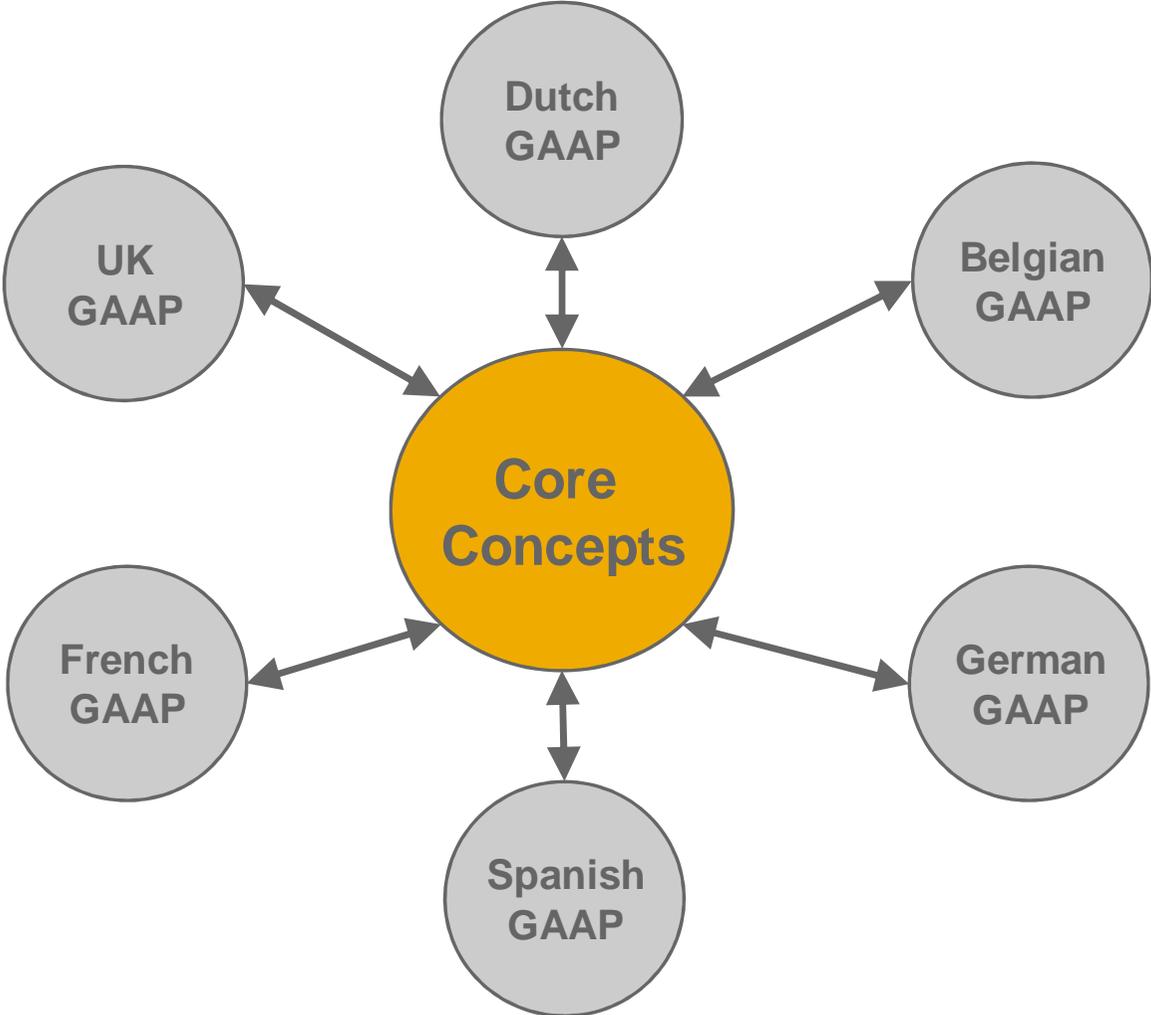


An example of Mapping





xEBR European Business Registers Mapping between taxonomies





Summary

- There are various ways to ensure interoperability of information in the web
 - By mapping knowledge representation systems (not at the core of MONNET)
 - My making knowledge representation systems multilingual, in order to support real interaction with the European citizen
- Thanks to the Montific project and for your attention
 - Contact: declerck@dfki.de
 - Webpage of Monnet: <http://www.monnet-project.eu/>.