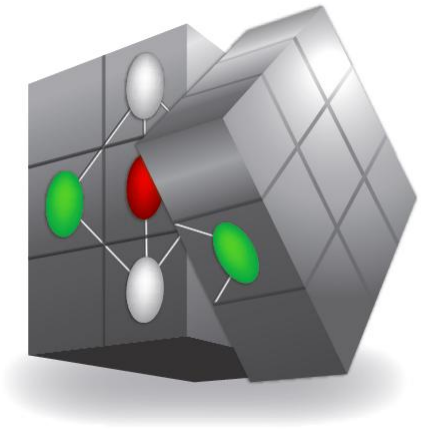
Slide with demo video, removed for th pdf-version of the slides

Content: CUBIST promotional video

Watch instead: https://www.youtube.com/watch?v=RC7Ncj2MYbQ



cubist

Your Business Intelligence

Dr. Frithjof Dau, Senior Researcher, SAP AG

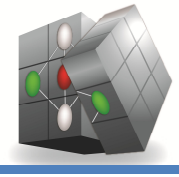
Fourth European Business Intelligence Summer School (eBISS 2014)

Agenda

- Project Setup and Key Technologies
  - First Introduction into CUBIST
  - Use Cases
  - Introduction into Semantic Technologies
  - Introduction into Formal Concept Analysis
  - Key Messages
- CUBIST Prototype
  - Architecture
  - Different Means to Access Information
  - Semantic Search
  - Query Generation
  - Explorative Search
  - Conceptual Scaling
  - Visual Analytics
- Outcome
  - User Evaluation
  - Our Take
  - Conclusions

Agenda

- Project Setup and Key Technologies
  - First Introduction into CUBIST
  - Use Cases
  - Introduction into Semantic Technologies
  - Introduction into Formal Concept Analysis
  - Key Messages
- CUBIST Prototype
  - Architecture
  - Different Means to Access Information
  - Semantic Search
  - Query Generation
  - Explorative Search
  - Conceptual Scaling
  - Visual Analytics
- Outcome
  - User Evaluation
  - Our Take
  - Conclusions



CUBIST – Project Details

Instrument

- Instrument: STREP
- Theme: ICT-2009-4.3
- Call: FP7 Call 5
- Lead: SAP Research
- Duration: 36 Months
- Start: 2010/10
- Effort: 403,00
- Budget/Funding: 4.357.195,41 / 3.029.836,00

Consortium

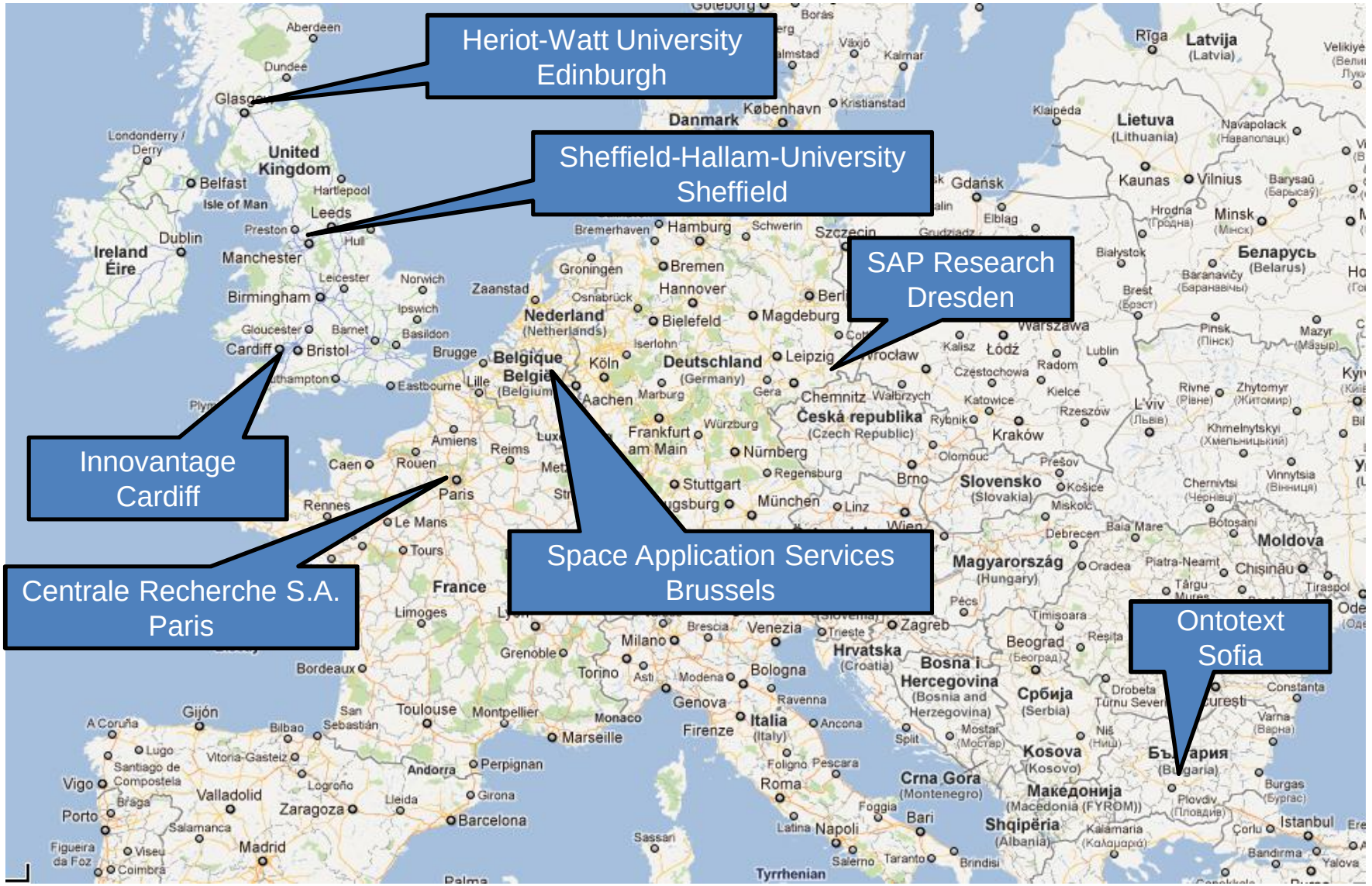
Technological Partners

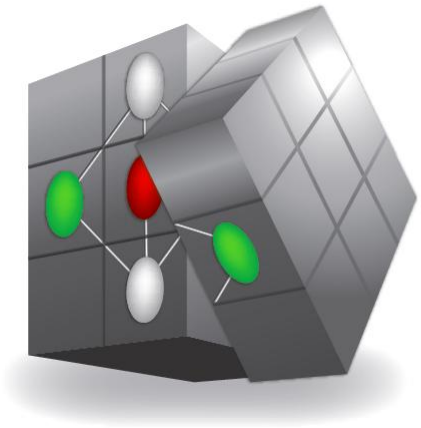
- SAP (Germany)
  - Coordinator and technological partner
- Ontotext (Bulgaria)
  - Expertise in Semantic Technologies
- Sheffield Hallam University (UK)
  - Expertise in FCA
- Centrale Recherche S.A. (France)
  - Expertise in FCA and Visual Analytics

Use Case Partners

- Heriot-Watt University (UK)
- Space Applications Services (Belgium)
- Innovantage (UK)

CUBIST – Partner



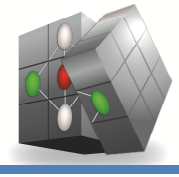


cubist

Your Business Intelligence

CUBIST in a nutshell: Developing an approach for semantic and user-friendly Business Intelligence by

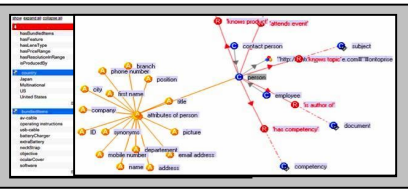
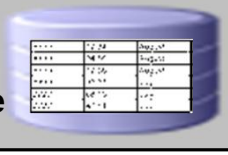
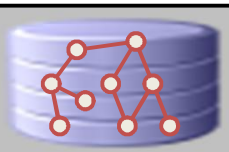
- augmenting Semantic Technologies with BI capabilities, and
- providing conceptually relevant and user friendly visual analytics.



Initial Motivation

- Increased proportion of unstructured data (>80%)
  - Not accessible for classical BI solutions
  - Can be better leveraged by means of Semantic Technologies (ST)
- Insufficient user interfaces for Business Intelligence (BI)
  - Improved visual analytics, based on Formal Concept Analysis (FCA), for qualitative Data Analysis
  - Complementing to existing approaches for quantitative Data Analysis

CUBIST Main Idea
From classical to semantic BI

Classical Business Intelligence		Semantic Business Intelligence
	**Output**	 
restricted queries / analytics	**User Interaction**	flexible and visual queries / analytics
Data Warehouse 	**Store**	**Triple Store** 
ETL	**Gathering Information**	Semantic ETL
 databases	**Data sources**	 databases  Forums, blogs  Office docs

CUBIST Main Idea
From classical to semantic BI

CUBIST: Developing an approach for semantic and user-friendly BI

Providing conceptually relevant and user friendly visual analytics.

- Formal Concept Analysis / Galois Lattices
- Faceted navigation
- Graph-based navigation

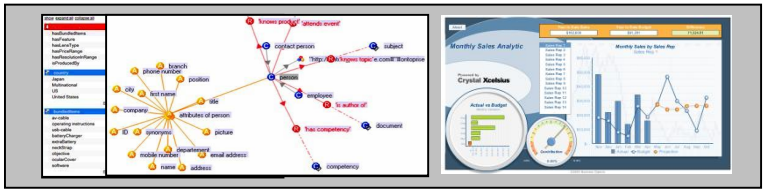
Augmenting Semantic Technologies with BI capabilities

- Triple store as persistency layer
- Flexible Data Warehouse design
- Extending SPARQL with OLAP functionalities
- Reasoning / Deriving implicit facts

Federating data from both unstructured and structured sources

- Enhanced ETL
- Text Mining
- Information Extraction

Semantic Business Intelligence



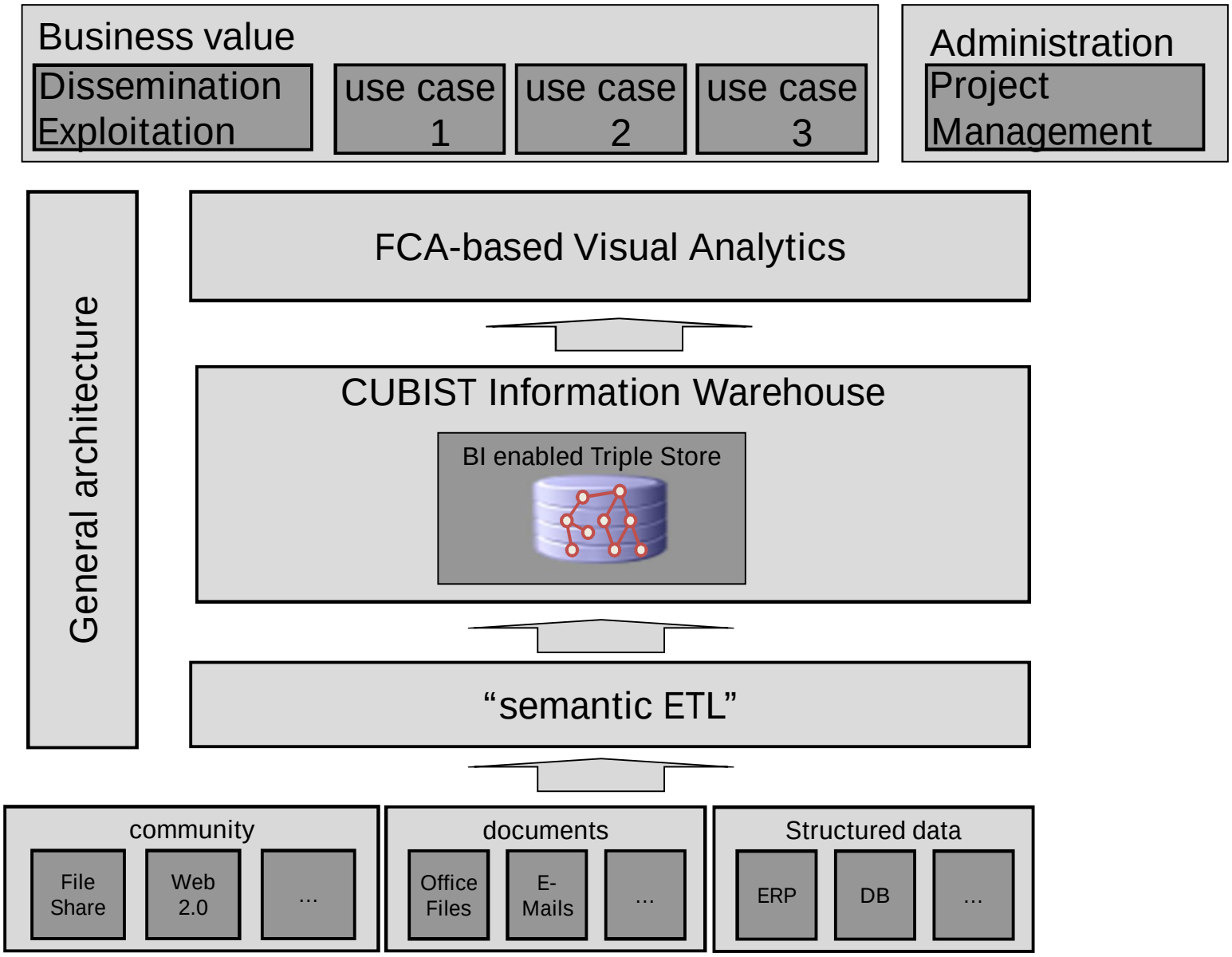
flexible and visual queries / analytics

Triple Store

Semantic ETL

databases | Forums, blogs | Office docs

CUBIST Highlevel Architecture



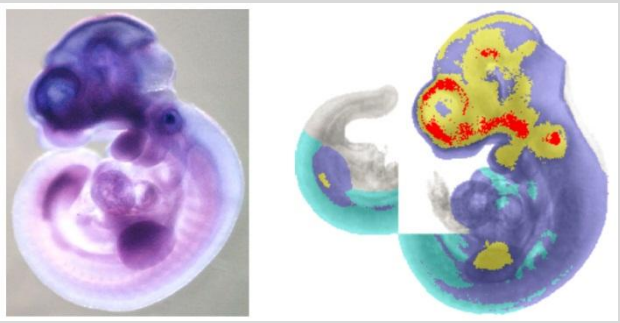
Agenda

- Project Setup and Key Technologies
  - First Introduction into CUBIST
  - Use Cases
  - Introduction into Semantic Technologies
  - Introduction into Formal Concept Analysis
  - Key Messages
- CUBIST Prototype
  - Architecture
  - Different Means to Access Information
  - Semantic Search
  - Query Generation
  - Explorative Search
  - Conceptual Scaling
  - Visual Analytics
- Outcome
  - User Evaluation
  - Our Take
  - Conclusions

CUBIST Use Cases

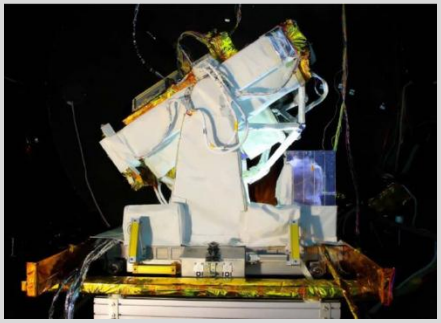
Heriot-Watt University

Analysis of gene expressions in mouse embryos



Space Applications Services

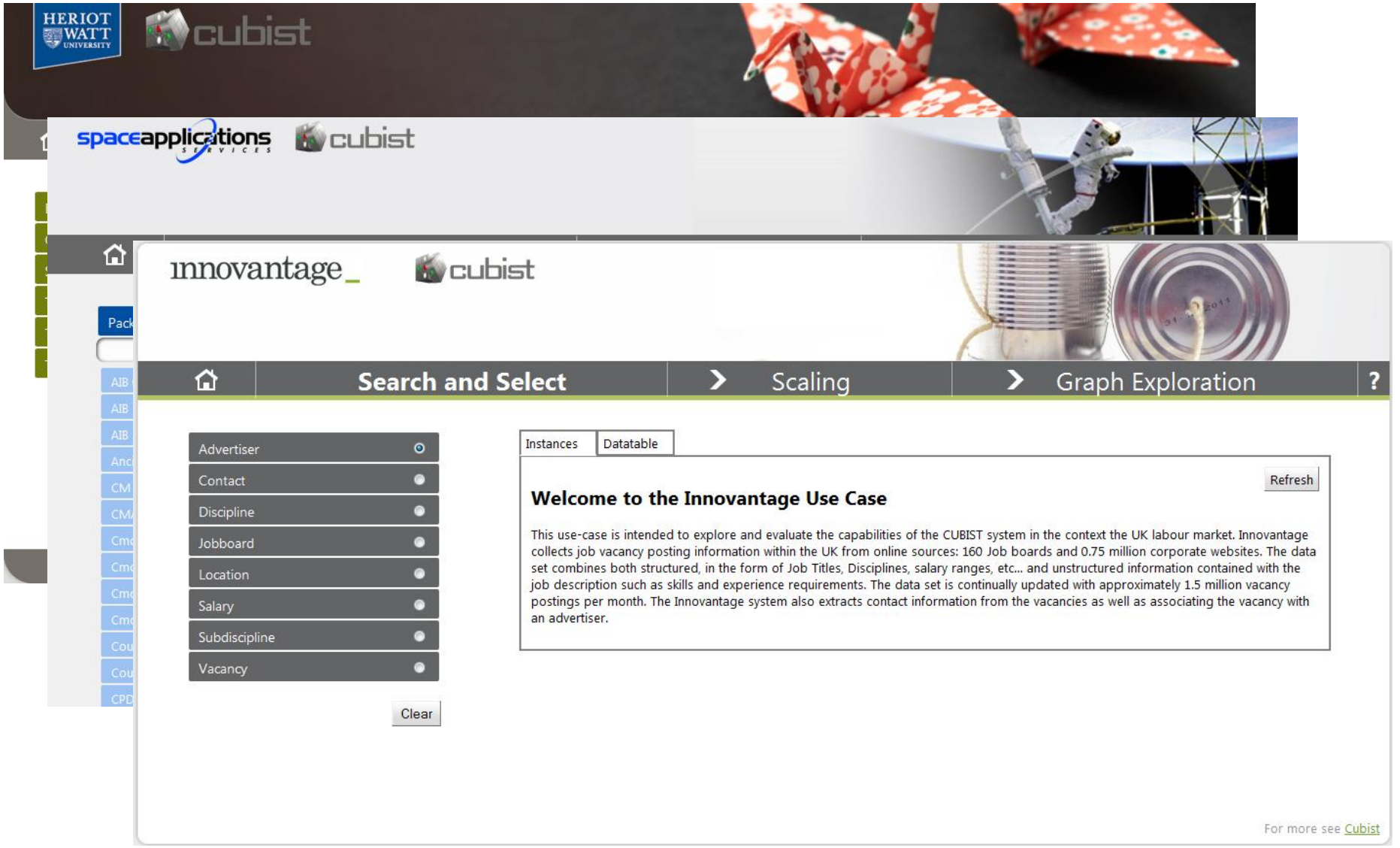
Analysis of logfiles of technical equipment in space



Innovantage

Analysis of the online recruitment activities of UK companies

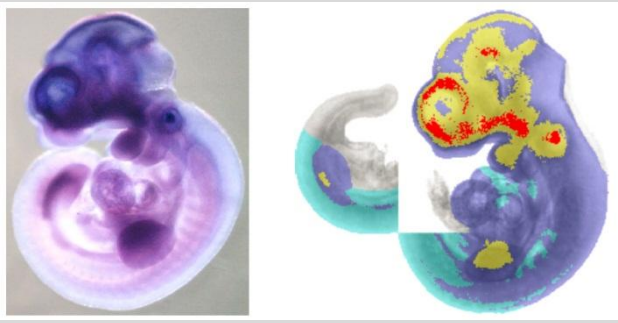
CUBIST Use Cases



CUBIST Use Cases

Heriot-Watt University

Analysis of gene expressions in mouse embryos



HWU Use Case

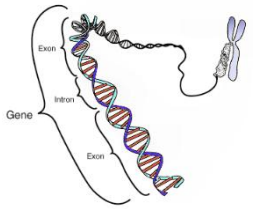
- Biological use case
- Conceptual approach to gene expression analysis enhanced by visual analytics
- Based on the *in situ* hybridisation gene expression data held within the EMAGE database
  - EMAGE (e-Mouse Atlas of Gene Expression is an online biological database of gene expression data in the developing mouse embryo.
  - EMAGE data is also text annotated to provide a text based description of the expression patterns.

HWU Use Case

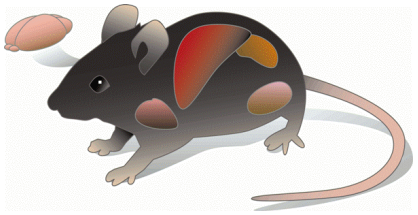
- In CUBIST, we dealt with textual annotations, e.g

Wnt1 is **detected** in the **neural extoderm**

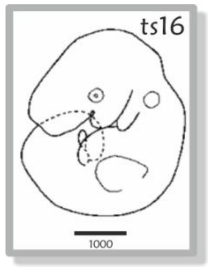
Gene Strength Tissue



weak
strong
not detected
etc



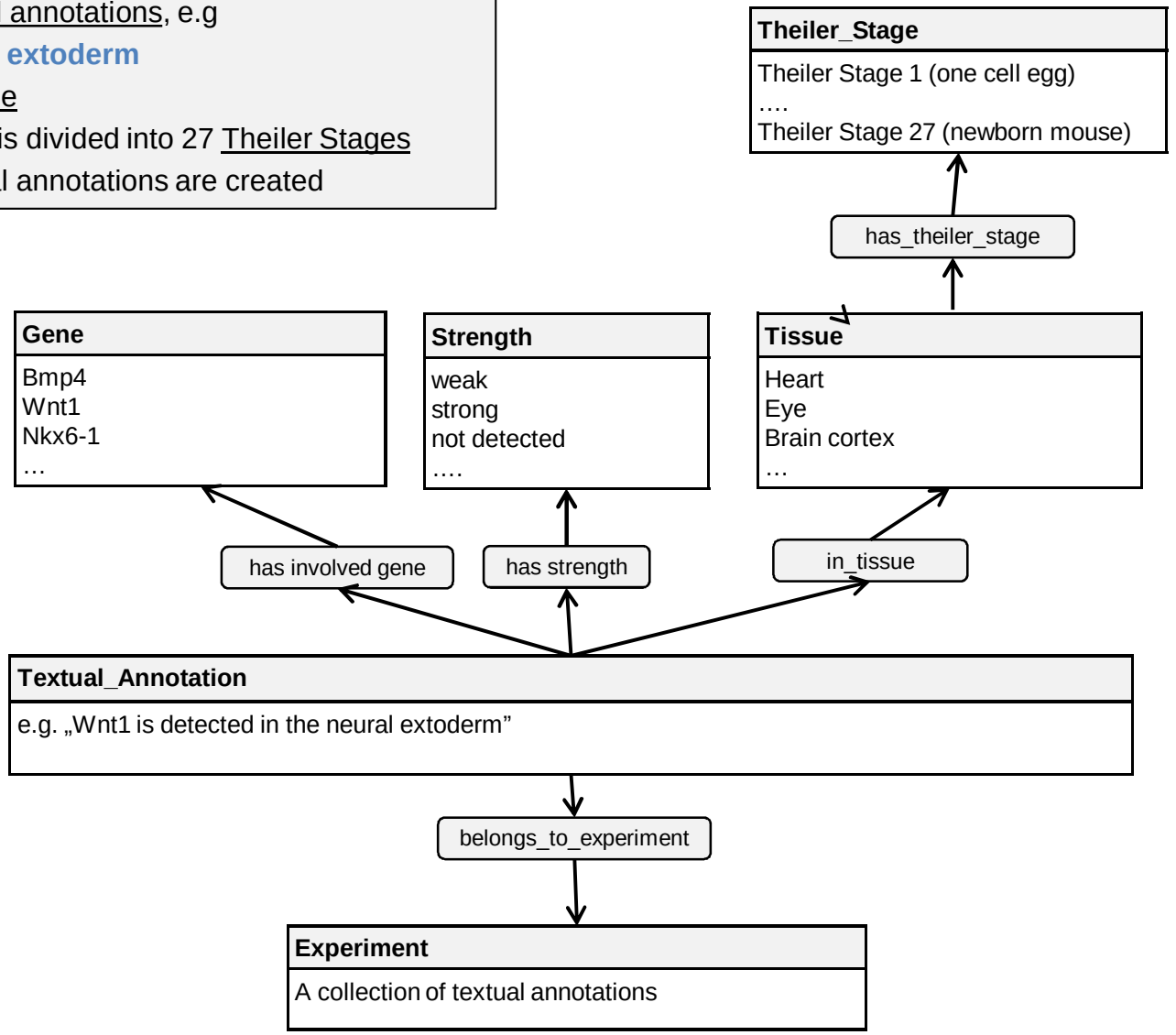
- The development of the mouse is divided into 27 Theiler Stages



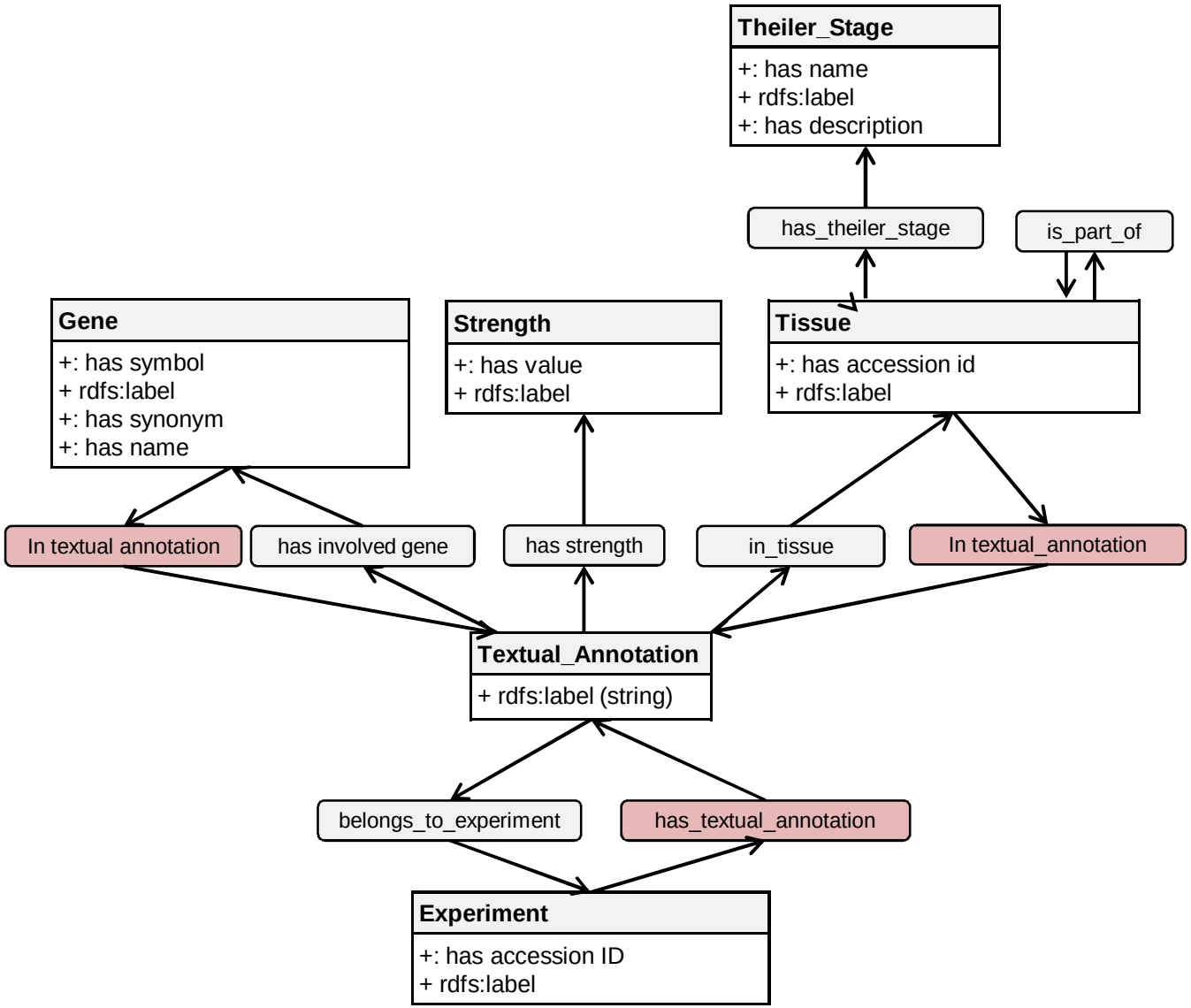
- In an experiment, several textual annotations are created

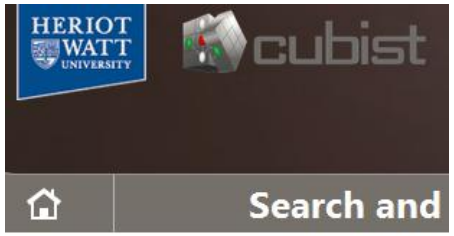
HWU Ontology (informal)

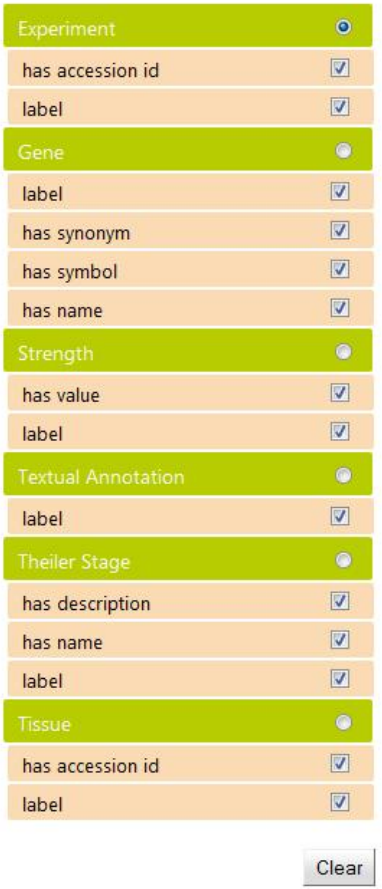
- In CUBIST, we dealt with textual annotations, e.g
  Wnt1 is **detected** in the **neural extoderm**
  Gene Strength Tissue
- The development of the mouse is divided into 27 Theiler Stages
- In an experiment, several textual annotations are created



HWU Ontology







HWU Use Case

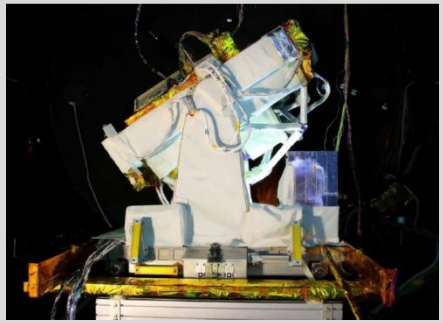
- Typical queries/information needs
  - Compare the gene expression profile of genes Bmp2, Bmp3 and Bmp4 in Theiler Stage 17
  - Compare the gene expression profile of the heart in in Theiler Stage 12

- Problems
  - No numbers: traditional BI means fall short here
  - No visual analytics tools for this use case

CUBIST Use Cases

Space Applications Services

Analysis of logfiles of technical equipment in space



Outline

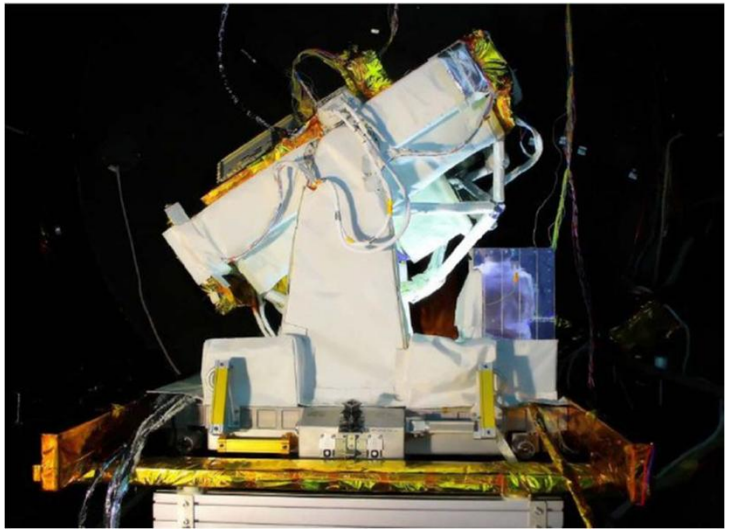
- **Space Applications Services NV (aka SpaceApps)** is an independent company whose aim is to be a leading provider of system and operations engineering as well as software engineering in the field of space and aerospace and to apply these capabilities to industrial applications.
- SpaceApps' expertise covers:
  - Space system engineering, specification, operations engineering, training and software development
  - Software Engineering
  - Research & Development
- SpaceApps' experience includes:
  - Control & Data Centers: complete ground segment and control centre solutions development & operation, for satellites & International Space Station (ISS) payloads.
  - Earth Observation Systems: semantic access to distributed EO data.
  - Knowledge Management: enterprise and scientific knowledge management solutions:

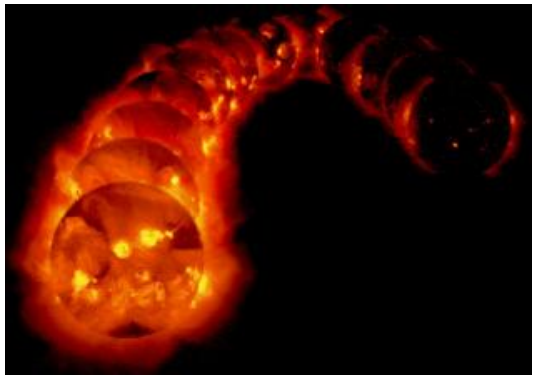
SAS Use Case

SAS Use Case

SAS Use Case

- External Payload installed on Columbus in February 2008.
- Integrated platform accommodating three instruments: SOVIM, SOLSPEC and SolACES.
- Measurement of the solar spectral irradiance throughout a large part of the electromagnetic spectrum.





- B.USOC (Belgian User Support and Operations Centre) ensure 24/7 operations support
- Team of 8 operators

SAS Use Case: Information Need

Forensic Analysis

A few months after the launch of the SOLAR payload, SOVIM, one of its three scientific instruments died because of an electric failure in a DC/DC converter. It is still unknown whether this failure could have been predicted given the previous telemetry stream. The objective of the CUBIST system would be to find patterns of failure in the flow of telemetry parameters with the aim to transpose these to the prediction of future failures.

SAS Use Case: Data Sources

- **Structured data sources**
  - Payload Telemetry
    - House keeping data (does not include Science data)
    - Processed parameters
    - 1 telemetry packet/second
    - 343 parameters/ telemetry packet

- **Unstructured data sources**
  - Columbus Operations Support Tools
  - System Problem reports
  - Payload Operations Data File
  - Daily Operations Report
  - SOLAR Predictor Tool
  - Local Bugs Database
  - Documentation

Slide with demo video, removed for th pdf-version of the slides

Content: SAS Current Analytics Demo

SAS Use Case: problems

- Typical queries/information needs
  - When was the earliest occurrence of SOVIM power status (SOLAR_PB3_28V_Out3) "ON" and SOVIM TM were halted or off nominal
  - Analyse correlations between errors and errors/platform TM/instrument TM/
- Problems
  - There is no single, unified interface for the SOLAR Operators to easily query all the relevant information and help predict & analyze instrument or payload failures
  - Today a lot of time and effort is spent on
    - Data or parameter retrieval
    - Post-analysis for both nominal operations and anomalies
    - Generation of supportive evidence for debriefing and decision making processes

SAS Use Case: Tool Need

As SOLAR Operators on console, we would like a **unified** tool (rather than multiple disconnected tools)

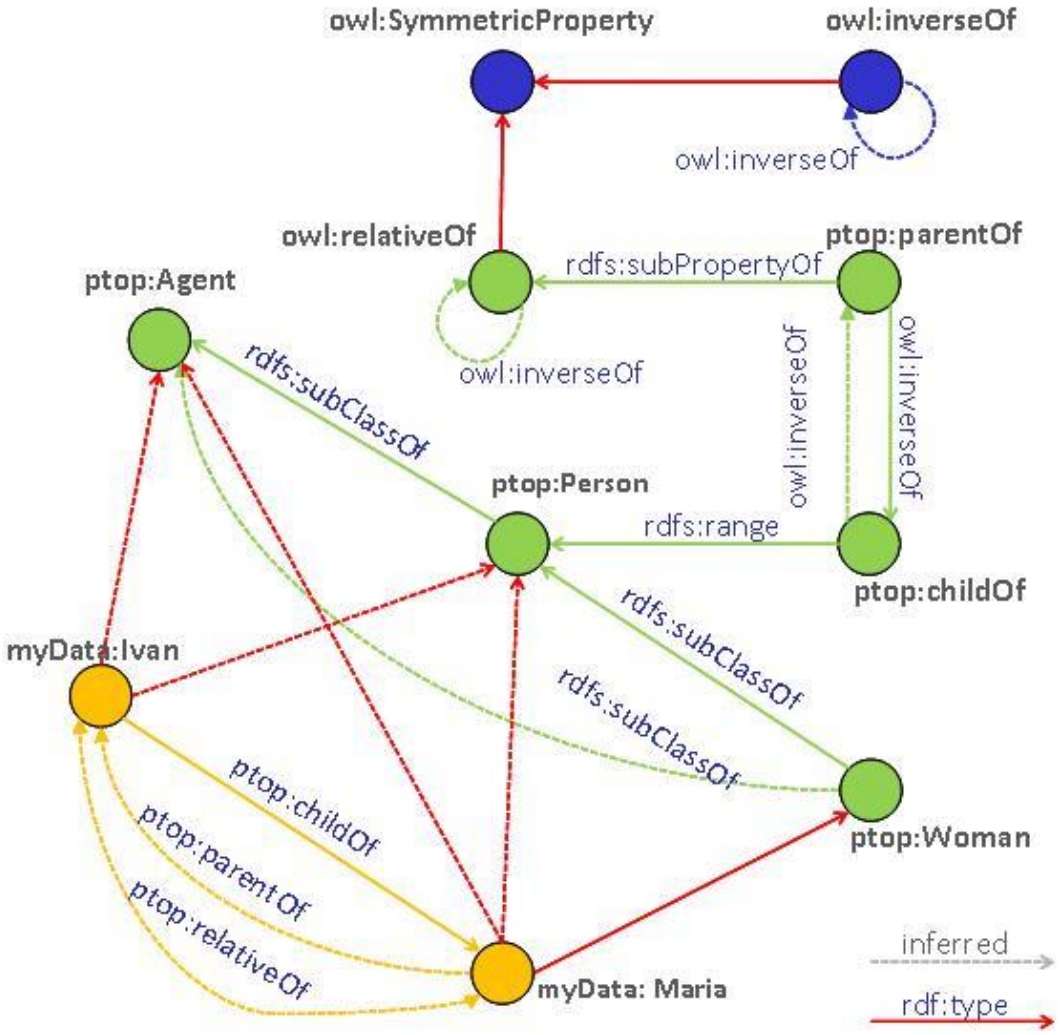
- exploiting structured telemetry data
- providing ways of visual analytics
- supporting us in the post-analysis and decision making

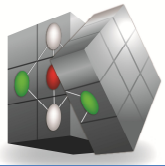
Agenda

- Project Setup and Key Technologies
  - First Introduction into CUBIST
  - Use Cases
  - Introduction into Semantic Technologies
  - Introduction into Formal Concept Analysis
  - Key Messages
- CUBIST Prototype
  - Architecture
  - Different Means to Access Information
  - Semantic Search
  - Query Generation
  - Explorative Search
  - Conceptual Scaling
  - Visual Analytics
- Outcome
  - User Evaluation
  - Our Take
  - Conclusions

Semantic Technologies

- Graph-based data model
  - (subject predicate object)
- Schema-free or schema-last approach
- (light-weight) reasoning
  - Hierarchy of types
  - Hierarchy of relations
  - Properties of relations





Let‘s borrow some slides …

The need for a smarter Web

- *"The Semantic Web is an extension of the current web in which information is given well-defined meaning, better enabling computers and people to work in cooperation."* (Tim Berners-Lee, 2001)

The Semantic Web vision (W3C)

- Extend principles of the Web from documents to data
- Data should be accessed using the general Web architecture (e.g., URI-s, protocols, ...)
- Data should be related to one another just as documents are already
- Creation of a common framework that allows:
  - Data to be shared and reused across applications
  - Data to be processed automatically
  - New relationships between pieces of data to be inferred

Ontologies as data models on the Semantic Web

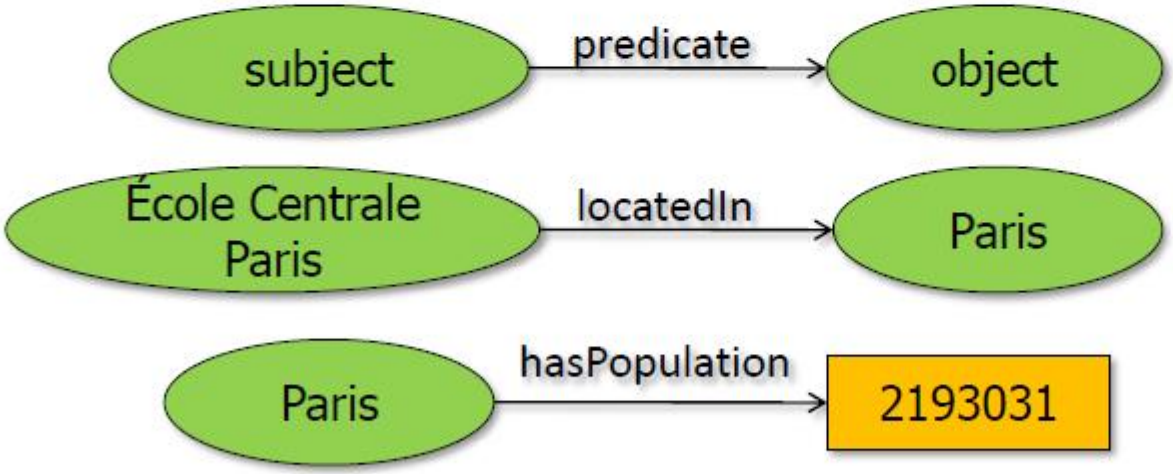
- An ontology is a *formal* specification that provides sharable and reusable knowledge representation
  - Examples – taxonomies, thesauri, topic maps, formal ontologies
- An ontology specification includes
  - Description of the *concepts* in some domain and their properties
  - Description of the possible *relationships* between concepts and the *constraints* on how the relationships can be used
  - Sometimes, the *individuals (*members of concepts)

Resource Description Framework (RDF)

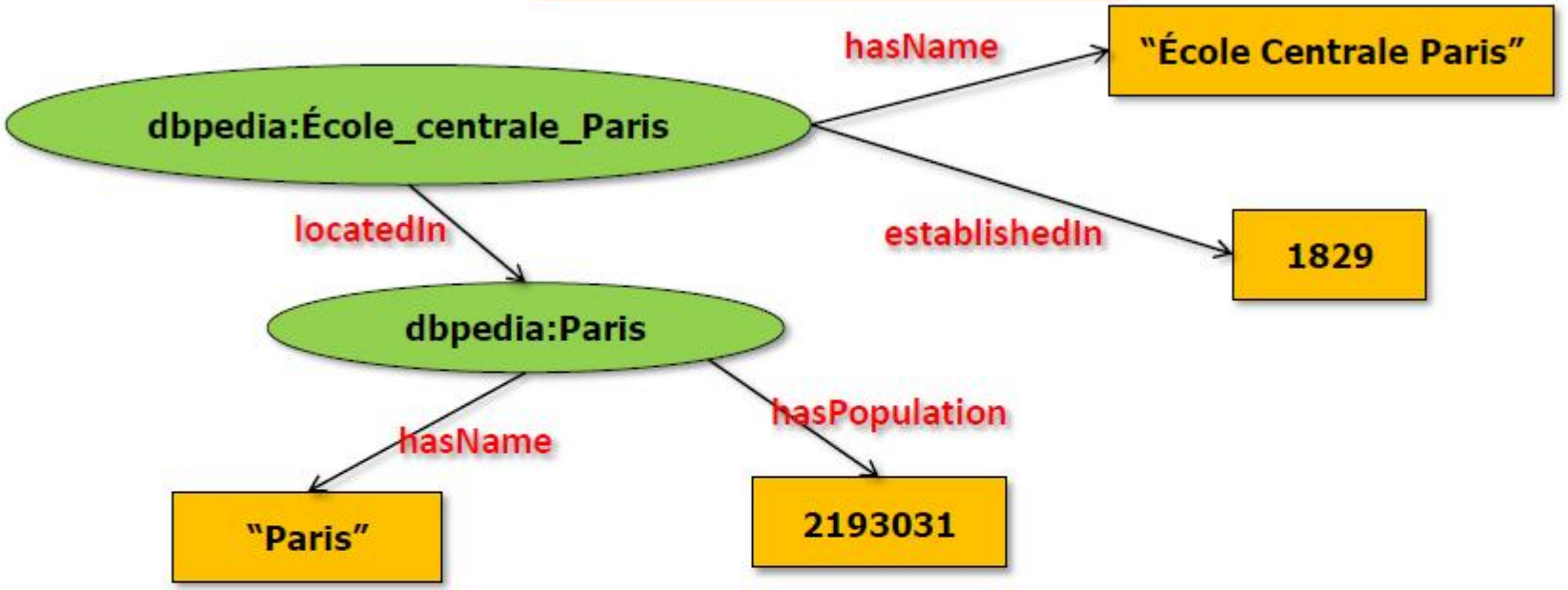
- A simple data model for
  - Formally describing the *semantics* of information
  - representing meta-data (data about data)
- A set of representation syntaxes
  - RDF/XML (standard), N-Triples, N3
- Building blocks
  - *Resources* (with unique identifiers)
  - *Literals*
  - Named *relations* between pairs of resources (or a resource and a literal)

RDF (2)

- Everything is a triple
  - **Subject** (resource), **Predicate** (relation), **Object** (resource or literal)
- The RDF graph is a collection of triples



RDF graph example (3)



Subject	Predicate	Object
http://dbpedia.org/resource/Paris	hasName	"Paris"
http://dbpedia.org/resource/Paris	hasPopulation	2193031
http://dbpedia.org/resource/École_centrale_Paris	locatedIn	http://dbpedia.org/resource/Paris
http://dbpedia.org/resource/École_centrale_Paris	hasName	"École Centrale Paris"
http://dbpedia.org/resource/École_centrale_Paris	establishedIn	1829

RDF advantages

- Global identifiers of all resources (URIs)
  - Reduces ambiguity
  - Makes incremental data integration easier
- Graph data model
  - Suitable for sparse, unstructured and semi-structured data
- Inference of implicit facts
- Schema agility
  - Lowers the cost of schema evolution

RDF Schema (RDFS)

- RDFS provides means for:
  - Defining *Classes* and *Properties*
  - Defining hierarchies (of classes and properties)
  - Domain/range of a property
- Entailment rules (axioms)
  - Infer new triples from existing ones

Web Ontology Language (OWL)

- More expressive than RDFS
  - Identity equivalence/difference
    - *sameAs, differentFrom, equivalentClass/Property*
- Complex class expressions
  - Class intersection, union, complement, disjointness
- More expressive property definitions
  - Object/Datatype properties
  - Cardinality restrictions
  - Transitive, functional, symmetric, inverse properties

SPARQL Protocol and RDF Query Language (SPARQL)

- SQL-like query language for RDF data
- Simple protocol for querying remote databases over HTTP
- Query types
  - *select* – query data by complex graph patterns
  - *ask* – whether a query returns results (result is true/false)
  - *describe* – returns all triples about a particular resource
  - *construct* – create new triples based on query results

From SPARQL 1.0 to SPARQL 1.1

- W3C recommendation
  - SPARQL 1.0: January 2008
  - SPARQL 1.1: March 2013
- HUGE step from 1.0 to 1.1
- New functionalities in SPARQL 1.1
  - **Aggregate functions**
  - **Subqueries**
  - Negation
  - Project expressions
  - Query language syntax
  - Property paths
  - Commonly used SPARQL functions
  - Basic federated query
- Aggregates, subqueries: Not used in CUBIST!

Linked Data

- *"To make the Semantic Web a reality, it is necessary to have a large volume of data available on the Web in a standard, reachable and manageable format. In addition the relationships among data also need to be made available. This collection of interrelated data on the Web can also be referred to as* ***Linked Data****. Linked Data lies at the heart of the Semantic Web: large scale integration of, and reasoning on, data on the Web."* (W3C)
- *Linked Data* is a set of principles that allows publishing, querying and browsing of RDF data, distributed across different servers
  - similar to the way HTML is currently published & consumed

SEMANTIC DATABASES (TRIPLESTORES)

Triplestores

- RDF databases
  - Store data according to the RDF data model
  - Provide inference of implicit triples (either at data loading time, or at query time)
  - SPARQL as a query language
- Many similarities to traditional DBMS approaches
  - … and many differences too

Triplestore advantages

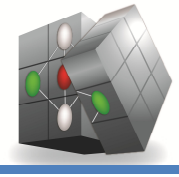
- Global identifiers of resources (entities)
  - Lowers the cost of data integration
- Inference of implicit facts
- Graph data model
  - Suitable for sparse, semi-structured and unstructured data
- Agile schema
  - New relations between entities may be easily added
- Exploratory queries against unknown schema
  - Query and data vocabulary may differ
- Compliance to standards (RDF, SPARQL)

Semantic Technologies & Triplestores for BI

- Speed-up data integration
  - RDF based ETL is more agile
- Lower the cost of data integration
  - Initial cost of using ontologies is higher
  - But the cost of ad-hoc ETL will be higher in the long term (too many data sources)
- Align & integrate legacy data silos
  - Querying & consuming data from disparate sources is easier with SPARQL

Semantic Technologies & Triplestores for BI (2)

- Infer implicit & hidden knowledge
  - Custom, user-defined rules as well
- Efficiently manage unstructured & semi-structured data together
  - graph data model
- Improve the quality of query results
  - Inference of implicit facts
  - SPARQL query vocabulary may differ from data vocabulary
  - Exploratory queries



Traditional BI vs BI in CUBIST
BO semantic layer vs CUBIST schema

"The semantic layer [in Business Objects products] is an abstraction layer between the database and the business user that frees the business user from the complexity of the data structures and technical names." *

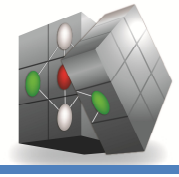
BI notion	CUBIST notion	comments
dimensions	classes or types	
measures, attributes	data properties, object properties	Measures in CUBIST can be numbers, dates, strings. • "raw" values are converted to context using conceptual scaling • FCA allows to combine different measures in one chart • Object properties can be used in CUBIST to analyze data as well, showing relationships (Clusters) between entities of different types
hierarchies	hierarchies of classes or properties	• In ST/CUBIST, we have hierarchies for types and properties • No need that hierarchies are trees. • Reasoning can be utilized
queries	analytics	

- Using ST, we essentially capture (apart from predefined calculations and functions) all notions of standard BI notions in the semantic layer
- in contrast to standard BI, we do not have two tiers (relational/star schema and a semantic layer on top of it). Instead, the schema of the repository directly serves as semantic layer

* http://www.sdn.sap.com/irj/scn/go/portal/prtroot/docs/library/uuid/c05314bb-e5a3-2e10-0e81-9e5a2db585df?QuickLink=index&overridelayout=true&51887500376956

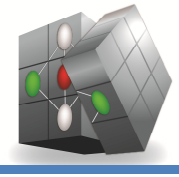
Agenda

- Project Setup and Key Technologies
  - First Introduction into CUBIST
  - Use Cases
  - Introduction into Semantic Technologies
  - Introduction into Formal Concept Analysis
  - Key Messages
- CUBIST Prototype
  - Architecture
  - Different Means to Access Information
  - Semantic Search
  - Query Generation
  - Explorative Search
  - Conceptual Scaling
  - Visual Analytics
- Outcome
  - User Evaluation
  - Our Take
  - Conclusions



What is Formal Concept Analysis?

- Formal Concept Analysis is the main means in CUBIST to analyze data.
- FCA is best suited for **qualitative** data analysis
  - It does not particularly target **quantitative** data analysis
  - But quantitative data analysis can be covered by FCA



FCA in three Minutes (i)

How can we describe the concept "BI products from SAP"?

- Extensionally by enumerating all ***objects***:
  - BO Xcelsius, BO Crystral Reports, …
- Intensionally through ***attributes***:
  - "is an SAP product", "is a BI tool", …

Generally, a ***concept*** is divided into two mutually dependent parts:

- Its ***extension*** are all objects that share all the attributes of the concept,
- Its ***intension*** are the attributes which precisely describe the objects of the concept.

The concepts form a hierarchy: A concept C1 is a ***subconcept*** of C2, iff

- the extension of C1 is a subset of the extension of C2
- the intension of C2 is a subset of the extension of C1

} equivalent

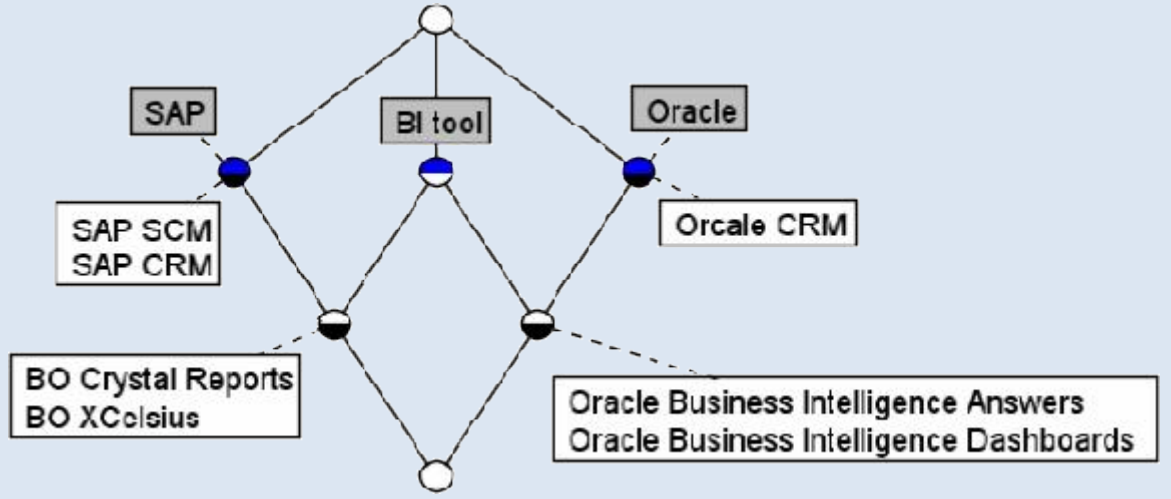
Theorem: For a given universe, the concept hierarchy is a complete lattice

FCA in three Minutes (ii)

A toy **formal context**

	SAP	Oracle	BI tool
BO XCelsius	X		X
BO Crystal Reports	X		X
SAP CRM	X		
SAP SCM	X		
Oracle Business Intelligence Dashboards		X	X
Oracle Business Intelligence Answers		X	X
Orcale CRM		X	
Oracle Fusion GRC		X	

Its derived **concept lattice**



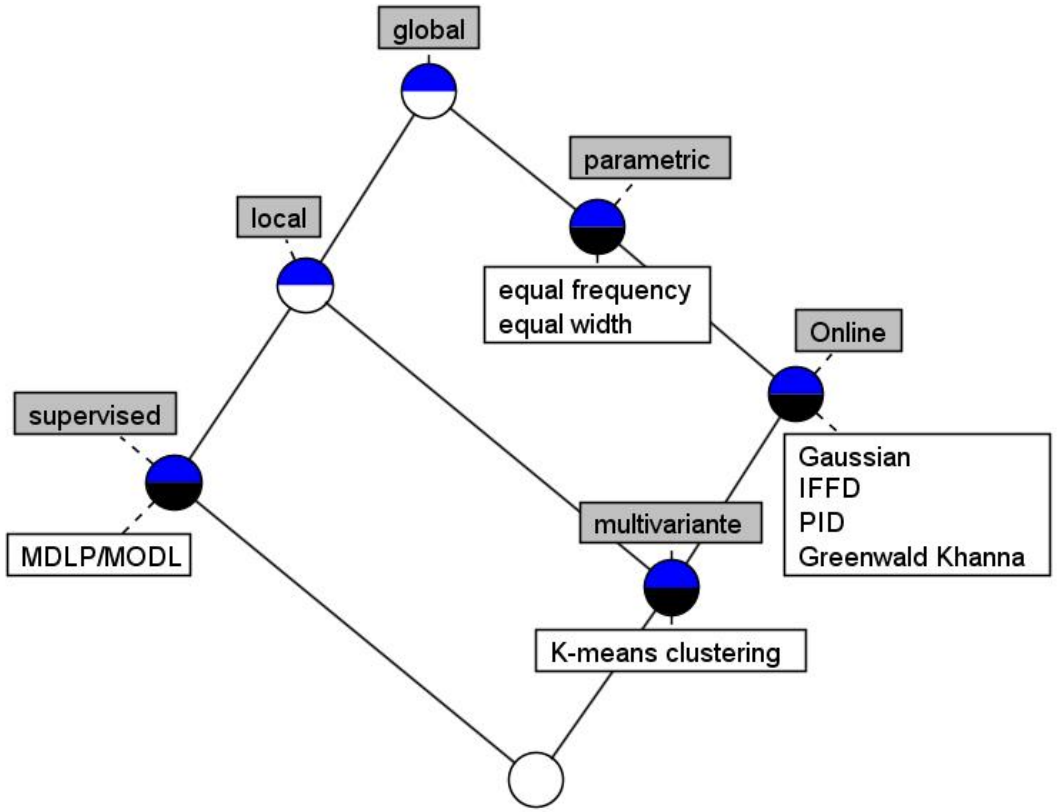
Example from Yesterday

Methods comparison

Method	Global / local	Multivariate	Parametric	Supervised	Online / stream
Equal Width	Global	No	Yes	No	No
Equal Freq	Global	No	Yes	No	No
Greenwald Khanna	Global	No	Yes	No	Yes
K-means clustering	Global and local	Yes	Yes	No	Yes / No
PID (Layer 1)	Global	No	Yes	No	Yes
MDLP / MODL	Global and local	No	No	Yes	No
IFFD	Global	No	Yes	No	Yes / No
Gaussian	Global	No	Yes	No	Yes

3 criteria were proposed by: Dougherty J, Kohavi R, Sahami M.
Supervised and unsupervised discretization of continuous features. ML199

A	B	C	D	E	F	G	H
	global	local	multivariante	parametric	supervised	Online	stream
equal width	X			X			
equal frequency	X			X			
Greenwald Khanna	X			X		X	X
K-means clustering	X	X	X	X		X	
PID	X			X		X	X
MDLP/MODL	X	X			X		
IFFD	X			X		X	
Gaussian	X			X		X	X



Small, Real Example Context: Feature Comparison Matrix

The table below is to be visualized as a concept lattice.

Table 1: SAP® Crystal Reports® Software Comparison Matrix

High-Productivity Report Creation Features	2008 and 2011	Version for Visual Studio 2010	Version for Eclipse	XI	10	9	8.5
Wizards and experts for report creation	D	D	D	D,P,S	A,D,P,S	A,D,P,S	D,P,S
Database expert for graphical table linking	D	D	D	D,P,S	A,D,P,S	A,D,P,S	D,P,S
Field explorer to manage report fields	D	D	D	D,P,S	A,D,P,S	A,D,P,S	D,P,S
Drill down in runtime	D	D	D	D,P,S	A,D,P,S	A,D,P,S	D,P,S
Autosave	D			D,P,S	A,D,P,S	A,D,P,S	D,P,S
Editable preview window	D		D	D,P,S	A,D,P,S	A,D,P,S	D,P
Browse field data	D	D	D	D,P,S	A,D,P,S	A,D,P,S	D,P,S
Move, resize, and multiselect objects	D	D	D	D,P,S	A,D,P,S	A,D,P,S	D,P,S
Custom templates	D			D,P,S	A,D,P,S	A,D,P,S	
Repository for component reuse	D			D,P	A,D,P	A,D,P	
Workbench tool for managing projects	D			D,P,S			
Automatic patch notification and installation	D	D	D	D,P,S			
Start page to stay connected with latest information	D	D	D	D,P,S			
HTML preview	D	D	D	D,P			
Extension points for custom wizards, toolbar buttons, and designer events			D				
Extension points for toolbar buttons	D						
Multilingual report design and view	D						
Dual monitor support	D	D	D				

SAP® Crystal Reports® Versions

Key Editions:
Advanced Developer, Developer, Professional, Standard

Version 8.5		D	P	S
Version 9	A	D	P	S
Version 10	A	D	P	S
Version XI		D	P	S
Version 2008 and 2011		D		
Version 2011 only		D 2011		
Version 2008 only		D 2008		

Source: **Comparison of features by version** for SAP Crystal Reports and SAP Crystal Server Software. Pdf-brochure, www.sap.com

A Feature Matrix is simply a Binary Relation

Table 1: SAP® Crystal Reports® Software Comparison Matrix

High-Productivity Report Creation Features	2008 and 2011	Version for Visual Studio 2010	Version for Eclipse	XI	10	9	8.5
Wizards and experts for report creation	D	D	D	D,P,S	A,D,P,S	A,D,P,S	D,P,S
Database expert for graphical table linking	D	D	D	D,P,S	A,D,P,S	A,D,P,S	D,P,S
Field explorer to manage report fields	D	D	D	D,P,S	A,D,P,S	A,D,P,S	D,P,S
Drill down in runtime	D	D	D	D,P,S	A,D,P,S	A,D,P,S	D,P,S
Autosave	D			D,P,S	A,D,P,S	A,D,P,S	D,P,S
Editable preview window	D		D	D,P,S	A,D,P,S	A,D,P,S	D,P
Browse field data	D	D	D	D,P,S	A,D,P,S	A,D,P,S	D,P,S
Move, resize, and multiselect objects	D	D	D	D,P,S	A,D,P,S	A,D,P,S	D,P,S
Custom templates	D			D,P,S	A,D,P,S	A,D,P,S	
Repository for component reuse	D			D,P	A,D,P	A,D,P	
Workbench tool for managing projects	D			D,P,S			
Automatic patch notification and installation	D	D	D	D,P,S			
Start page to stay connected with latest information	D	D	D	D,P,S			
HTML preview	D	D	D	D,P			
Extension points for custom wizards, toolbar buttons, and designer events			D				
Extension points for toolbar buttons	D						
Multilingual report design and view	D						
Dual monitor support	D	D	D				

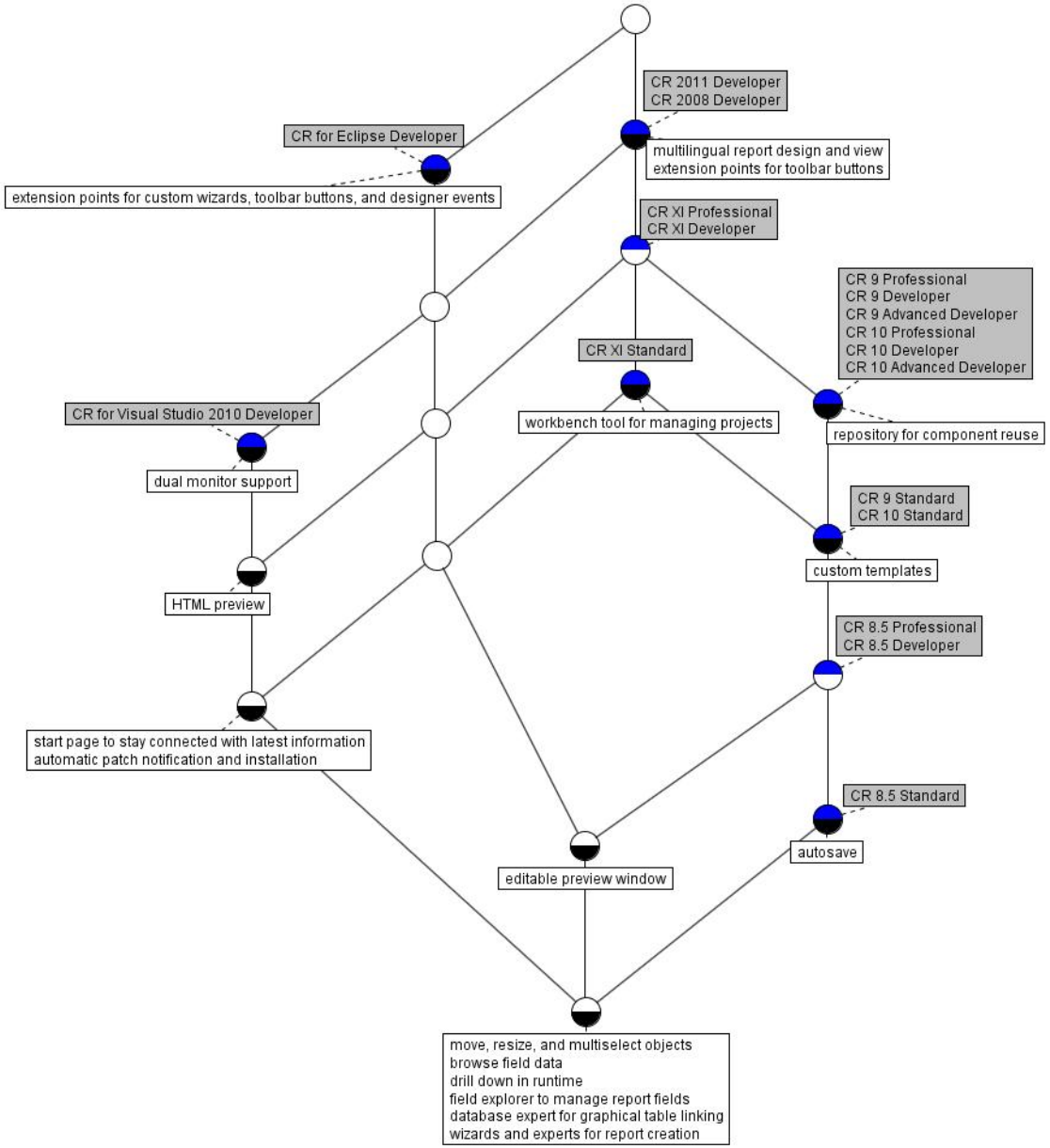
SAP® Crystal Reports® Versions

Key Editions:
Advanced Developer, Developer, Professional, Standard

Version 8.5		D	P	S
Version 9	A	D	P	S
Version 10	A	D	P	S
Version XI		D	P	S
Version 2008 and 2011		D		
Version 2011 only		D 2011		
Version 2008 only		D 2008		

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
	CR 2008 Developer	CR ...	CR f...	CR fo...	CR XI...	CR X...	CR XI...	CR 1...	CR 10 ...	CR 10 ...	CR 1...	CR 9 ...	CR 9 ...	CR 9 P...	CR 9 ...	CR 8....	CR 8.5...	CR 8....
wizards and experts for report creation	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
database expert for graphical table linking	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
field explorer to manage report fields	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
drill down in runtime	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
autosave	X	X			X	X	X	X	X	X	X	X	X	X	X	X	X	X
editable preview window	X	X		X	X	X	X	X	X	X	X	X	X	X	X	X	X	
browse field data	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
move, resize, and multiselect objects	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
custom templates	X	X			X	X	X	X	X	X	X	X	X	X	X			
repository for component reuse	X	X			X	X		X	X	X		X	X	X				
workbench tool for managing projects	X	X			X	X	X											
automatic patch notification and installation	X	X	X	X	X	X	X											
start page to stay connected with latest information	X	X	X	X	X	X	X											
HTML preview	X	X	X	X	X	X												
extension points for toolbar buttons	X	X																
multilingual report design and view	X	X																
dual monitor support	X	X	X	X														
extension points for custom wizards, toolbar buttons, and designer eve...				X														

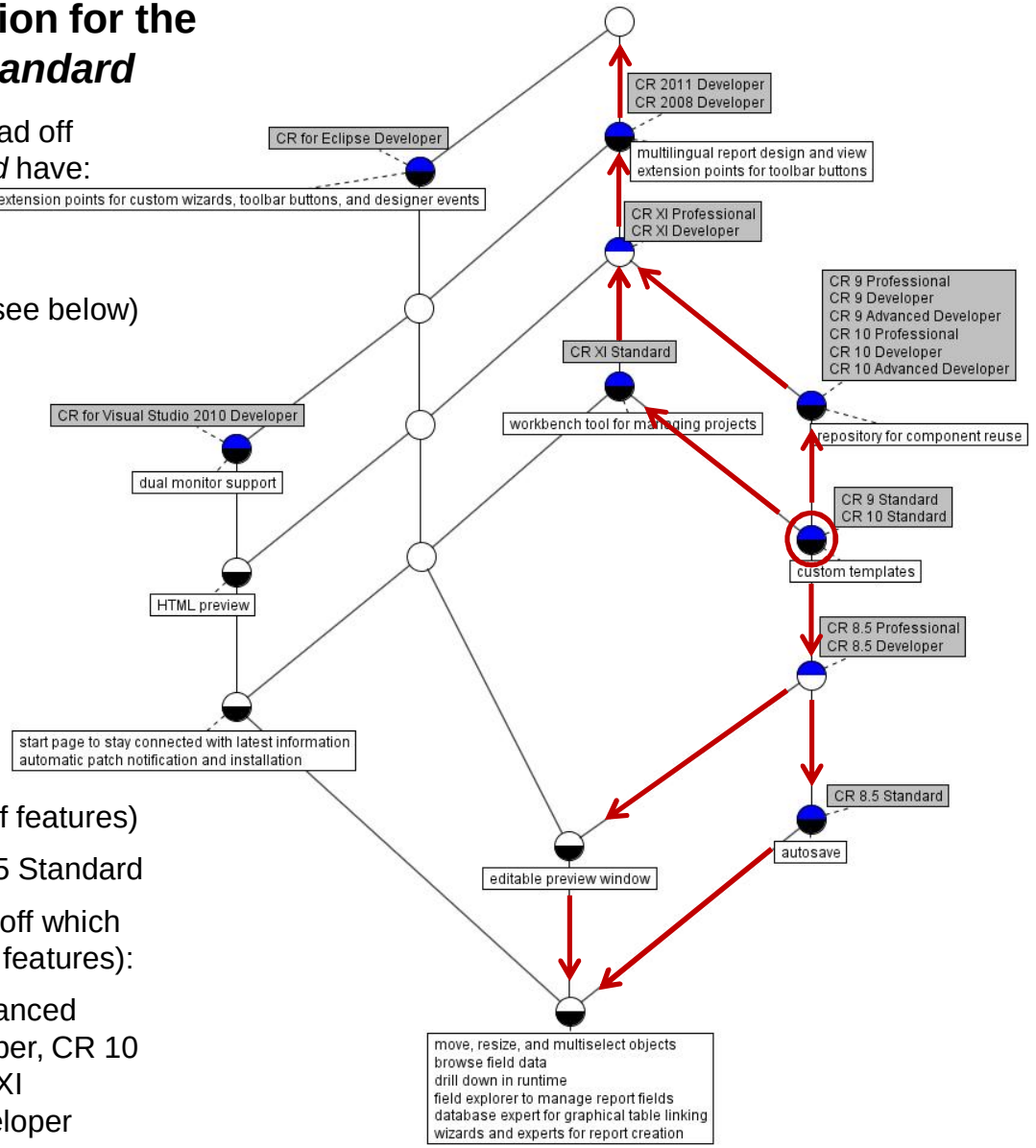
Feature Comparison Matrix: Concept Lattice



Feature Comparison Matrix: Reading the Concept Lattice

Here is how you read off the information for the versions *CR 9 Standard* and *CR 10 Standard*

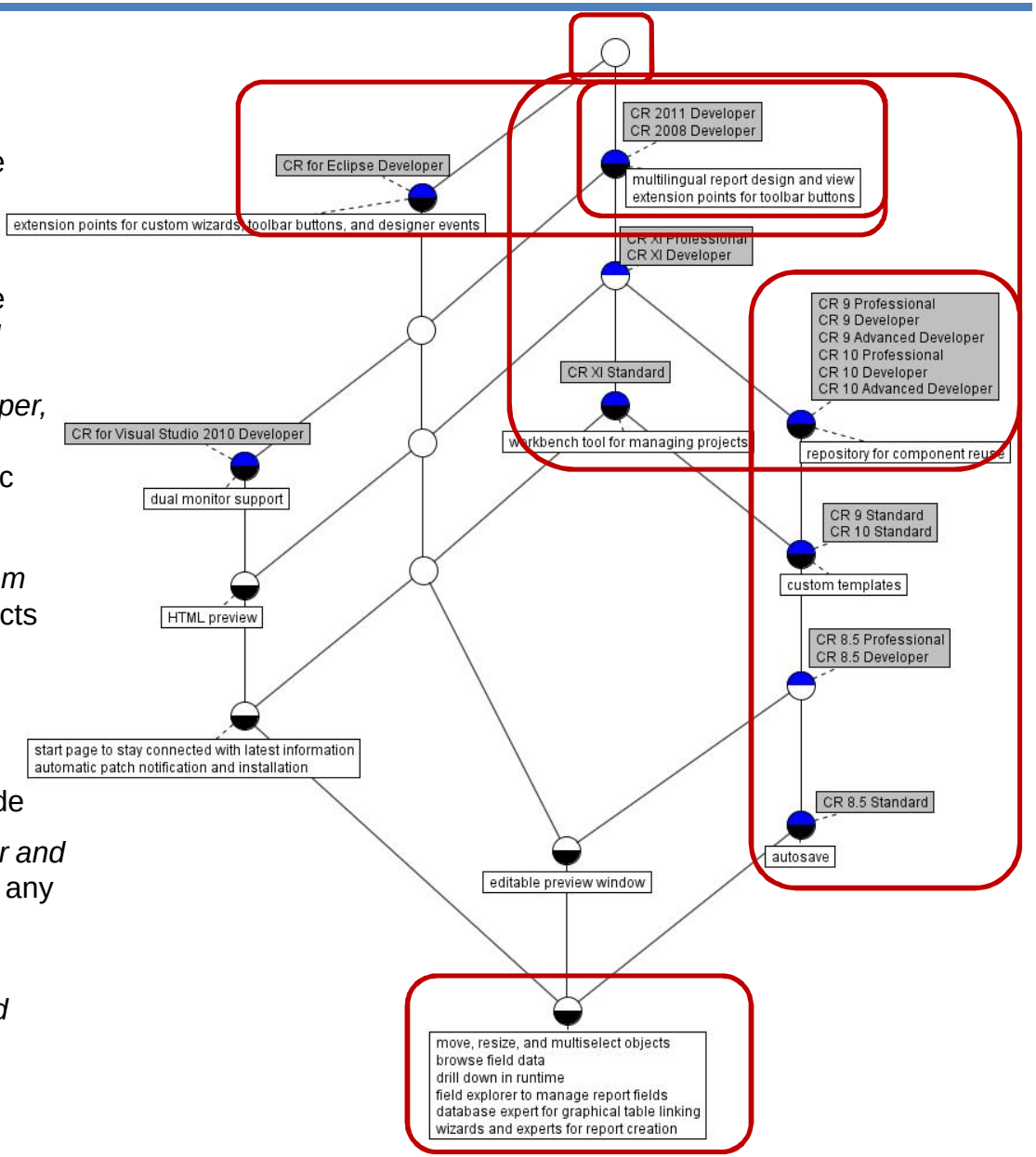
- Following all possible paths downwards, we can read off which features *CR 9 Standard* and *CR 10 Standard* have:
  - custom templates
    - indeed the distinguishing feature of these versions, compared to “weaker” versions (see below)
  - Editable preview window
  - Autosave
  - Move, resize, and multiselect objects;
  - Browse field data
  - Drill down in runtime
  - Field explorer to manage report fields
  - Database expert for graphical table linking
  - Wizards and experts for report creation
- Following all possible paths downwards, we can read off versions are weaker (i.e., have a subset of features)
  - CR 8.5 Professional, CR 8.5 Developer, CR 8.5 Standard
- Following all possible paths upwards, we can read off which versions are stronger (i.e., they have a superset of features):
  - CR 9 Professional, CR 9 Developer, CR 9 Advanced Developer, CR 10 Professional, CR 10 Developer, CR 10 Advanced Developer, CR XI Professional, CR XI Developer, CR 2008 Developer, CR 2011 Developer



Feature Comparison Matrix: Reading the Concept Lattice

Some more things one can read off

- *CR 2011 Developer* and *CR 2008 Developer* have exactly the same features
  - Because they are on the same node
- *CR 2011 Developer* and *CR 2008 Developer* have more features than *CR XI Professional* and *CR XI Developer*, which in turn have more features than *CR XI Standard, CR 9 Professional, CR 9 Developer, CR 9 Advanced Developer, CR 10 Professional, CR 10 Developer, CR 10 Advanced Developer*, etc
  - Reading the lattice downwardly
- *Autosave* is featured in more products than *Custom templates*, which in turn is featured in more products *than repository for component reuse*, etc
  - Reading the lattice upwardly
- There is no product having all features
  - As there is no product name on the top node
- But *CR for Eclipse Developer, CR 2011 Developer and CR 2008 Developer* are the best products (i.e. for any of those, there is no product with a superset of features)
- *Move, resize, and multiselect objects, browse field data*, etc are featured in all products

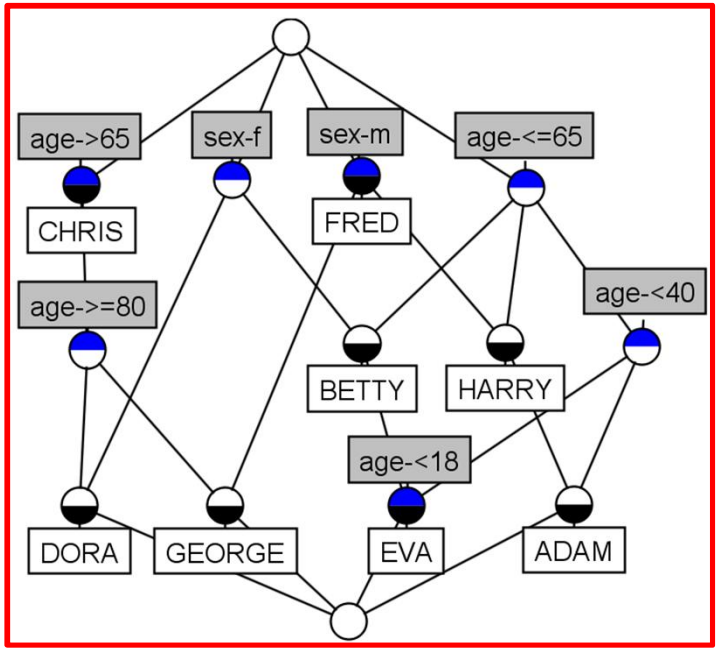


Conceptual Scaling
From many-valued to single-valued contexts

- FCA genuinely deals with boolean data only
- Conceptual scaling is a means to “translate” non-boolean data attributes if entites into formal contexts
- Conceptual scales can be manually or semi-automatically created
- Example: Entities with two data-properties
  - sex (two values, nonimal data)
  - age (integer, ordinal data)

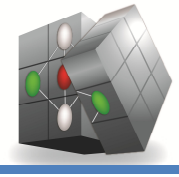
	sex	age
ADAM	m	21
BETTY	f	50
CHRIS	?	66
DORA	f	88
EVA	f	17
FRED	m	?
GEORGE	m	90
HARRY	m	60

	sex		age				
	m	f	*<18*	*<40*	*<=65*	*>65*	*>=80*
ADAM	X			X	X		
BETTY		X			X		
CHRIS						X	
DORA		X				X	X
EVA		X	X	X	X		
FRED	X						
GEORGE	X					X	X
HARRY	X				X		



Agenda

- Project Setup and Key Technologies
  - First Introduction into CUBIST
  - Use Cases
  - Introduction into Semantic Technologies
  - Introduction into Formal Concept Analysis
  - Key Messages
- CUBIST Prototype
  - Architecture
  - Different Means to Access Information
  - Semantic Search
  - Query Generation
  - Explorative Search
  - Conceptual Scaling
  - Visual Analytics
- Outcome
  - User Evaluation
  - Our Take
  - Conclusions



What the next slides are about …

The next slides provide a few thoughts on different kinds of analyzing some data, in order to compare the following **Visual Analytics** means:

1. Traditional BI Visual means (here: a bar chart)
2. A graph-based visualization (here: force-based layout)
3. A visualization based on Formal Concept Analysis (here: concept lattices)

Toy Example Data Set

Skill	Persons with that Skill
IE	Anja, Ben, Ernst, Fred, Ken
ETL	Chris, Fred, Mark
BI	Ben, Chris, Fred, Lemmy, Mark, Naomi
ST	Anja, Diana, Ernst, Fred, Gerald, Harriet, Ken, Owen
FCA	Anja, Diana, Gerald, Harriet, Ian, John, Ken, Owen
VIZ	Anja, Diana, Ian

Possible Information Needs:

1. Show me the count of people for a given skill
2. Show me the skills and how many people share some skills, in order to get an idea on how strongly skills are related
3. Show me the skills and people such that I get an idea of the distribution of skills among people and dependencies between skills

Converting the Data (Analytic Model)

Raw Data

Skill	Persons with that Skill
IE	Anja, Ben, Ernst, Fred, Ken
ETL	Chris, Fred, Mark
BI	Ben, Chris, Fred, Lemmy, Mark, Naomi
ST	Anja, Diana, Ernst, Fred, Gerald, Harriet, Ken, Owen
FCA	Anja, Diana, Gerald, Harriet, Ian, John, Ken, Owen
VIZ	Anja, Diana, Ian

Bar Chart Data

Skill	#People
IE	5
ETL	3
BI	6
ST	8
FCA	8
VIZ	3

Counting the number of people per skill

Graph Data

	IE	ETL	BI	ST	FCA	VIZ
IE						
ETL	1					
BI	1	3				
ST	4	1	1			
FCA	2	0	0	5		
VIZ	1	0	0	2	3	

Counting the number of people who share two skills

FCA Data (Formal Context)

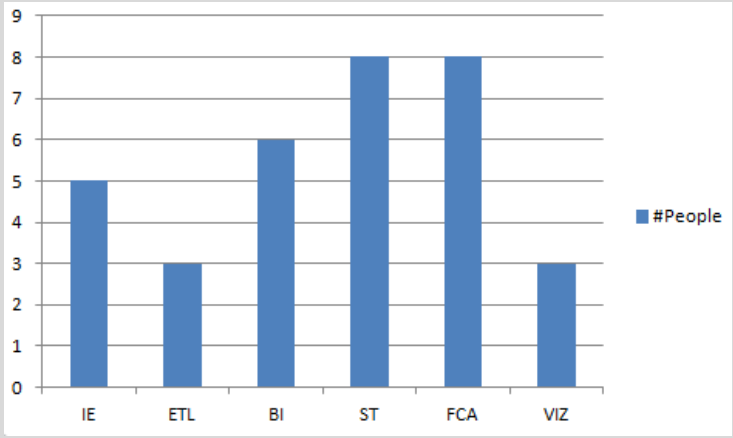
	IE	ETL	BI	ST	FCA	VIZ
Anja	X			X	X	X
Ben	X		X			
Chris		X	X			
Diana				X	X	X
Ernst	X			X		
Fred	X	X	X	X		
Gerald				X	X	
Harriet				X	X	
Ian					X	X
John					X	
Ken	X			X	X	
Lemmy			X			
Mark	X	X	X			
Naomi	X		X			
Owen				X	X	

Visualizing the Data

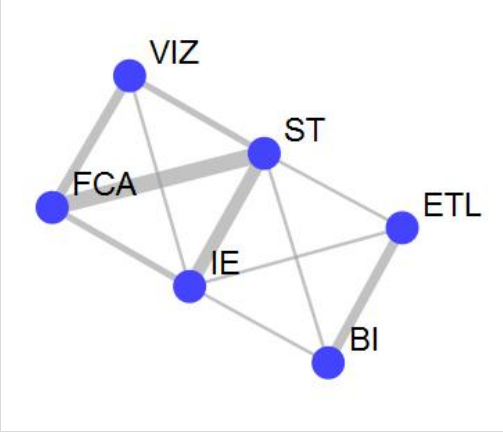
Raw Data

Skill	Persons with that Skill
IE	Anja, Ben, Ernst, Fred, Ken
ETL	Chris, Fred, Mark
BI	Ben, Chris, Fred, Lemmy, Mark, Naomi
ST	Anja, Diana, Ernst, Fred, Gerald, Harriet, Ken, Owen
FCA	Anja, Diana, Gerald, Harriet, Ian, John, Ken, Owen
VIZ	Anja, Diana, Ian

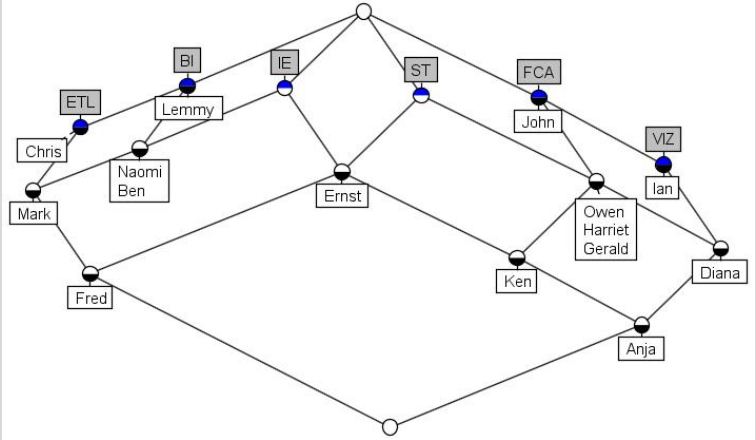
Bar Chart



Graph

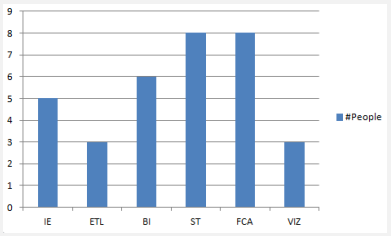


FCA Concept Lattice



Comparison

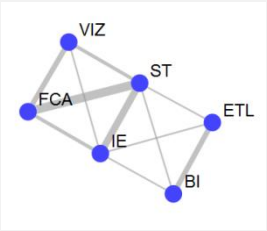
Bar Chart



- Many well-known visualizations
- Good (readable and comprehensible) layouts
- Good for analyzing numbers

- Loss of information (what people)
- Misleading for overlapping attributes (counting people manyfold)
- Not utilizing relationships between entities

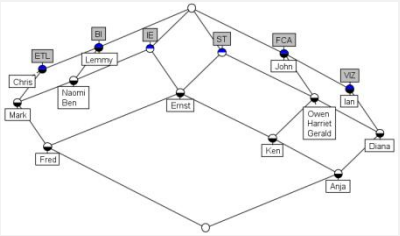
Graph



- Attractive visualizations
- (Relatively) easy to understand
- Utilizing and showing links between entities (skills)

- Loss of information (what people)
- Bad for analyzing numbers

FCA lattice



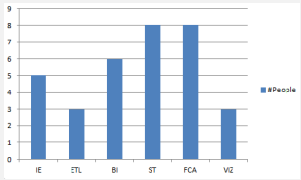
- No loss of information
- Meaningful clusters in one node
- Showing dependencies between entities (both people and skills)

- Number of nodes might explode
- Finding good layout is unsolved (nice layout in example is accidential and has been manually created)
- Unfamiliar means for analytics
- Scalability
- Bad for analyzing numbers

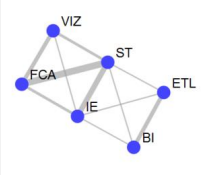
General Conclusion

Remember the information needs from the beginning

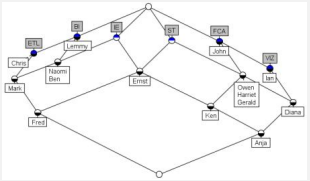
Show me the count of people for a given skill



Show me the skills and how many people share some skills, in order to get an idea on how strongly skills are related



Show me the skills and people such that I get an idea of the distribution of skills among people and dependencies between skills



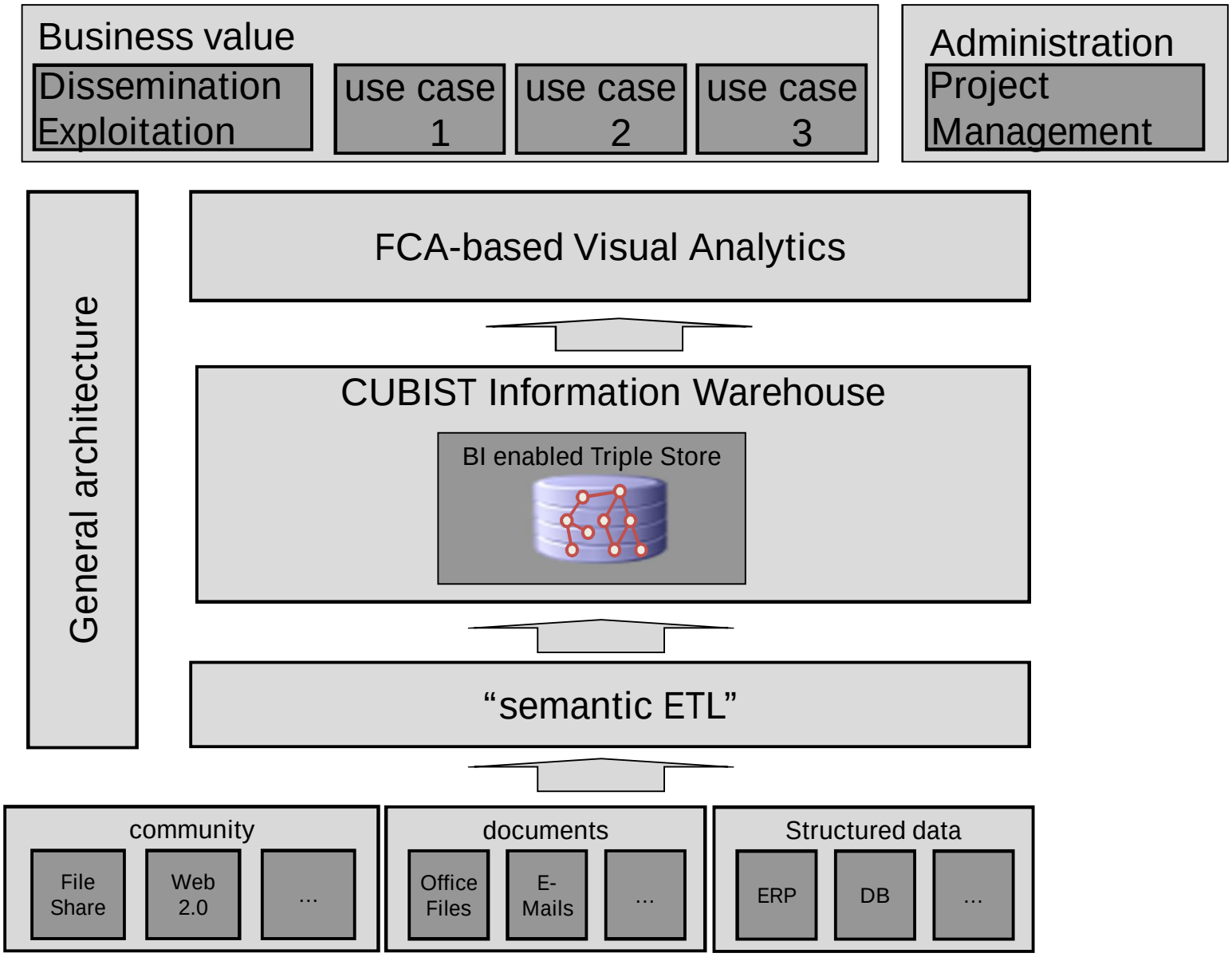
Conclusion

- Each visualization has ist own strengths and weaknesses
- Each type of visualization is suited for a specific type of information needs
- Thus the visualizations are complementing
- Thus future BI tools should provide all types of visualizations
  - For example, side by side with linking-and-brushing

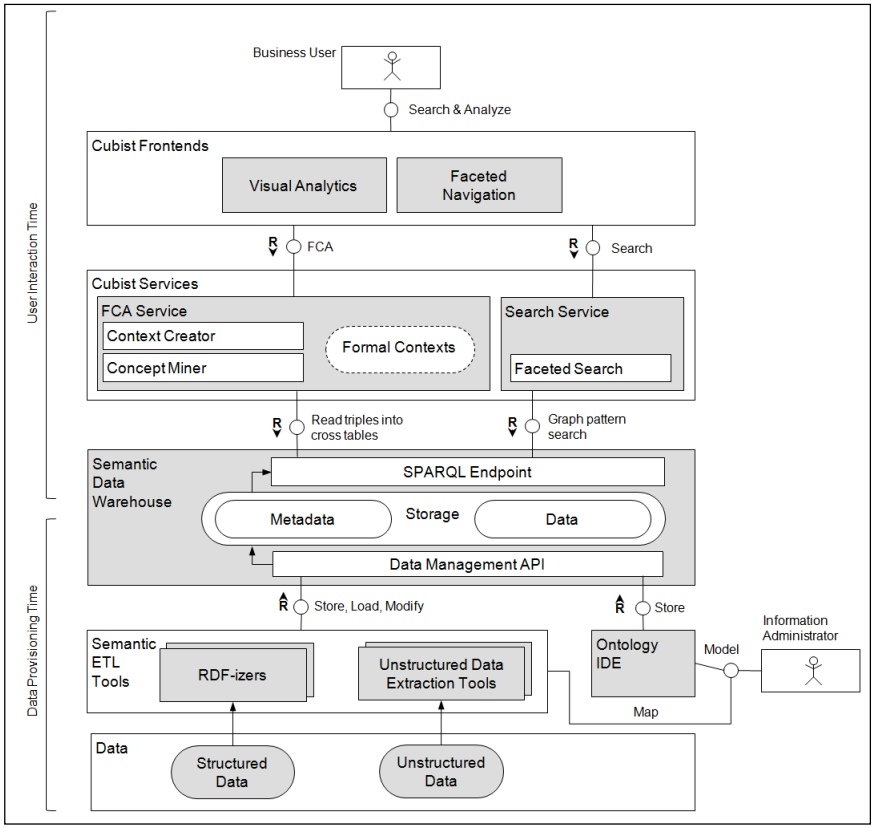
Agenda

- Project Setup and Key Technologies
  - First Introduction into CUBIST
  - Use Cases
  - Introduction into Semantic Technologies
  - Introduction into Formal Concept Analysis
  - Key Messages
- CUBIST Prototype
  - Architecture
  - Different Means to Access Information
  - Semantic Search
  - Query Generation
  - Explorative Search
  - Conceptual Scaling
  - Visual Analytics
- Outcome
  - User Evaluation
  - Our Take
  - Conclusions

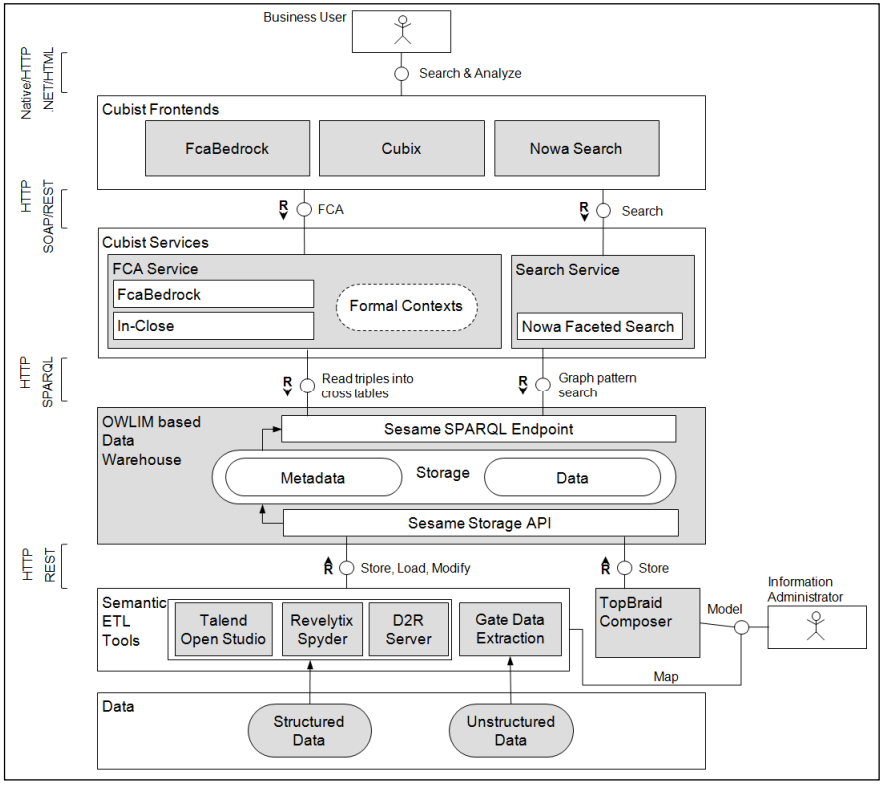
CUBIST Highlevel Architecture



CUBIST Prototype Architecture

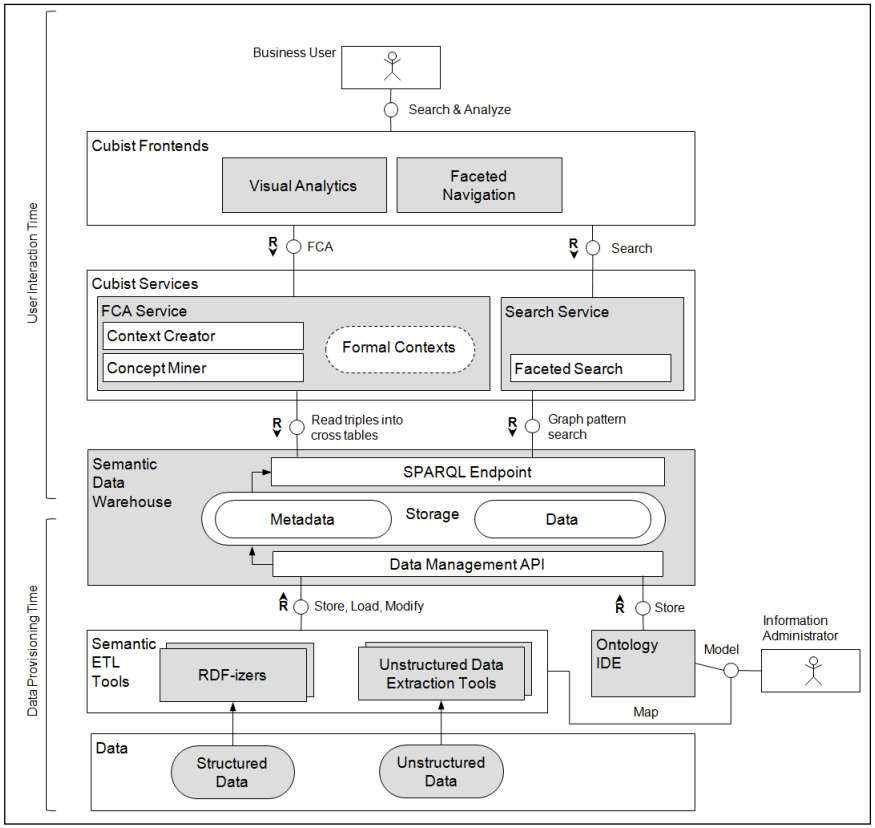


Reference Architecture

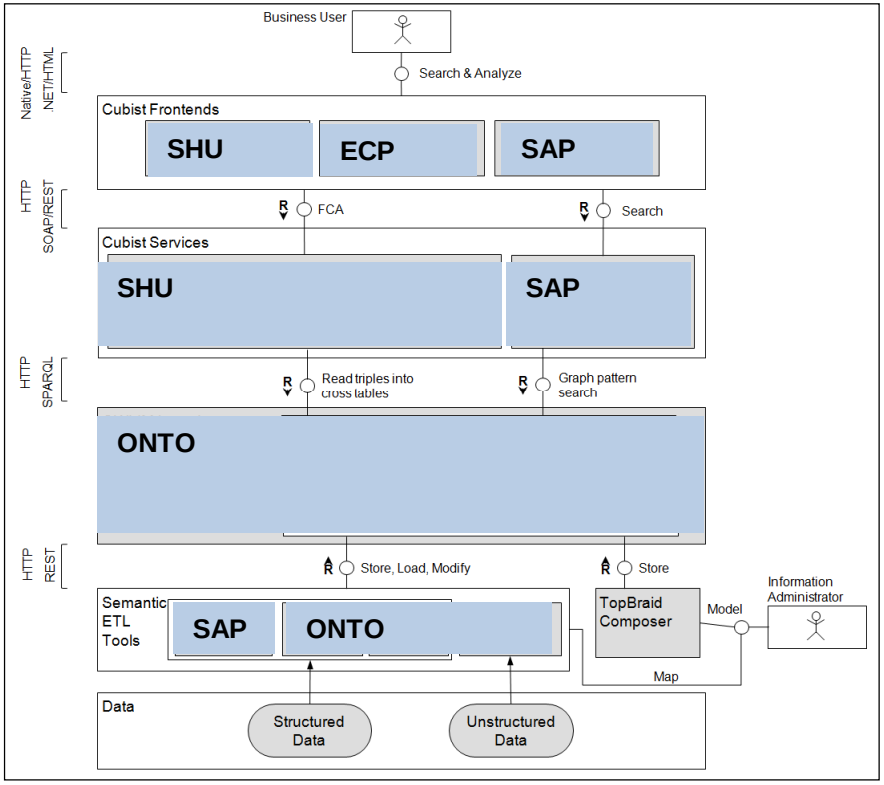


Implementation Architecture

CUBIST Prototype Architecture
Partner Contributions



Reference Architecture



Implementation Architecture

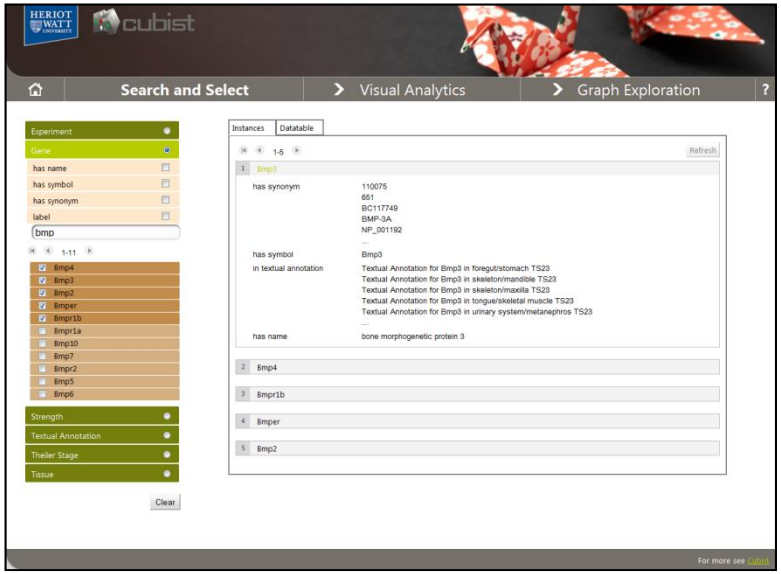
Agenda

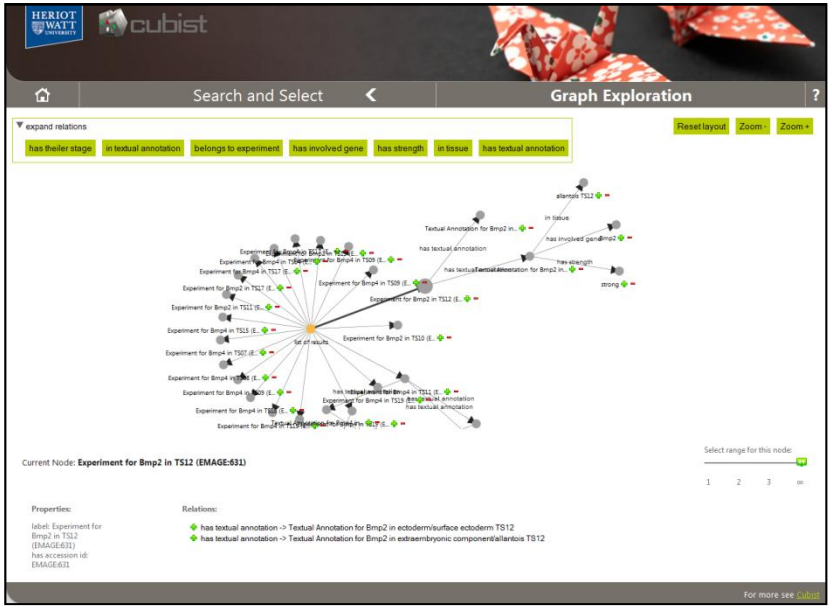
- Project Setup and Key Technologies
  - First Introduction into CUBIST
  - Use Cases
  - Introduction into Semantic Technologies
  - Introduction into Formal Concept Analysis
  - Key Messages
- CUBIST Prototype
  - Architecture
  - Different Means to Access Information
  - Semantic Search
  - Query Generation
  - Explorative Search
  - Conceptual Scaling
  - Visual Analytics
- Outcome
  - User Evaluation
  - Our Take
  - Conclusions

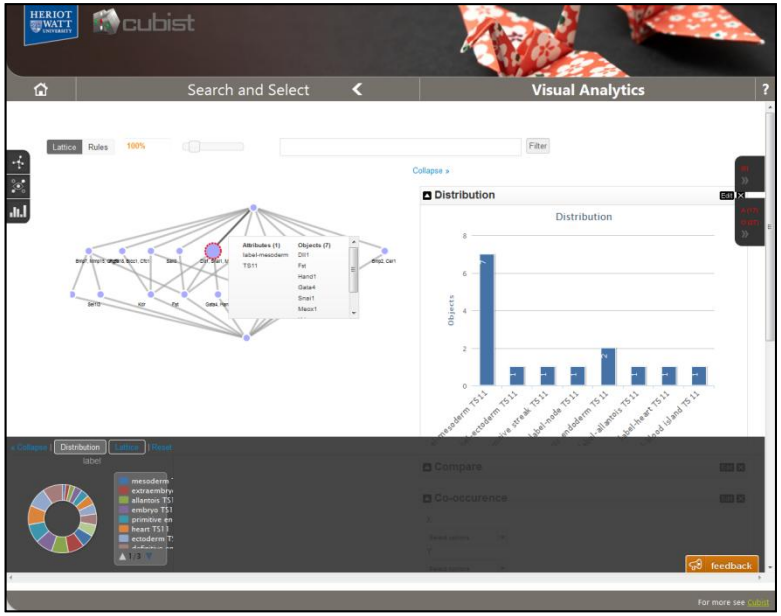
CUBIST Functionalities

Comprehensive Information Access Means

- **factual search**
  searching for specific entities
- **explorative search**
  exploring the information space
- **visual analytics**
  analyzing sets of entities, with traditional and novel diagrams

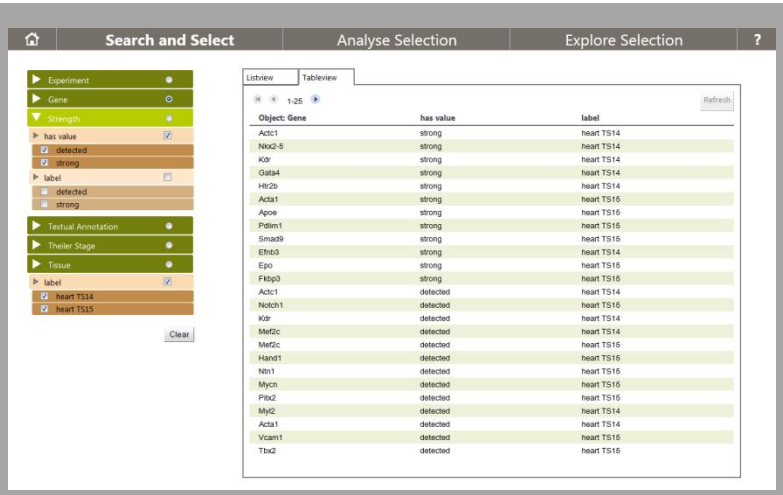




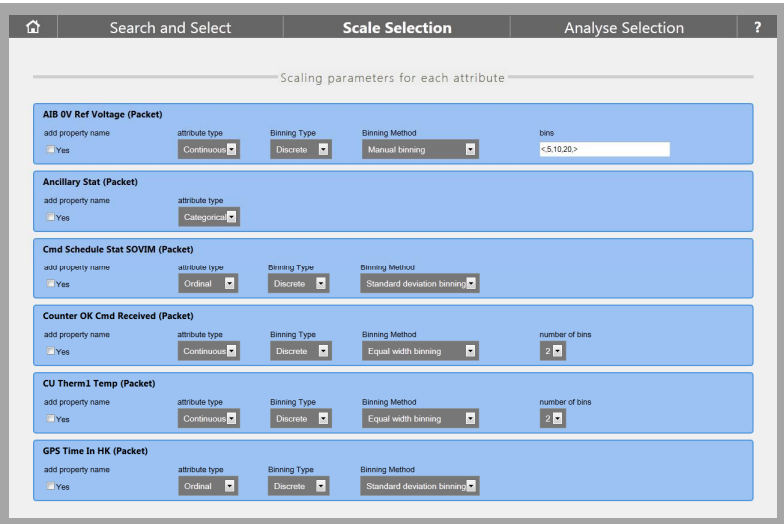


CUBIST Functionalities

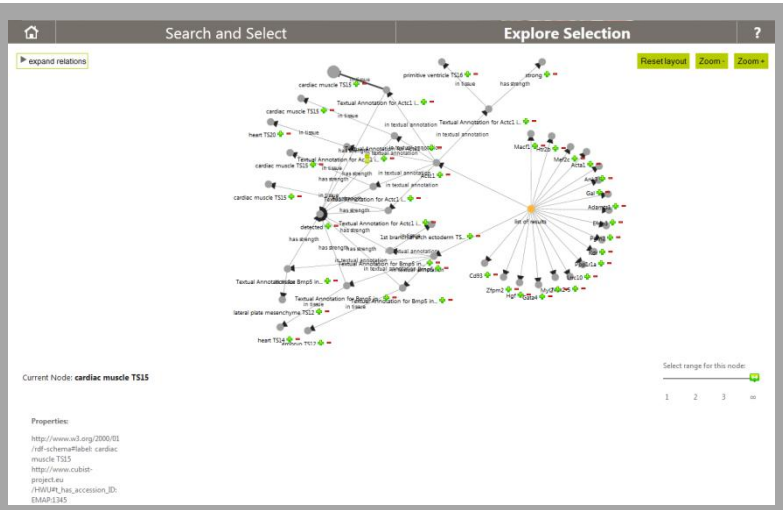
Comprehensive Information Access Means



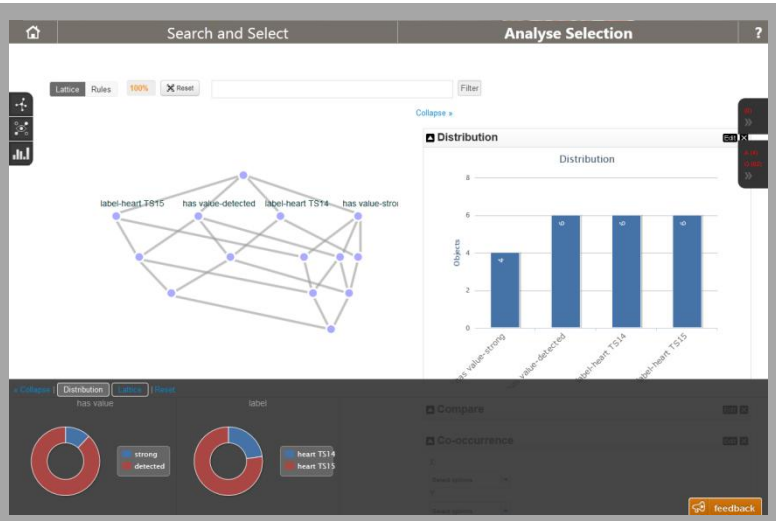
extended faceted & sem. search



conceptual scaling



graph-based exploration

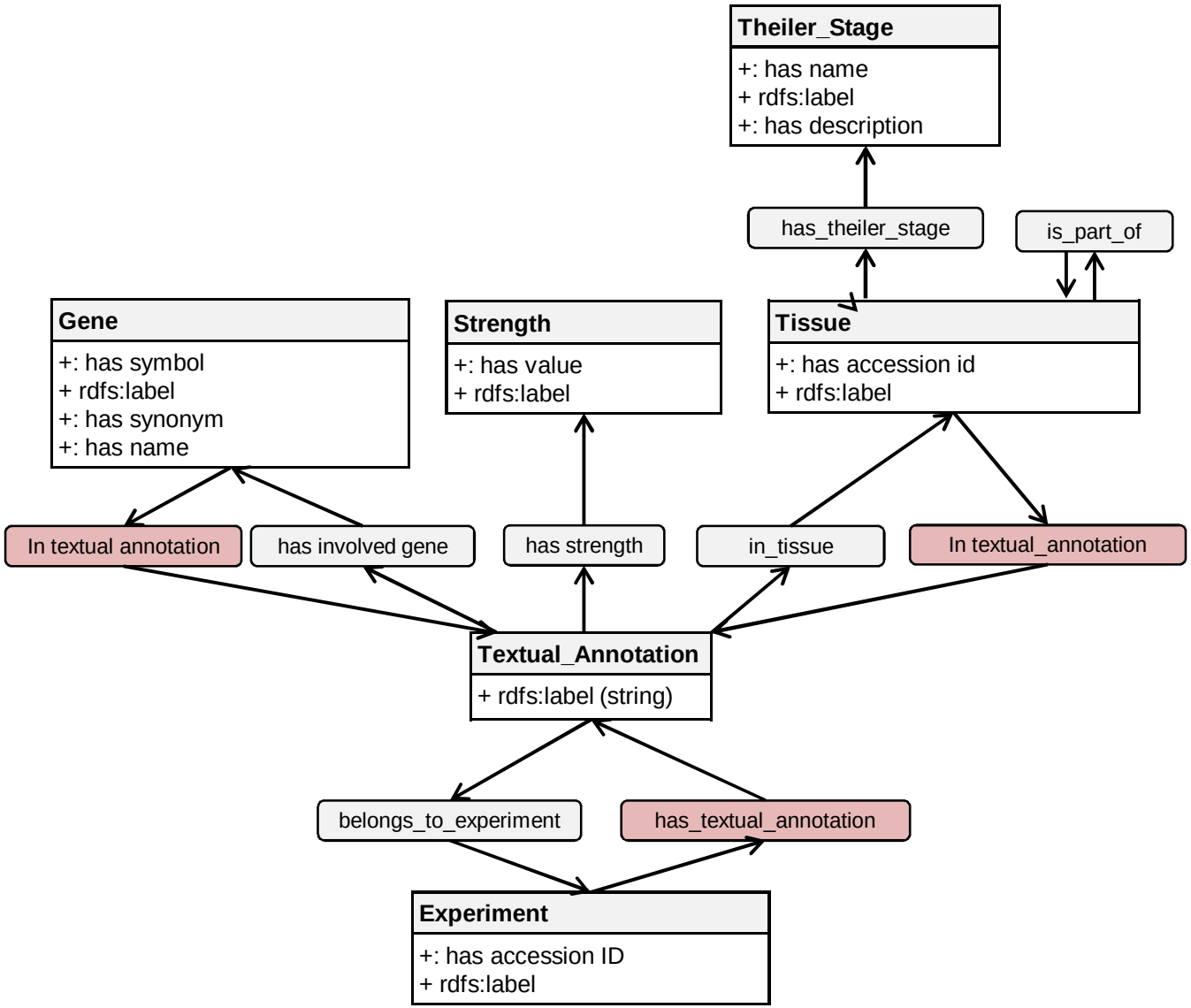


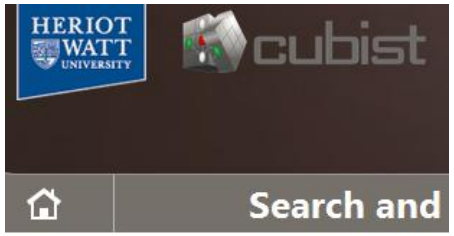
visual analytics

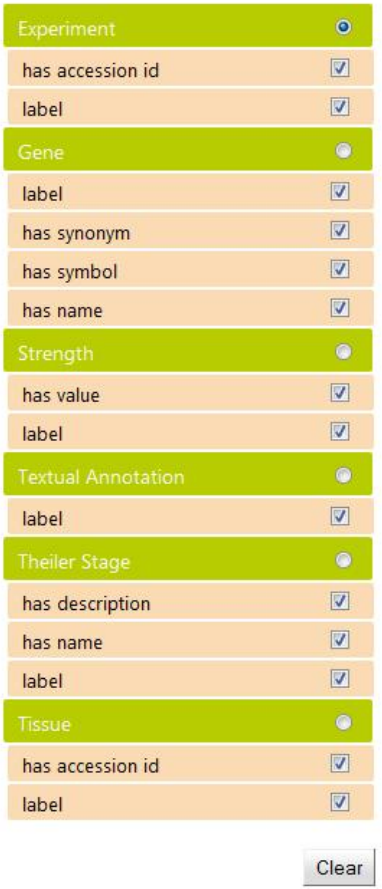
Agenda

- Project Setup and Key Technologies
  - First Introduction into CUBIST
  - Use Cases
  - Introduction into Semantic Technologies
  - Introduction into Formal Concept Analysis
  - Key Messages
- CUBIST Prototype
  - Architecture
  - Different Means to Access Information
  - Semantic Search
  - Query Generation
  - Explorative Search
  - Conceptual Scaling
  - Visual Analytics
- Outcome
  - User Evaluation
  - Our Take
  - Conclusions

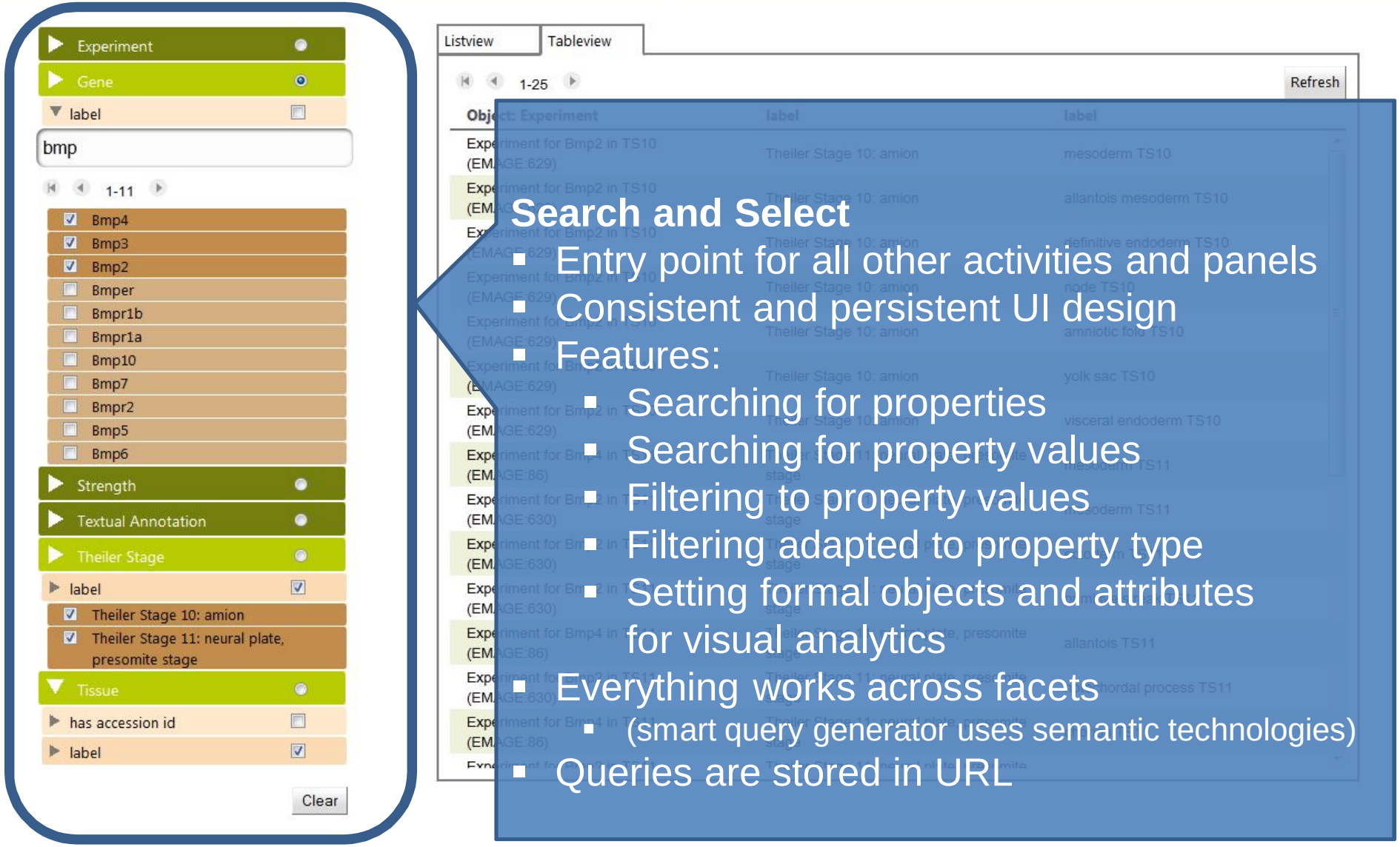
HWU Ontology







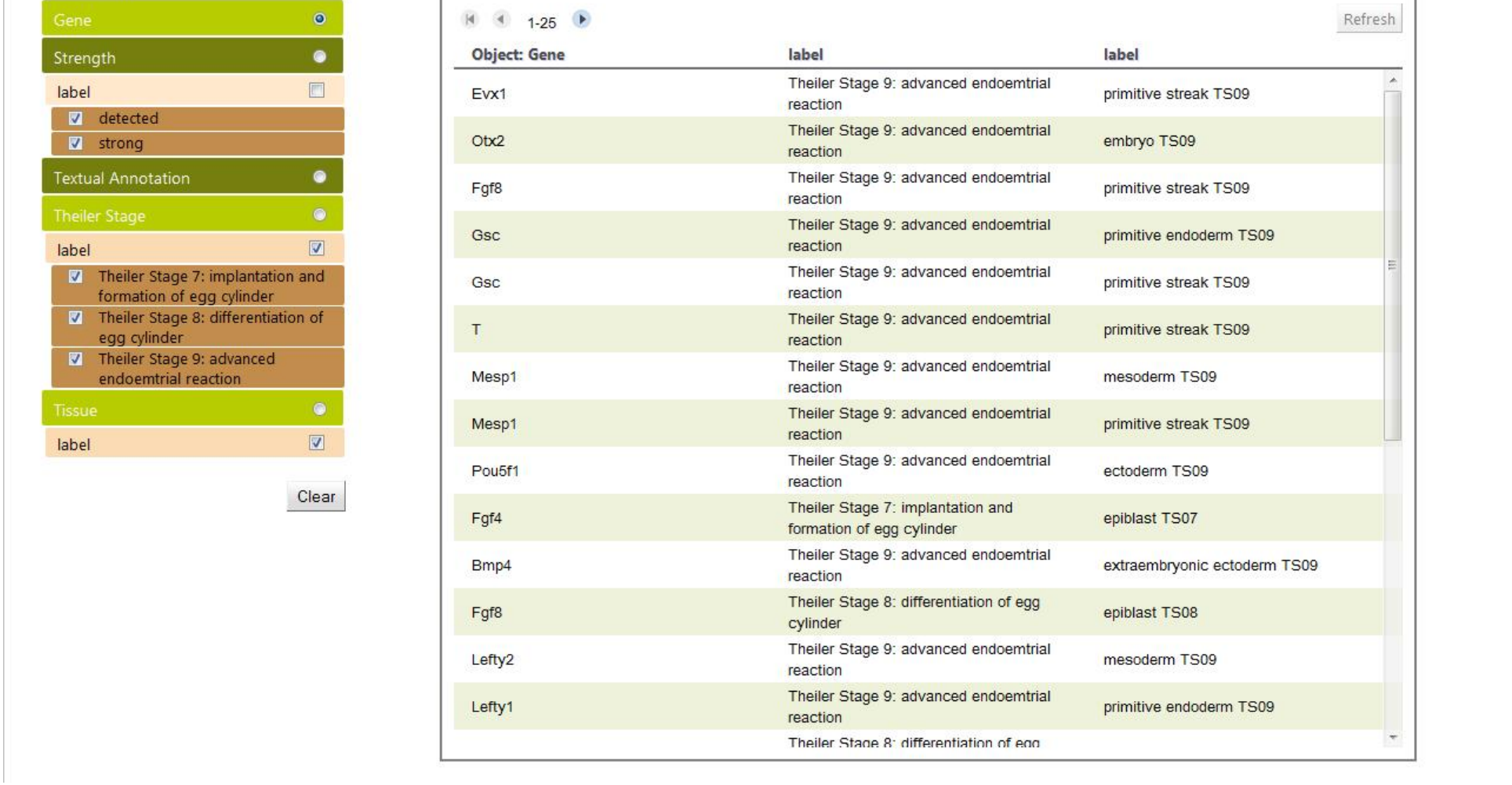
Defining a Data Set Overview



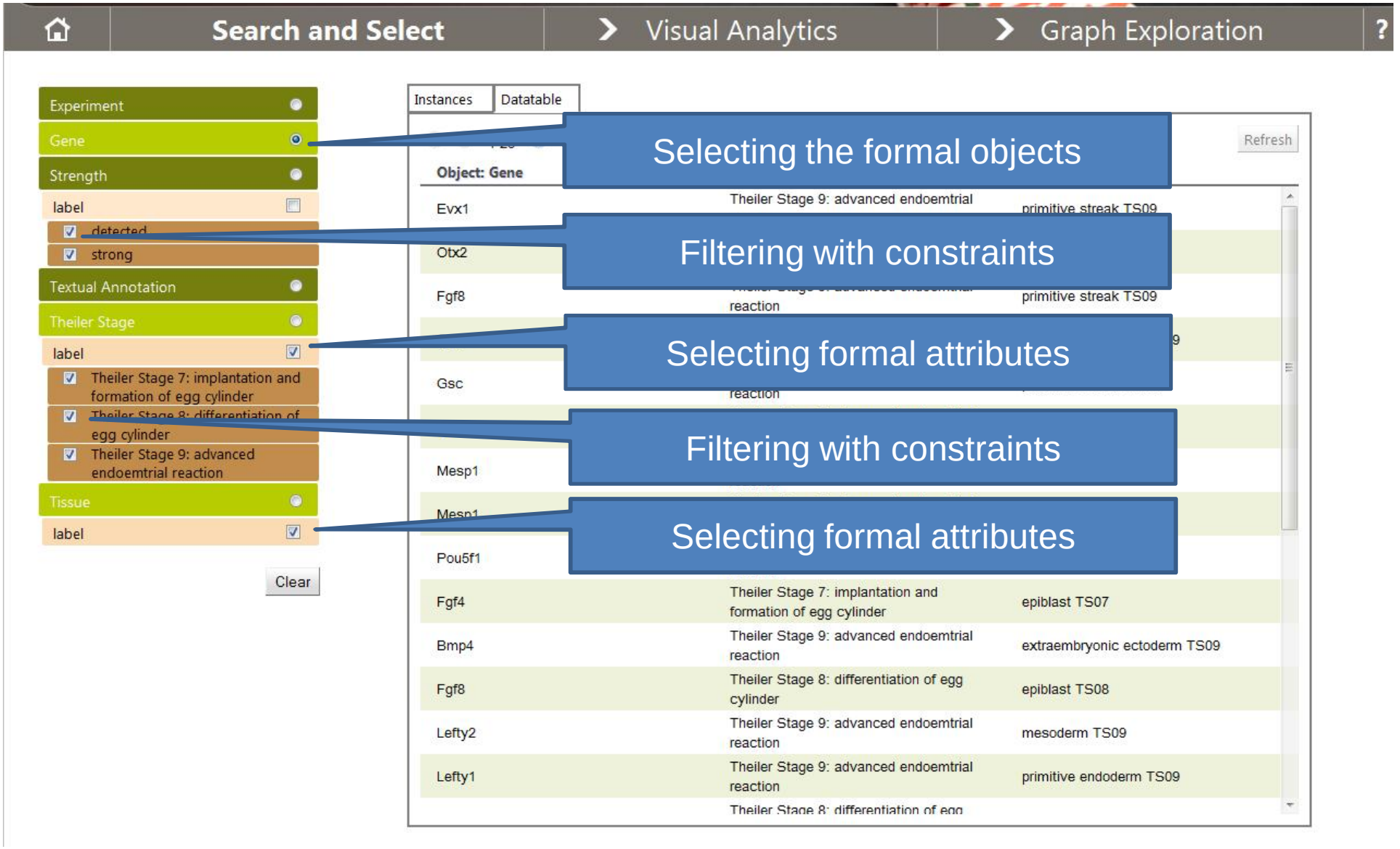
Search and Select

- Entry point for all other activities and panels
- Consistent and persistent UI design
- Features:
  - Searching for properties
  - Searching for property values
  - Filtering to property values
  - Filtering adapted to property type
  - Setting formal objects and attributes for visual analytics
- Everything works across facets
  - (smart query generator uses semantic technologies)
- Queries are stored in URL

Defining a Data Set

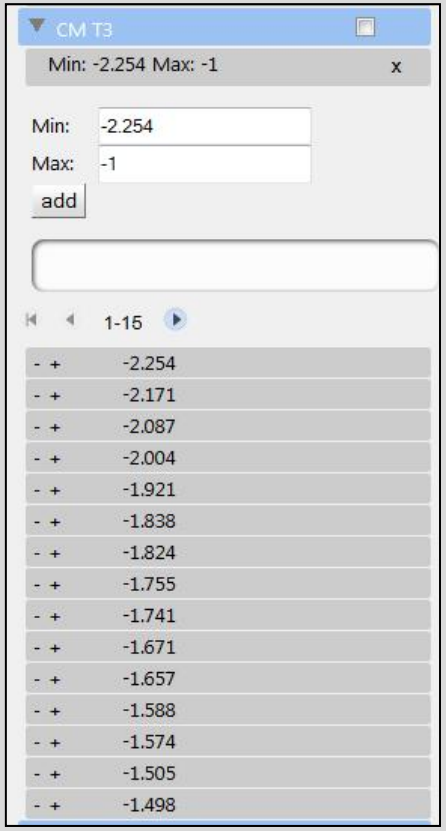


Defining a Data Set

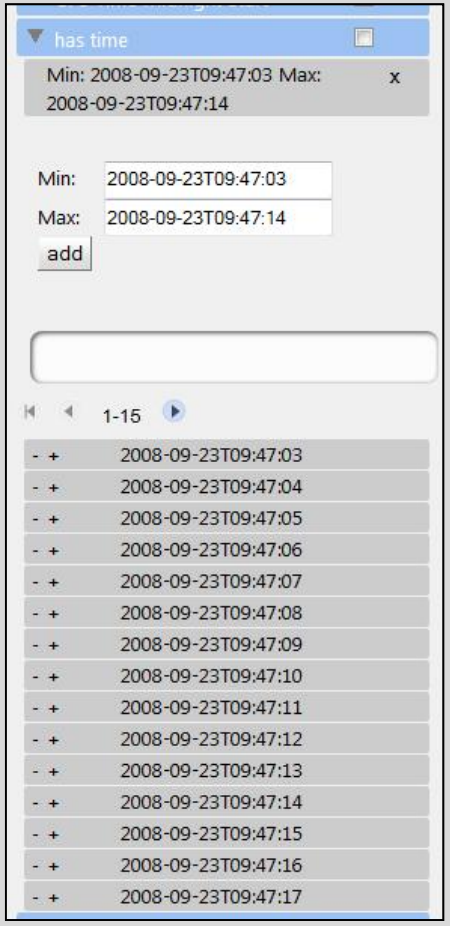


Definining a Dataset
Filtering Dependent on Type

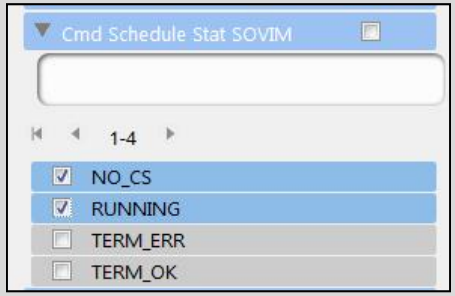
Integer



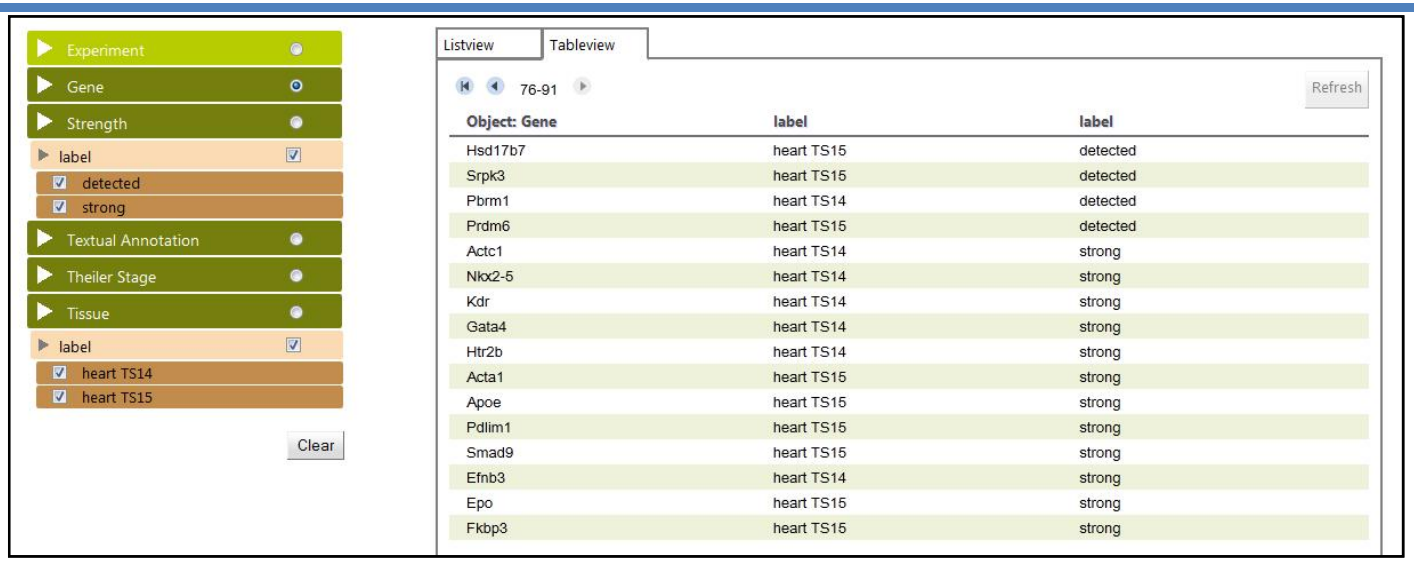
Date/Time

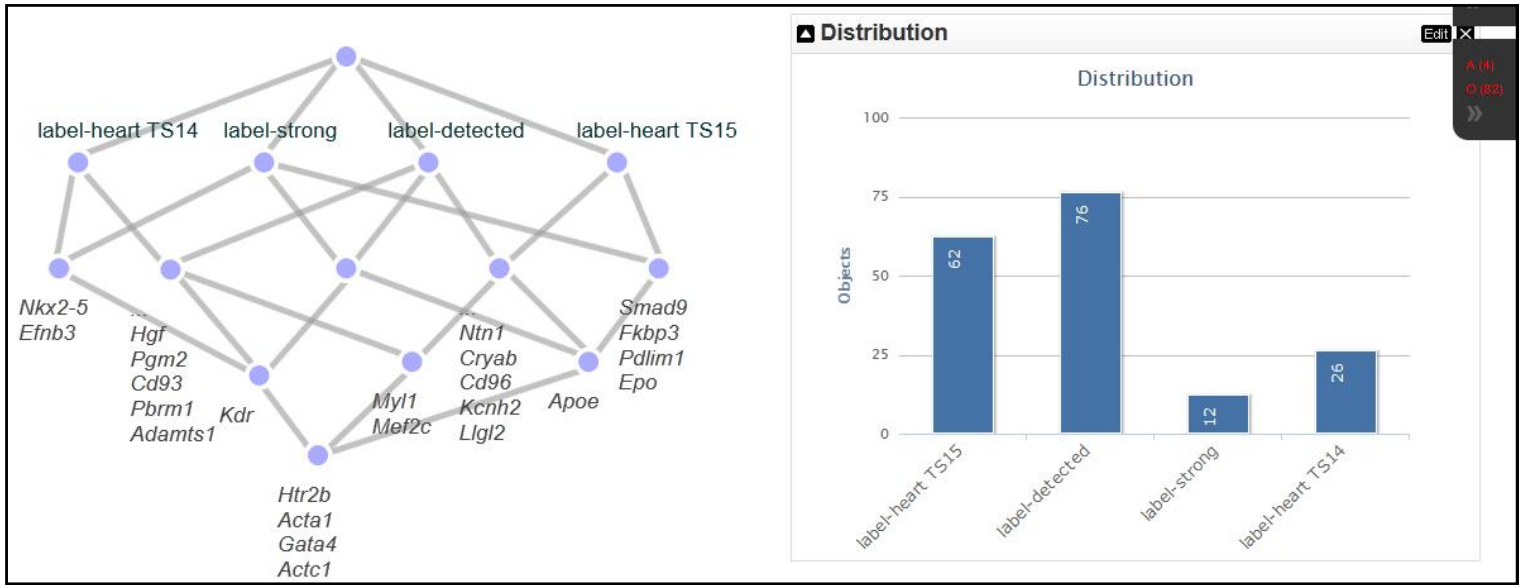


String

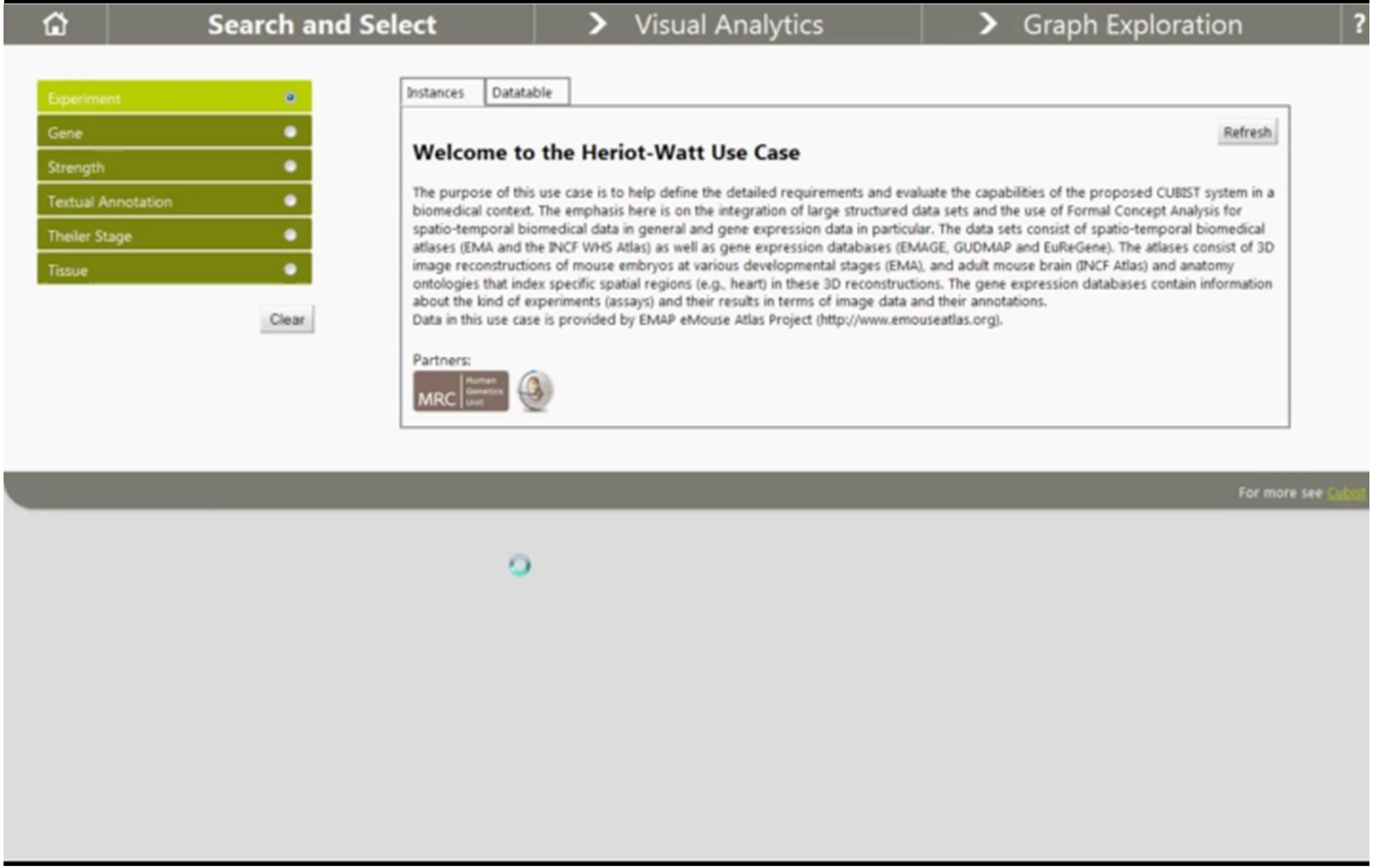


BI as a Self Service





Semantic Search and Instance View Demo



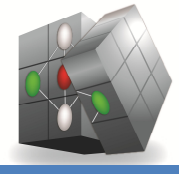
Slide with demo video, removed for th pdf-version of the slides

Content: Semantic Search and Instance View Demo

Watch instead: https://www.youtube.com/watch?v=Kuu756nr1_I

Agenda

- Project Setup and Key Technologies
  - First Introduction into CUBIST
  - Use Cases
  - Introduction into Semantic Technologies
  - Introduction into Formal Concept Analysis
  - Key Messages
- CUBIST Prototype
  - Architecture
  - Different Means to Access Information
  - Semantic Search
  - Query Generation
  - Explorative Search
  - Conceptual Scaling
  - Visual Analytics
- Outcome
  - User Evaluation
  - Our Take
  - Conclusions



Faceted/Semantic Search

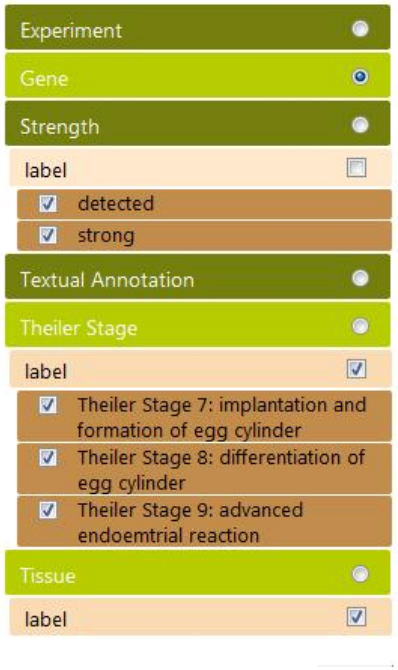
Ontological elements in UI

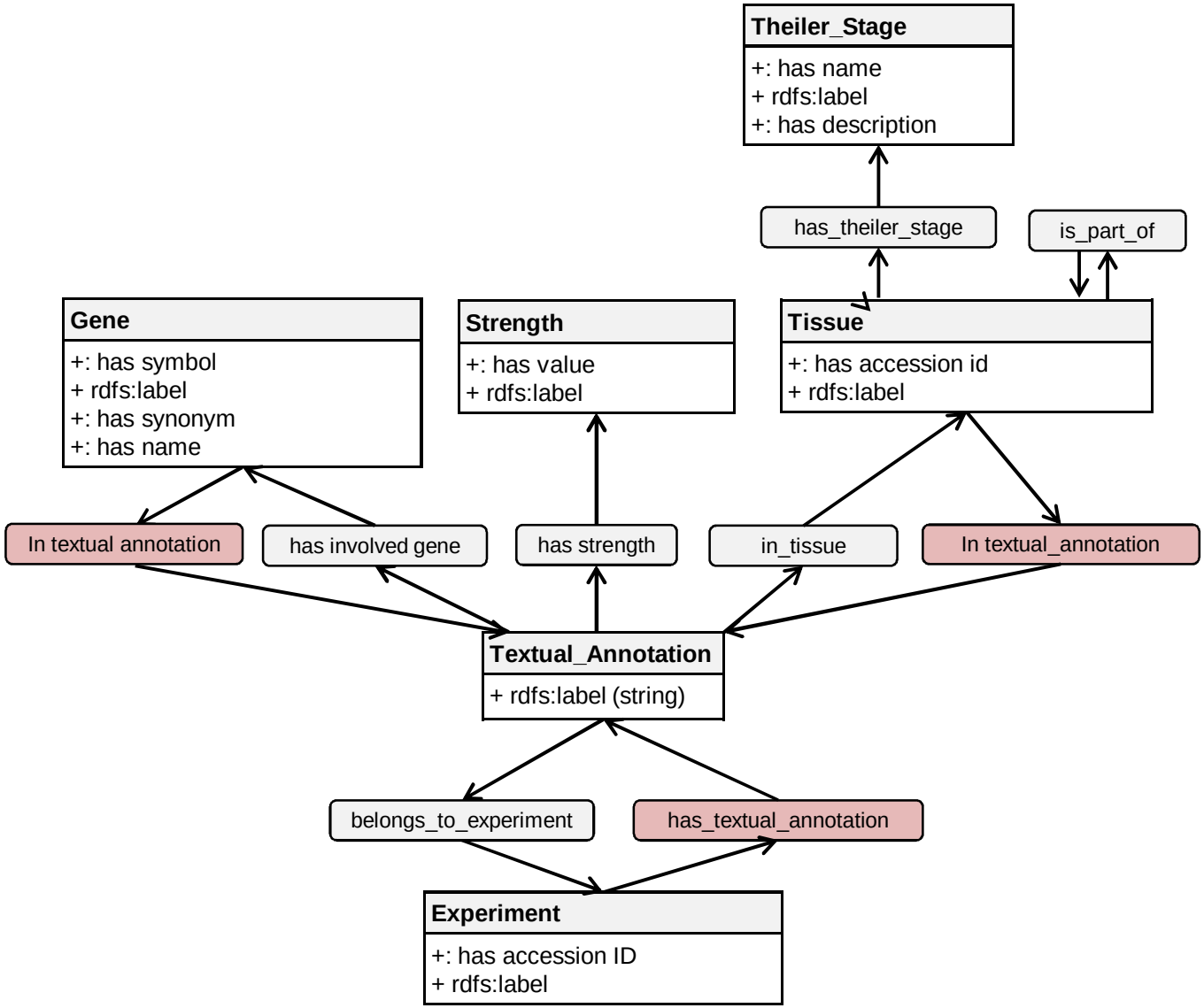
- Types are in UI displayed as facets
- Datatype properties are displayed as attributes
- Object properties are hidden

Ontological elements for query generation

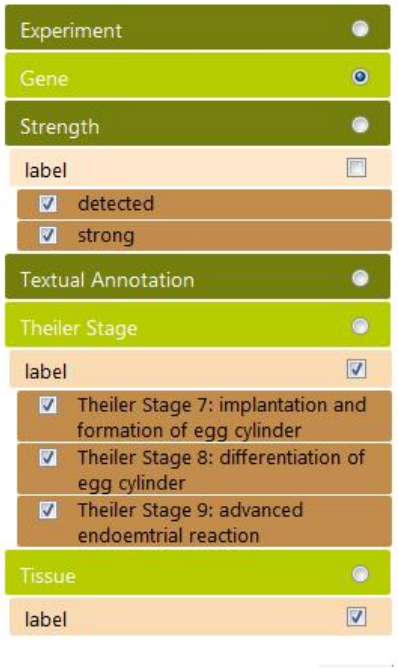
- Smart query generation taking ontology into account
- Types and object properties form the “query graph”
- Query graph can contain more types than selected in UI
- Datatype properties are used for filtering and formal attributes

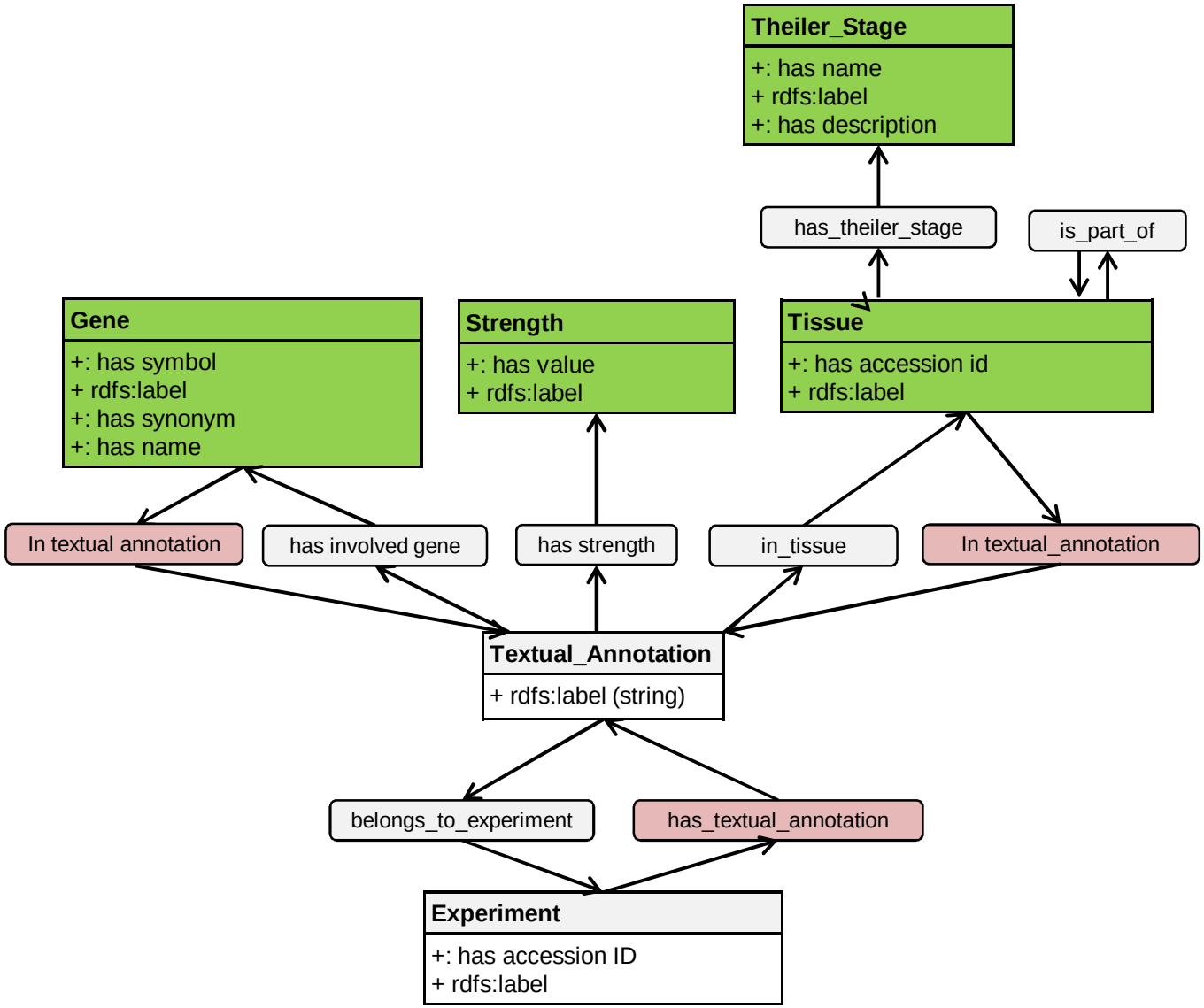
Defining a Date Set: Generating Query
Step1: Find minimal connected subgraph



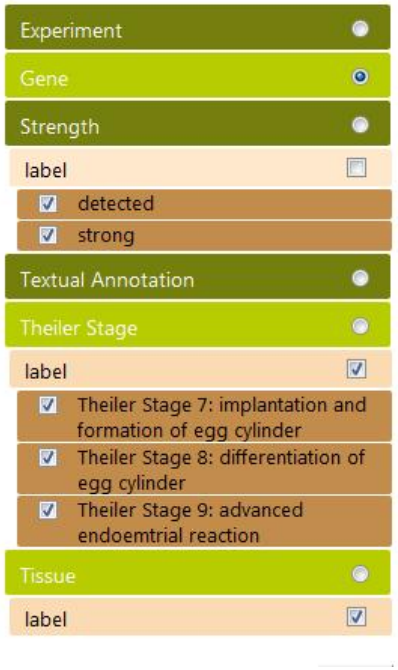


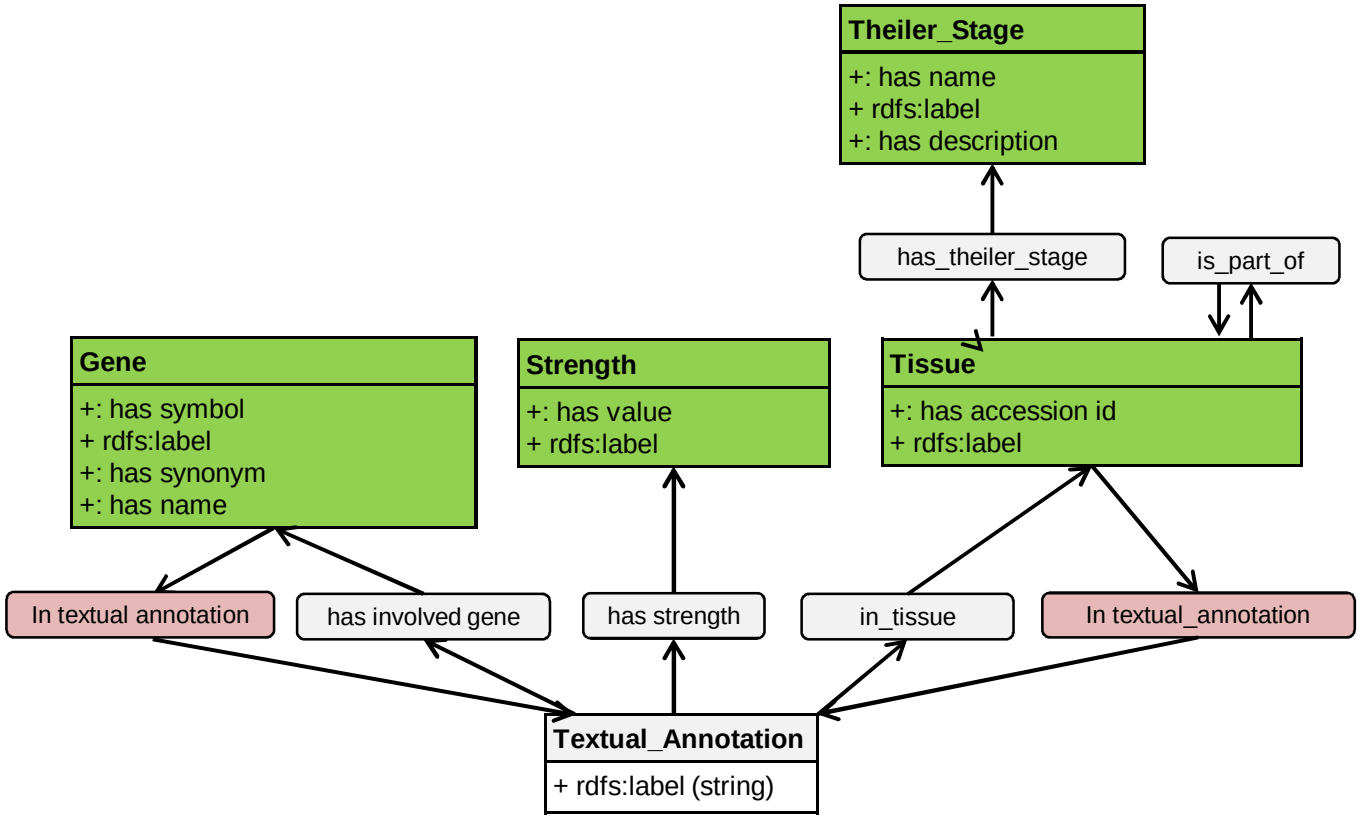
Defining a Date Set: Generating Query
Step1: Find minimal connected subgraph





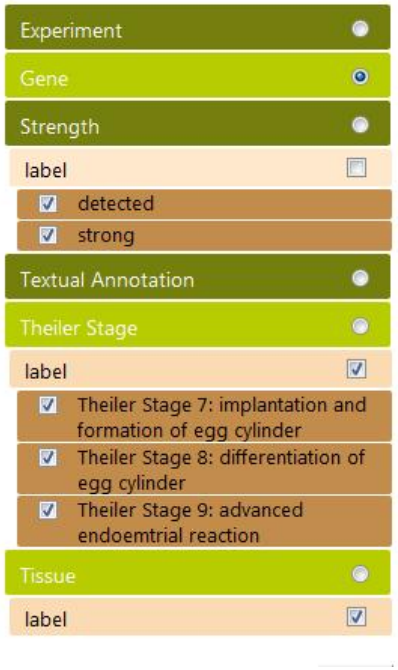
Defining a Date Set: Generating Query
Step1: Find minimal connected subgraph

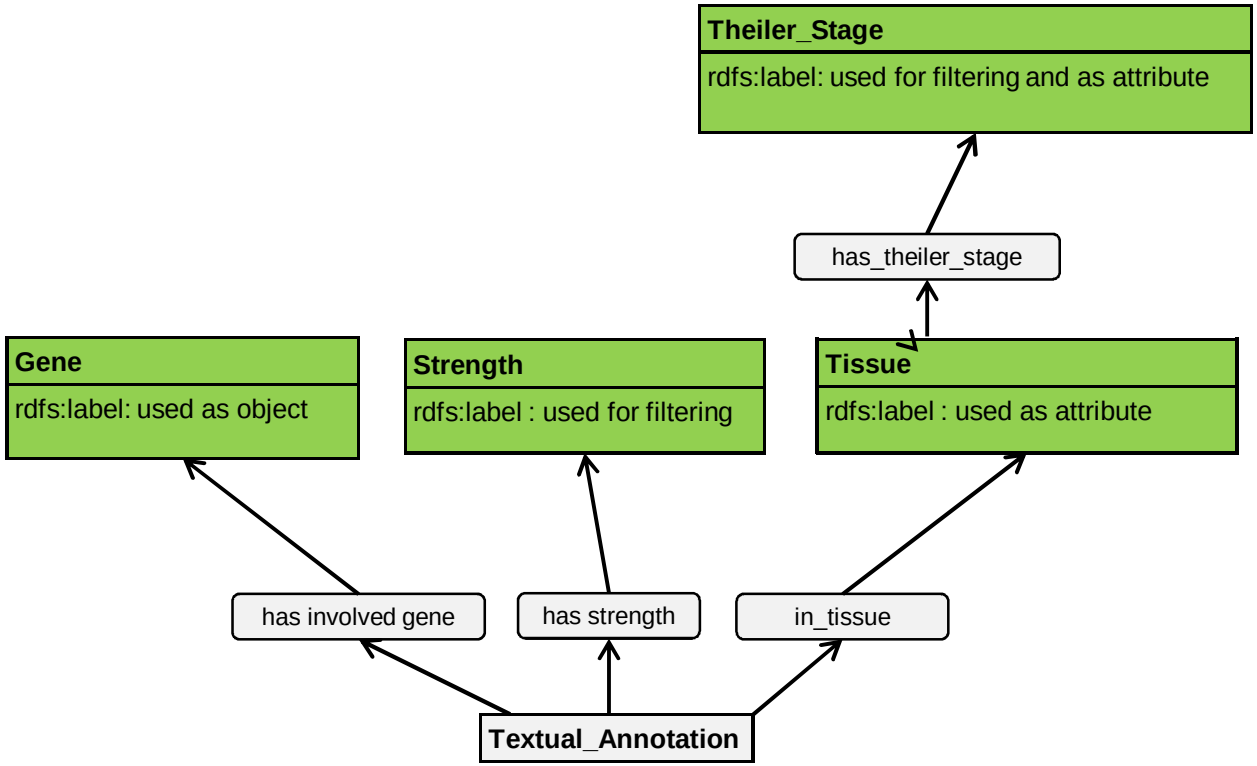




Defining a Date Set: Generating Query

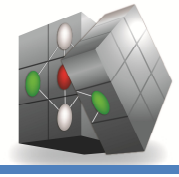
Step2: Use attributes as query variables or for filtering





Agenda

- Project Setup and Key Technologies
  - First Introduction into CUBIST
  - Use Cases
  - Introduction into Semantic Technologies
  - Introduction into Formal Concept Analysis
  - Key Messages
- CUBIST Prototype
  - Architecture
  - Different Means to Access Information
  - Semantic Search
  - Query Generation
  - Explorative Search
  - Conceptual Scaling
  - Visual Analytics
- Outcome
  - User Evaluation
  - Our Take
  - Conclusions



Graph Exploration View

- Used for exploring the information space
- Enties -> nodes, semantic relationship between entities -> edges
- highly interactive

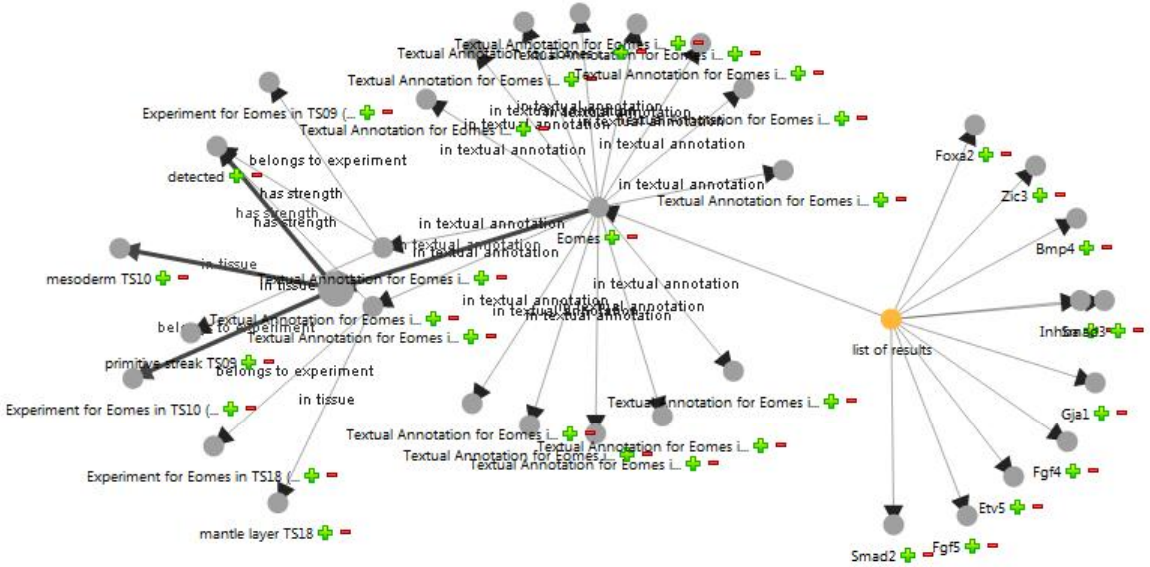
Graph Exploration View Screenshot

Search and Select | Graph Exploration | ?

expand relations

has textual annotation | has theiler stage | in tissue | belongs to experiment | has involved gene | has strength | in textual annotation

Reset layout | Zoom - | Zoom +



Current Node: **Textual Annotation for Eomes in embryo/mesoderm TS10**

Select range for this node:

1 2 3 ∞

Properties:

http://www.w3.org/2000/01/rdf-schema#label: Textual Annotation for Eomes in embryo/mesoderm TS10

Relations:

- has strength -> detected
- has involved gene -> Eomes
- in tissue -> mesoderm TS10
- belongs to experiment -> Experiment for Eomes in TS10 (EMAGE:3793)

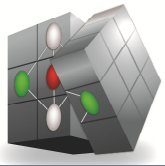
For more see Cubist

Slide with demo video, removed for th pdf-version of the slides

Content: Graph Exploration Demo

Watch instead: https://www.youtube.com/watch?v=Kuu756nr1_I

Functionalities within the Graph Exploration View



Extending the Graph Visualization:

- single relation for a single node
- all relations for a single node
- all relations of one type for all nodes

Restricting the Graph Visualization:

- removing adjecent nodes for a given node
- removing a single node
- only showing nodes within a given range for given node

User Interactions with the Graph Visualization

Manipulating the Graph Visualization:

- zoom in / zoom out
- automatically refreshing layout
- moving complete graph
- moving single node

Searching the Graph Visualization:

- highlighting adjacent nodes for a given node

Agenda

- Project Setup and Key Technologies
  - First Introduction into CUBIST
  - Use Cases
  - Introduction into Semantic Technologies
  - Introduction into Formal Concept Analysis
  - Key Messages
- CUBIST Prototype
  - Architecture
  - Different Means to Access Information
  - Semantic Search
  - Query Generation
  - Explorative Search
  - Conceptual Scaling
  - Visual Analytics
- Outcome
  - User Evaluation
  - Our Take
  - Conclusions

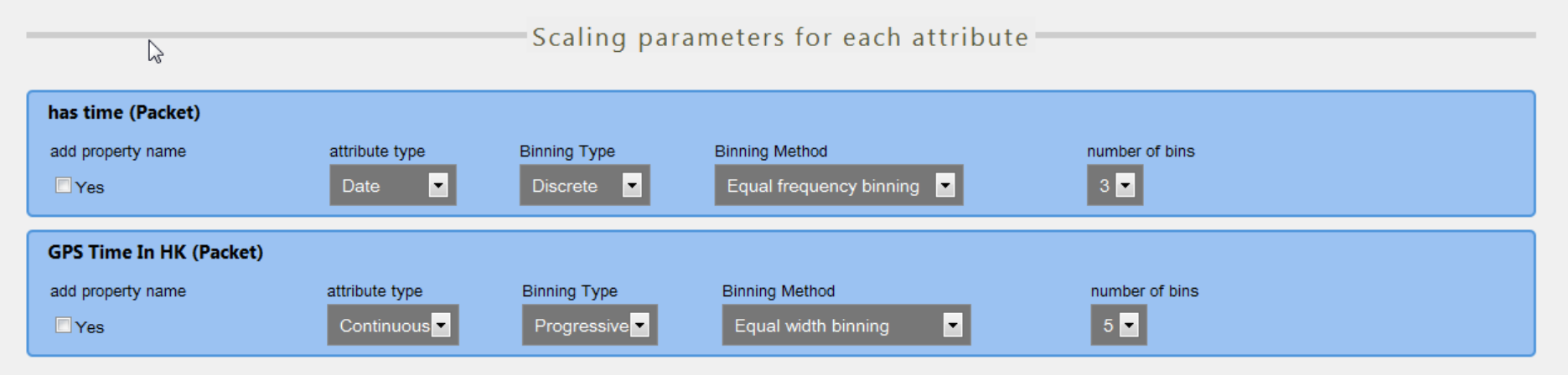
Conceptual Scaling in CUBIST

- Scaling in CUBIST essentially works on linearly ordered datatypes (date-time, int, …)
- Essentially, the set of all values is divided into intervals
- E.g. intervals of equal length, intervals with same number of (materialized) values, standard deviation …

Conceptual Scaling in CUBIST
Called “Binning” in CUBIST

Conceptual Scaling Options

- **Attribute Types**
  - Categorical (aka “no scaling”)
  - Boolean
  - Continuous (discretising the data)
  - Date (using standard ranges like month, week)
  - Ordinal (like categorical, where order is important)
- **Binning Type**
  - Discrete
  - Progressive
- **Binning Method**
  - Equal frequency binning
  - Equal width binning
  - Standard deviation binning
  - Manual binning
- **Number of Bins**



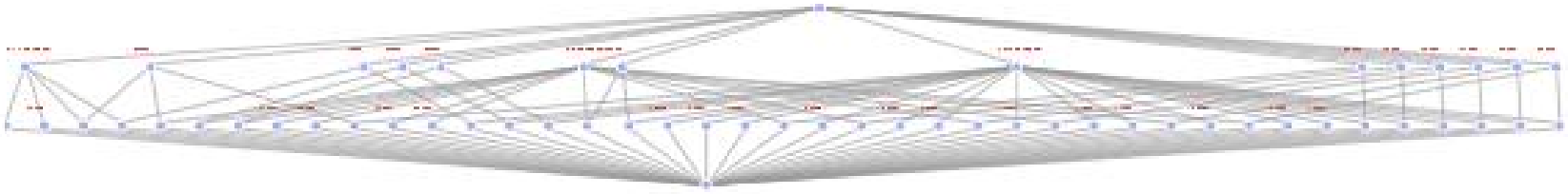
Innovantage Example
Without Binning / Conceptual Scaling

- Advertiser
- Contact
- Discipline
- Jobboard
- label
  - (Jobboard) Jobsite
- Location
- Salary
- min salary
  - Min: 20000 Max: 100000 x
- Subdiscipline
- Vacancy
- job title
  - C# Developer
  - C++ Developer
  - Java Developer

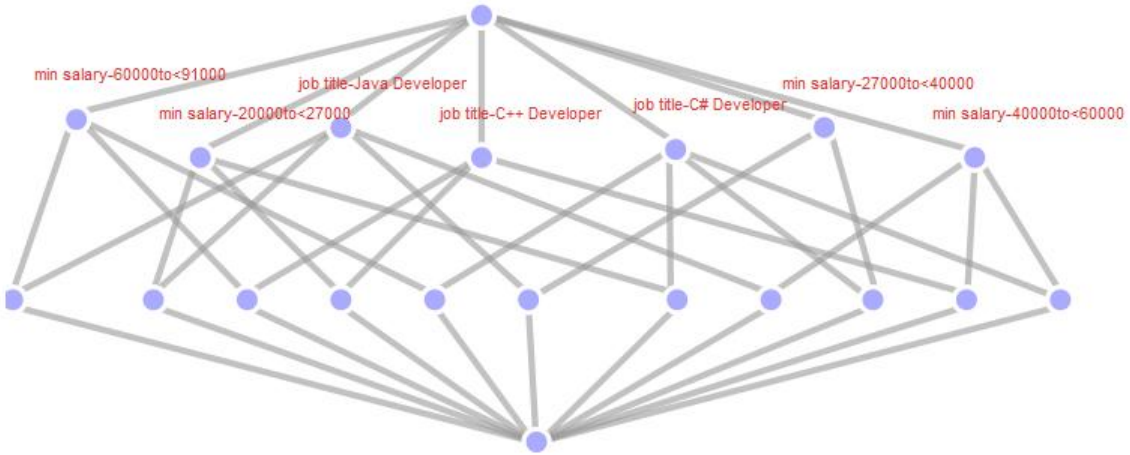
Instances | Datatable

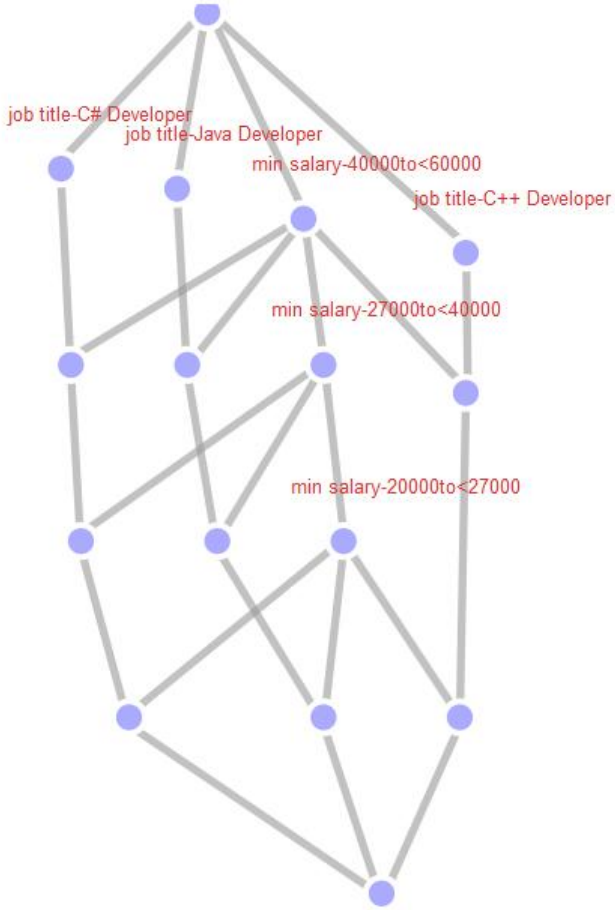
1-25 Refresh

Object: Vacancy	job title	min salary
(Vacancy: 62384928) C++ Developer	C++ Developer	20000
(Vacancy: 62161385) Java Developer	Java Developer	21000
(Vacancy: 62384979) C# Developer	C# Developer	22000
(Vacancy: 62723704) C# Developer	C# Developer	23000
(Vacancy: 62216155) C# Developer	C# Developer	23000
(Vacancy: 62255107) C# Developer	C# Developer	23000
(Vacancy: 62514244) C# Developer	C# Developer	23000
(Vacancy: 62515608) C# Developer	C# Developer	23000
(Vacancy: 62215202) Java Developer	Java Developer	24000
(Vacancy: 62432922) Java Developer	Java Developer	24000
(Vacancy: 62104022) Java Developer	Java Developer	25000
(Vacancy: 62160520) Java Developer	Java Developer	25000
(Vacancy: 62160866) Java Developer	Java Developer	25000
(Vacancy: 62215770) Java Developer	Java Developer	25000
(Vacancy: 62255790) Java Developer	Java Developer	25000
(Vacancy: 62297939) Java Developer	Java Developer	25000

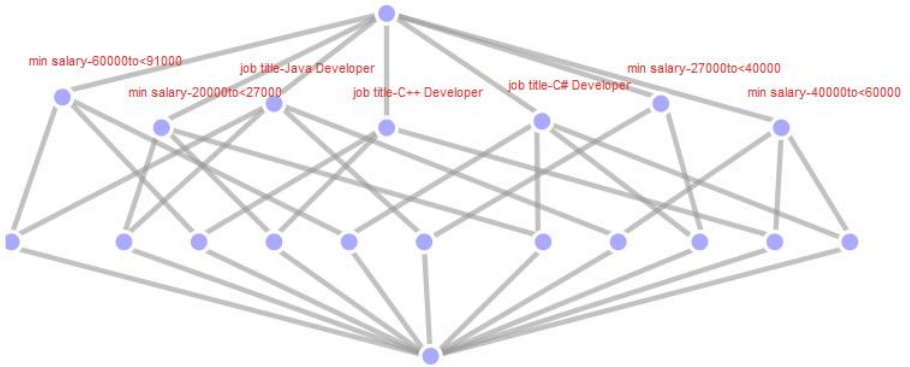


Binning Type:
Discrete vs. Progressive

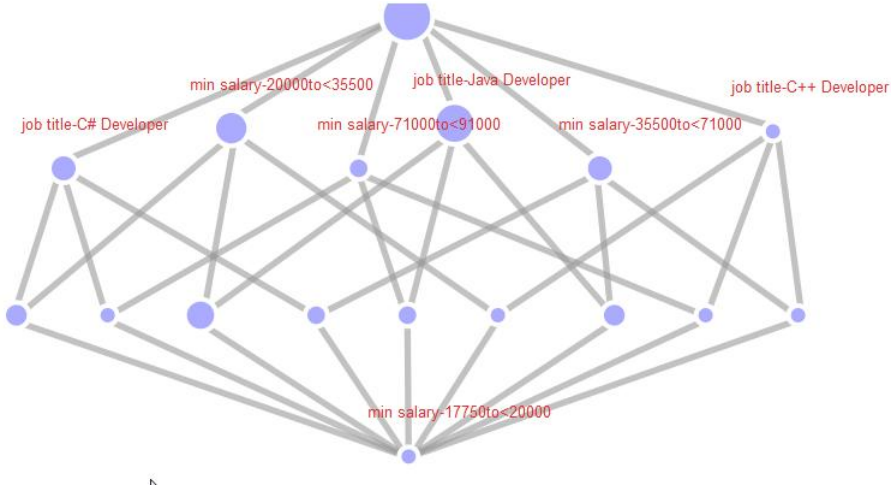




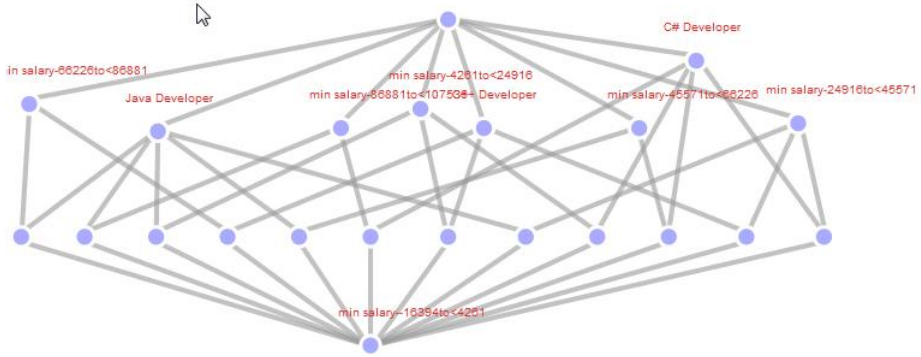
Binning methods



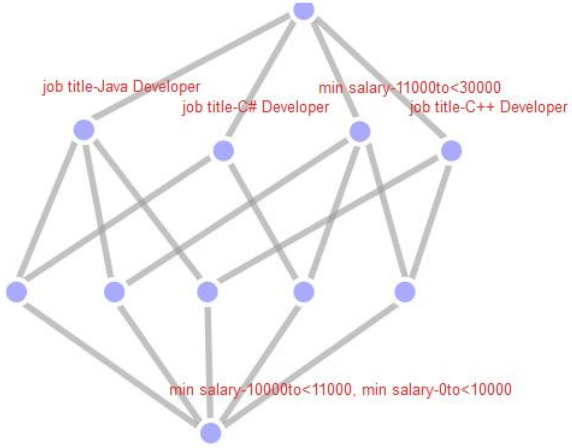
Equal frequency binning



Equal width binning

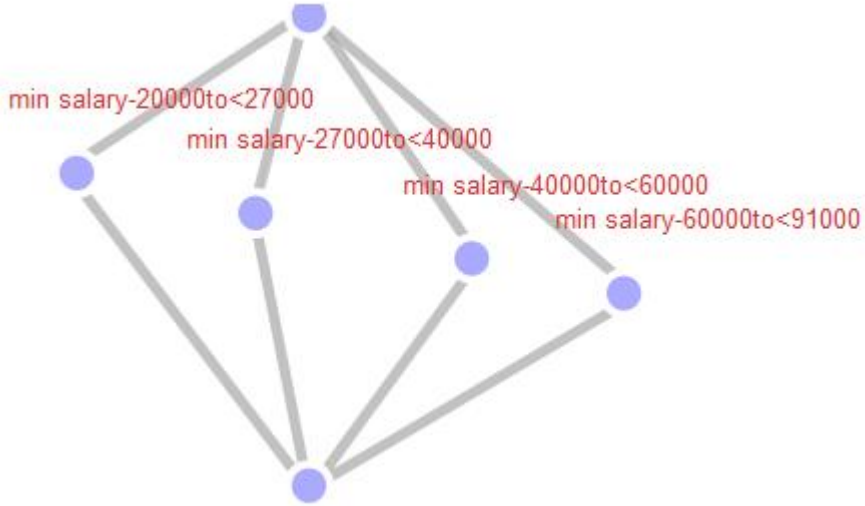


Standard deviation binning

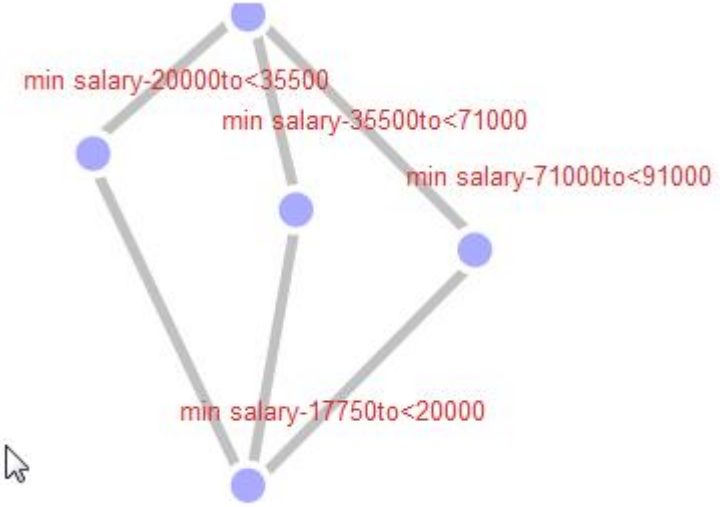


Manual binning

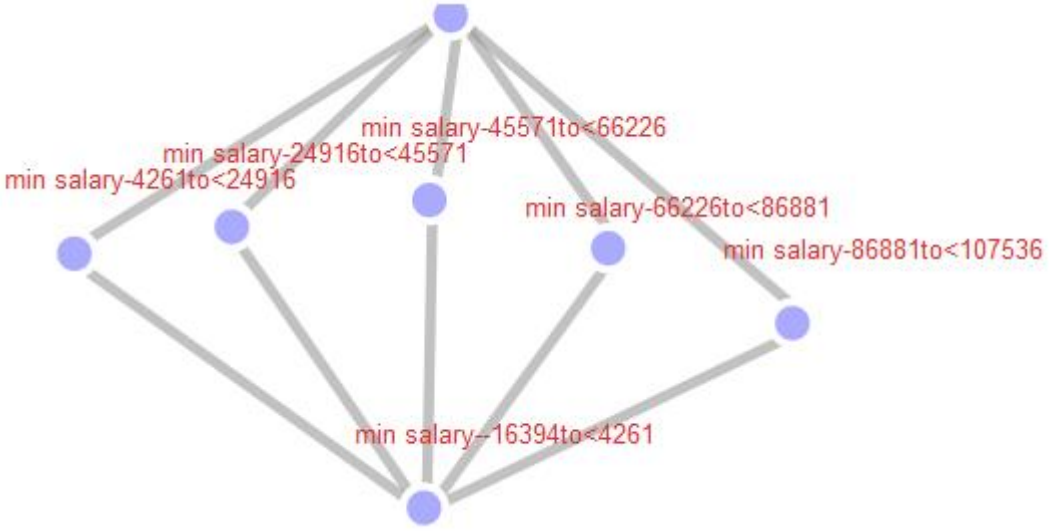
Binning methods



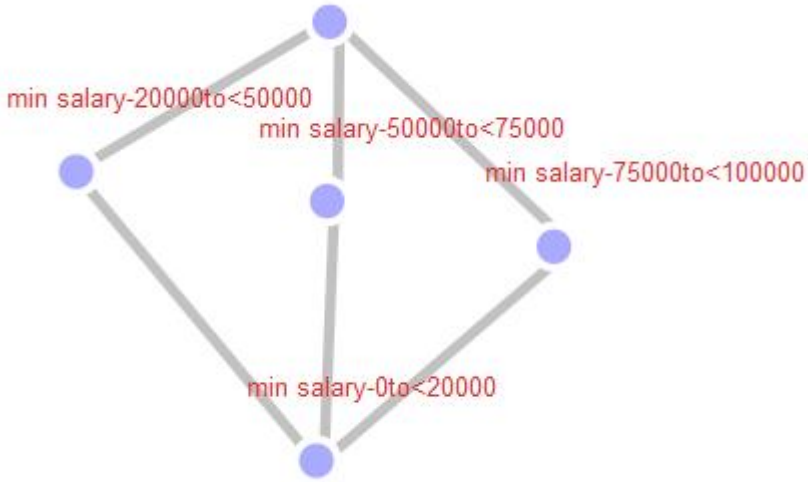
Equal frequency binning



Equal width binning



Standard deviation binning



Manual binning

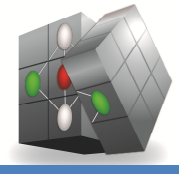
Agenda

- Project Setup and Key Technologies
  - First Introduction into CUBIST
  - Use Cases
  - Introduction into Semantic Technologies
  - Introduction into Formal Concept Analysis
  - Key Messages
- CUBIST Prototype
  - Architecture
  - Different Means to Access Information
  - Semantic Search
  - Query Generation
  - Explorative Search
  - Conceptual Scaling
  - Visual Analytics
- Outcome
  - User Evaluation
  - Our Take
  - Conclusions

Visual Analytics

- Visual analytics focuses on massive and dynamic volumes of information
- Supports human judgment
- by means of visual representations and interaction techniques in the analysis process [Keim *et al. 2001*]

- Visual Analytics in CUBIST combines:
  - Traditional BI (charts)
  - Graph-based visualization (graphs)
  - Concept visualization (concept lattice)



Visual Analytics

Summary

- Visual Analytics for lattices and rules
- Comprehensive set of visualizations
- Comprehensive formatting
- Filtering
- Combination of FCA, graphs, and traditional BI
- Highly interactive
- Linking and Brushing

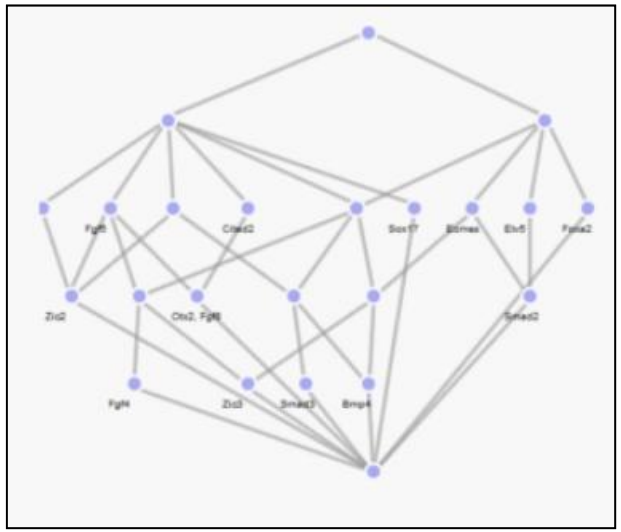
Lattices

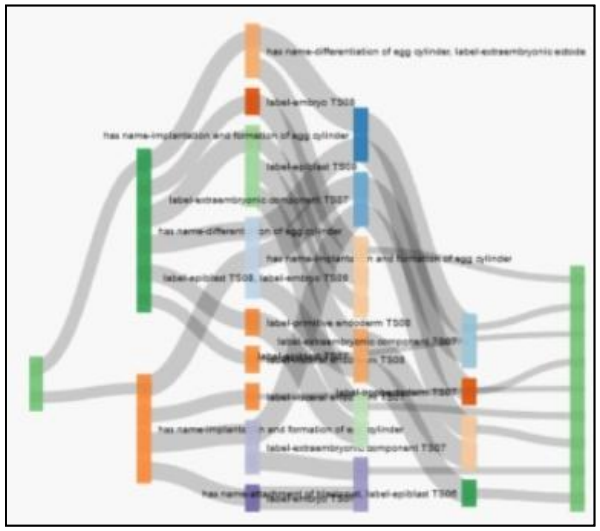
- Several metrics for attributes (color, size) of nodes and edges
- Filtering
- Additional Graphs
  - Distribution
  - Co-Occurrence
  - Concept comparison
  - Attribute graph
- Several Visualizations
  - Hasse-Diagram
  - Sankey
  - Sunburst
  - Tree
  - ICicle

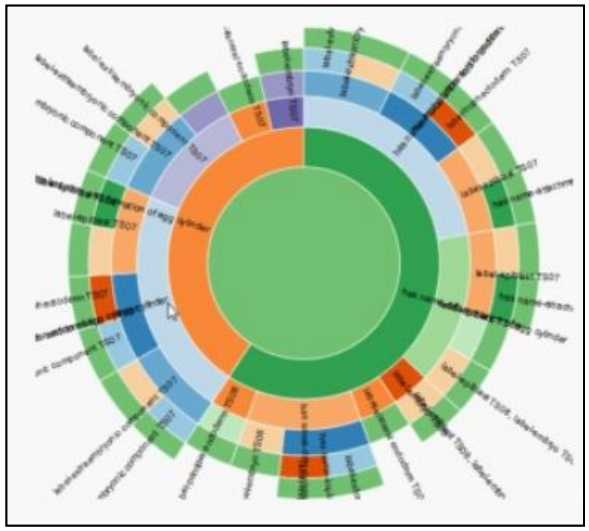
Rules

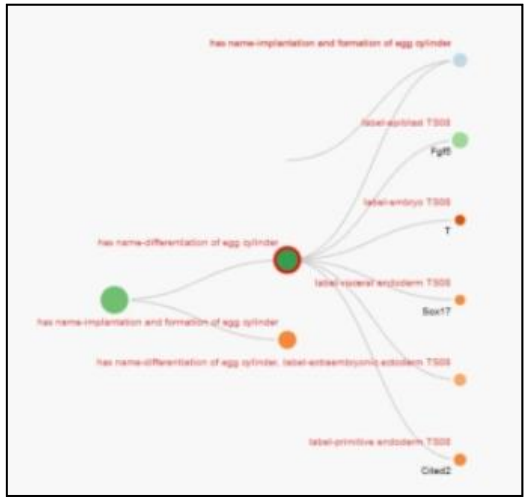
- Two Visualizations
  - Matrix
  - Radial
- Filtering with different metrics
  - Selection with scatter-plot

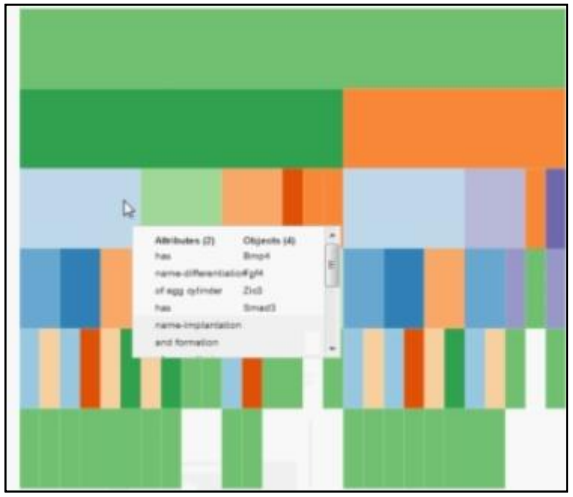
CUBIST Visualizations





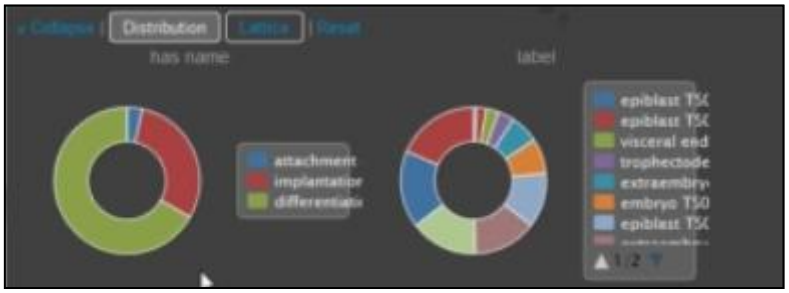


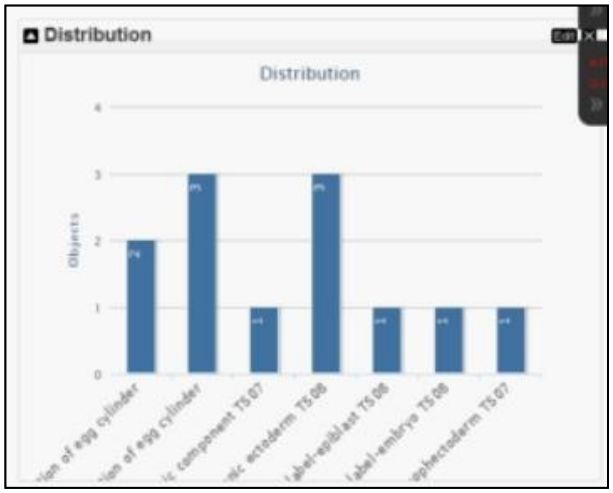


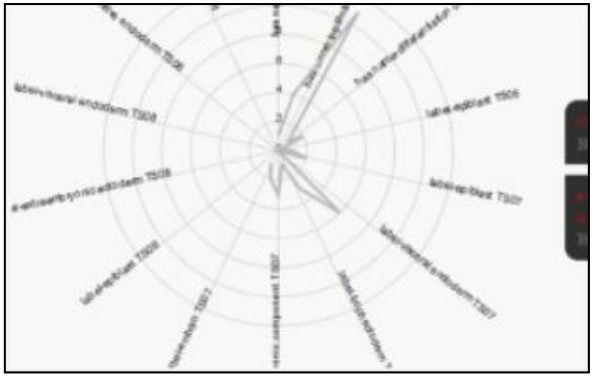


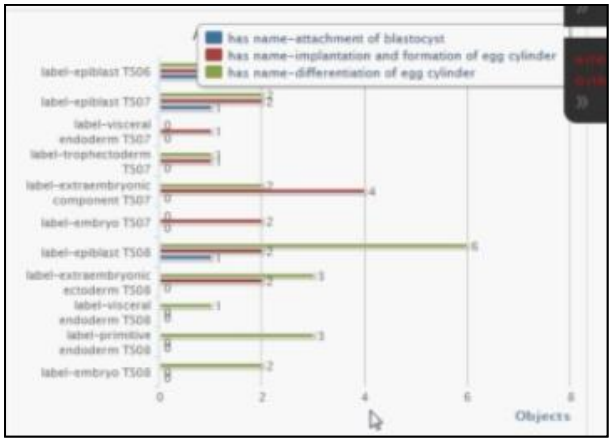
- Several Visualizations
  - Hasse-Diagram
  - Sankey
  - Sunburst
  - Tree
  - ICicle

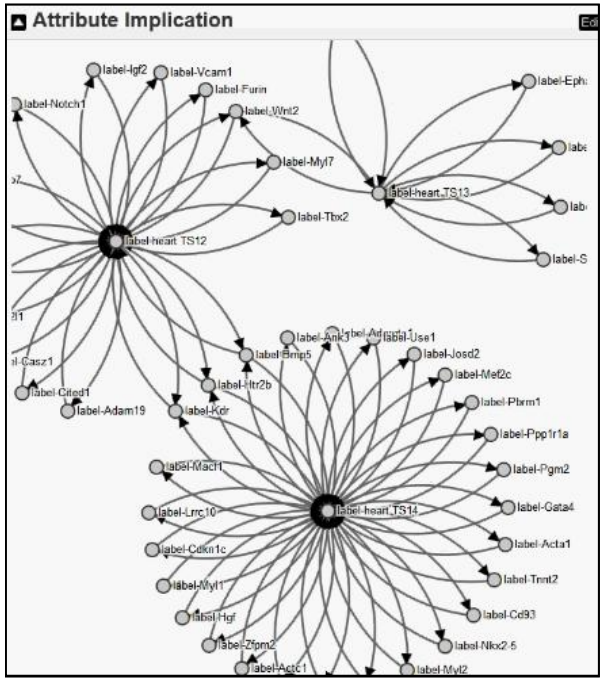
CUBIST functionalities





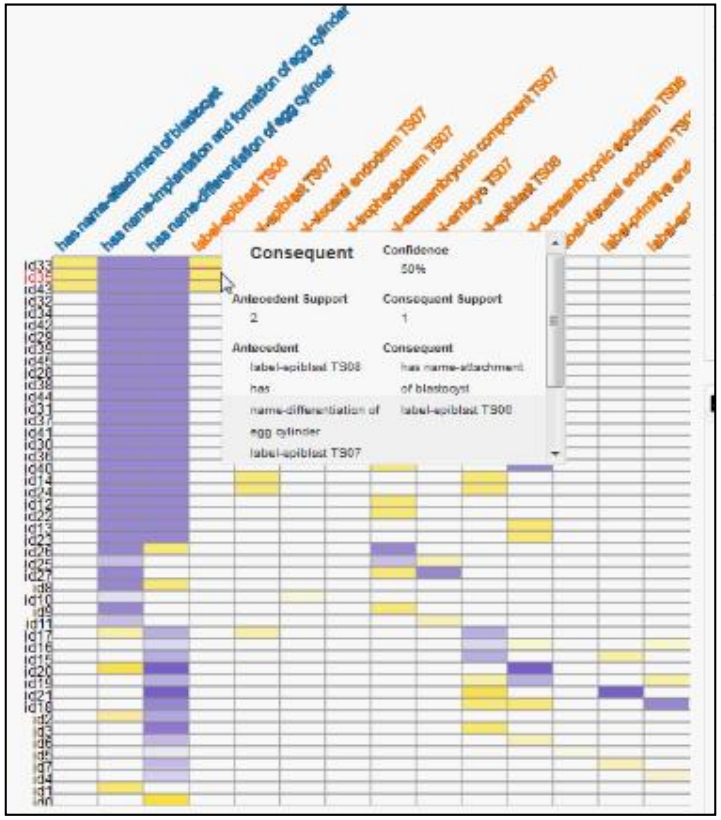


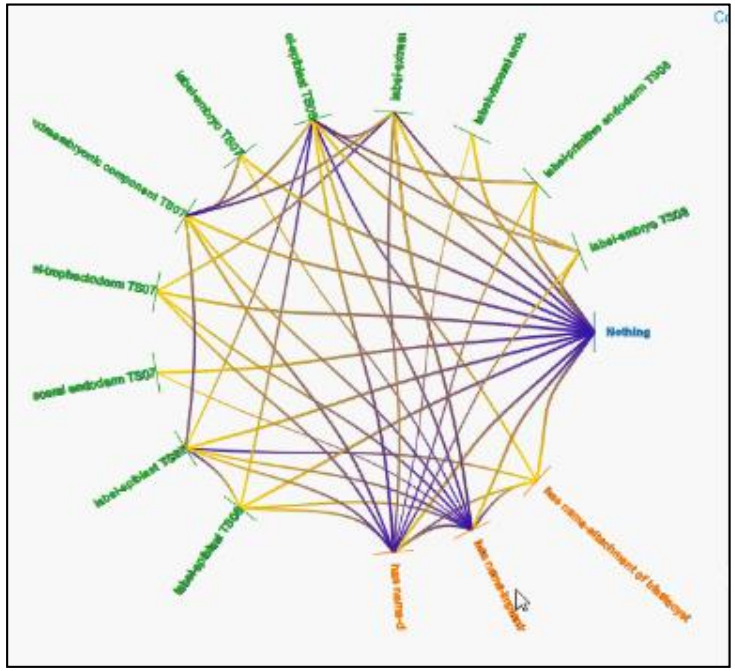


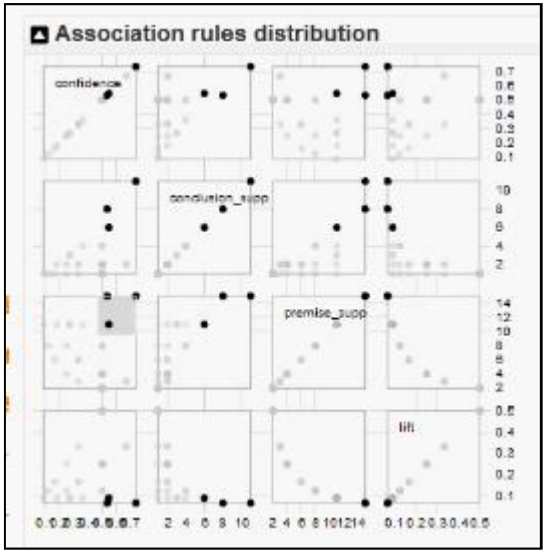


- Filtering
- Additional Graphs
  - Distribution
  - Co-Occurrence
  - Concept comparison
  - Attribute graph

CUBIST functionalities







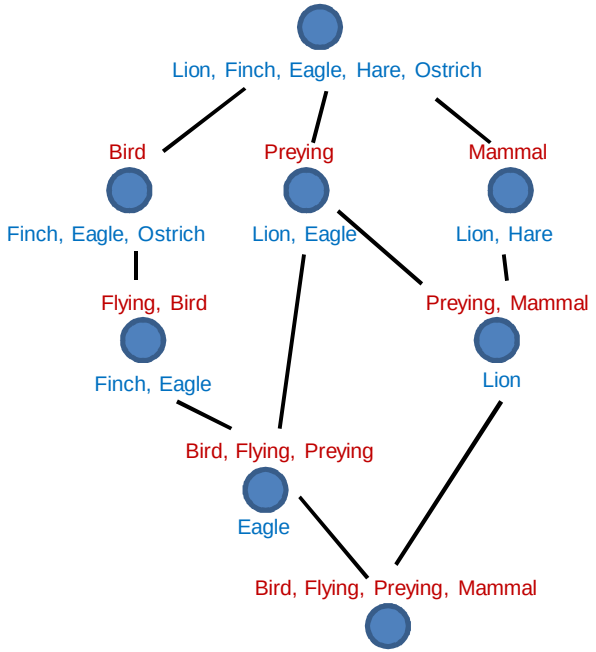
Rules

- Two Visualizations
  - Matrix
  - Radial
- Filtering with different metrics
  - Selection with scatter-plot

Association Rules

Displays patterns of co-occurrence between data under the form:
Premise => Conclusion

Concept lattice



Association rules

Conf.	#	Attributes		Attributes	#
100%	2	Flying	=>	Bird	3
50%	2	Preying	=>	Flying, Bird	1
50%	2	Preying	=>	Mammal	1

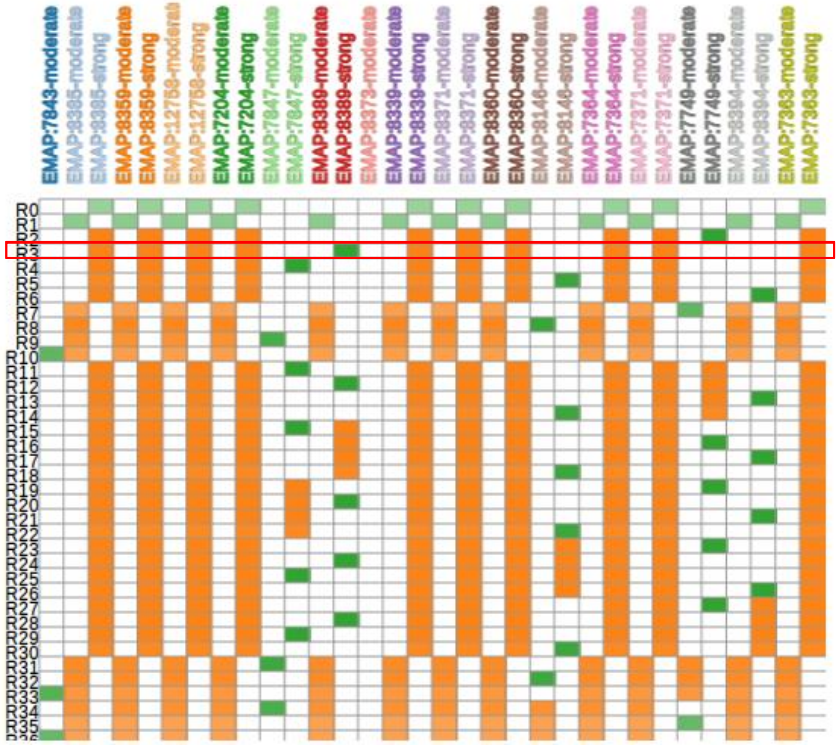
Visualization of Association Rules

New visual metaphors for association rules

```
5 < 1 > age-30to<40 education-HS-grad sex-Male US-citizen =[100%]=> < 1 > employment-Unskilled;
6 < 1 > age-30to<40 employment-Clerical =[100%]=> < 1 > education-Bachelors sex-Male US-citizen;
7 < 1 > age-30to<40 employment-Managerial =[100%]=> < 1 > education-Masters sex-Female;
8 < 1 > age-30to<40 employment-Unskilled sex-Male US-citizen =[100%]=> < 1 > education-HS-grad;
9 < 1 > age-30to<40 sex-Female =[100%]=> < 1 > education-Masters employment-Managerial;
10 < 2 > age-30to<40 US-citizen =[100%]=> < 2 > sex-Male;
11 < 1 > age-40to<50 =[100%]=> < 1 > employment-Clerical sex-Female;
12 < 1 > age->=50 education-Bachelors US-citizen =[100%]=> < 1 > employment-Managerial sex-Female;
13 < 1 > age->=50 education-HS-grad sex-Male US-citizen =[100%]=> < 1 > employment-Managerial;
14 < 1 > age->=50 employment-Managerial sex-Male US-citizen =[100%]=> < 1 > education-HS-g
15 < 1 > age->=50 employment-Unskilled sex-Male US-citizen =[100%]=> < 1 > education-11th;
16 < 3 > age->=50 =[100%]=> < 3 > US-citizen;
17 < 1 > education-Bachelors sex-Male US-citizen =[100%]=> < 1 > age-30to<40 employment-Cle
18 < 3 > education-Bachelors =[100%]=> < 3 > US-citizen;
19 < 1 > education-Masters =[100%]=> < 1 > age-30to<40 employment-Managerial sex-Female;
20 < 1 > education-11th =[100%]=> < 1 > age->=50 employment-Unskilled sex-Male US-citizen;
21 < 1 > education-HS-grad employment-Unskilled sex-Male US-citizen =[100%]=> < 1 > age-30to
22 < 2 > education-HS-grad =[100%]=> < 2 > sex-Male US-citizen;
```

List of rules - Conexp

Matrix view - Cubix



Slide with demo video, removed for th pdf-version of the slides

Content: Graph Exploration Demo

Watch instead: https://www.youtube.com/watch?v=Kuu756nr1_I

Agenda

- Project Setup and Key Technologies
  - First Introduction into CUBIST
  - Use Cases
  - Introduction into Semantic Technologies
  - Introduction into Formal Concept Analysis
  - Key Messages
- CUBIST Prototype
  - Architecture
  - Different Means to Access Information
  - Semantic Search
  - Query Generation
  - Explorative Search
  - Conceptual Scaling
  - Visual Analytics
- Outcome
  - User Evaluation
  - Our Take
  - Conclusions

User Evaluation Methods

- A walk-through for use-case-specific tasks using the prototype by the test users, utilizing the think-aloud-method
- Structured interviews conducted with the test users
- Questionnaires with Likert-scales filled by the test users

1 Interview

- Name: Click here to enter text.
- Use Case (HWU; SAS; INN): Click here to enter text.
- Age: Click here to enter text.
- Gender: Click here to enter text.
- Profession: Click here to enter text.
- Computer Usage per day in hour: Click here to enter text.
- Date of Test: DD/MM/2013 Click here to enter text.
- Location of Test: [City, State] Click here to enter text.
- Please rate your overall computer skills?

  ☐ Very good (e.g. programming, security, data modeling, …).
  ☐ Good (e.g. frequently using spreadsheet applications, advanced in office tools, analysis tools,…)
  ☐ Standard (e.g. surfing, e-mail and writing simple documents, but not much more).
  ☐ Bad.

1.1 For the tasks as conducted:

1. Please shortly describe the tasks you conducted with CUBIST:

   Click here to enter text.

   What do you expect from a system to fulfill these tasks?

   Click here to enter text.

2. Did the system offer you the right information to fulfill your analytical tasks?
   a. If yes, what kind of information and system functionality provided did you find especially helpful?
      Click here to enter text.

	strongly agree			neutral	strongly disagree		n/a
The purpose and function of the component is clear.	☐	☐	☐	☐	☐	☐	☐
The component is easy to understand and use.	☐	☐	☐	☐	☐	☐	☐
The interface is appealing and attractive.	☐	☐	☐	☐	☐	☐	☐
The component is useful.	☐	☐	☐	☐	☐	☐	☐
For some kinds of information needs or queries, particularly this component (or similar components based on the same approach) is useful.	☐	☐	☐	☐	☐	☐	☐
I have similar functionalities in the tools I usually use.	☐	☐	☐	☐	☐	☐	☐

User Evaluation Methods

- Two test users per partner, i.e. six test users in total
- We distinguished between HWU/SAS and INN

For the "Search and Select" Component	With Innovantage		Without Innovantage	
The purpose and function of the component is clear.	6	0,56	4	0,75
The component is easy to understand and use.	6	0,28	4	0,58
The interface is appealing and attractive.	6	-0,06	4	0,42
The component is useful.	6	0,89	4	0,83
For some kinds of information needs or queries, particularly this component is useful.	6	0,72	4	0,75
I have similar functionalities in the tools I usually use.	5	0,20	3	-0,11

Evaluation Workflow

Preparation Phase

- D1.4.1: Directives for the Evaluation of the UC Prototypes
- interview questions
- questionnaire
- tutorial video
- information document

Evaluation Phase

- D7.4.1: Evaluation of UC Prototype against User Expectations
- D8.4.1: Evaluation of UC Prototype against User Expectations
- D9.4.1: Evaluation of UC Prototype against User Expectations

Analysis Phase

- D1.4.2: Evaluation of Final CUBIST Prototype

Evaluation of Overall Prototype

- Overall positively rated
- Useful
- Novel
- Expert tool
- Achieving ease of use requires learning
- Better suited for “non-traditional information needs”
- CUBIST has components/panels which support factual search, explorative search and visual analytics
  - Each component is useful for specific tasks and appreciated
  - Integration of components pay off
  - Usability of integration is challenging

For the overall prototype	With Innovantage			Without Innovantage		
The CUBIST software was easy to use and work with.		6	0,17		4	0,25
Neglecting the prototypic character, I would like to use CUBIST in future again.		5	0,33		3	0,89
In future, I would prefer CUBIST to other analytical tools I currently use.		4	-0,08		2	0,33
Using CUBIST software could make my work more effective and efficient.		5	0,20		3	0,67
The integration of different components was helpful for fulfilling my tasks.		6	0,67		4	0,67
The different components and the visualizations in CUBIST are well integrated.		6	0,44		4	0,50
It is clear how the different components interact.		6	0,22		4	0,50
The navigation/interaction functionalities were easy to understand and apply.		6	0,44		4	0,58
It was easy to follow the system's steps when using the interaction functionalities.		5	0,13		3	0,67

Comparison of the components

- **“Search and Select”**
  - Most useful
  - Positive tendency to being easily used
  - Appealing
  - Not very novel
- **“Explore Selection”**
  - Very useful
  - Clear purpose
  - Appealing and attractive
  - Most novel
- **“Navigate in Data”**
  - Slightly useful
  - Purpose is not too clear
  - Bot novel as all
- **“Analyse Selection”**
  - Useful particularly in the “non-traditional-BI-use cases”
  - Novel
  - Ease of use, and the appeal and attractiveness: badly rated

Evaluation of Search and Select

- Very easy to use
  - Allows easy browsing through data
  - Allows easy searching (filtering) for specific events
- Storing queries in URLs is helpful.
- Concrete tips on how to still improve the interface
  - actual minimum and maximum values in the filter ranges
  - “select all” option in the filter;
  - distinction between selected and not selected parameters
  - greying out facets with no data.

For the "Search and Select" Component	With Innovantage		Without Innovantage	
The purpose and function of the component is clear.	6	0,56	4	0,75
The component is easy to understand and use.	6	0,28	4	0,58
The interface is appealing and attractive.	6	-0,06	4	0,42
The component is useful.	6	0,89	4	0,83
For some kinds of information needs or queries, particularly this component is useful.	6	0,72	4	0,75
I have similar functionalities in the tools I usually use.	5	0,20	3	-0,11

Evaluation of Explore Selection

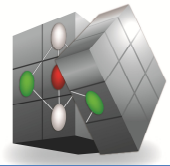
- Not evaluated by SAS
- Useful
- Clear purpose
- Novel

For the "Explore Selection" Component	With Innovantage		Without Innovantage	
The purpose and function of the component is clear.	4	0,50	2	0,67
The component is easy to understand and use.	4	0,50	2	0,67
The interface is appealing and attractive.	4	0,17	2	0,67
The component is useful.	4	0,67	2	0,67
For some kinds of information needs or queries, particularly this component is useful.	4	0,67	2	0,67
I have similar functionalities in the tools I usually use.	3	-0,56	1	-1,00

Evaluation of Visual Analytics

- Very novel
- integration of different visulisations helps to fulfill tasks (for HWU/SAS)
- Hasse-diagrams pay off
  - Even diagrams which are in the beginning hard to understand
- Interaction, particularly filtering, appraised
- not very appealing
- Not easy to use for novices

For the "Analyse Selection"-Component	With Innovantage		Without Innovantage	
The purpose and function of the component is clear.	5	0,20	4	0,42
The component is easy to understand and use.	5	-0,27	4	-0,25
The interface is appealing and attractive.	5	0,00	4	0,17
The component is useful.	5	0,33	4	0,50
For some kinds of information needs or queries, particularly this component is useful.	5	0,33	4	0,50
I have similar functionalities in the tools I usually use.	4	-0,42	3	-0,89
The visualizations were easy to understand.	5	-0,07	4	-0,17
There are visualizations available that did fit my tasks very well.	4	0,42	3	0,44
The integration of different visualizations was helpful for fulfilling my task.	4	0,42	3	0,67
It is clear how the different visualization interact.	5	-0,13	4	-0,08



Two nice quotations ☺

I’m a big fan of Formal Concept Analysis, and the lattice visualization.

Saliha Klei (certified SOLAR operator at SAS)

I like to see more, this is fantastic!

Chris Armit (chief editor of EMAGE)

Agenda

- Project Setup and Key Technologies
  - First Introduction into CUBIST
  - Use Cases
  - Introduction into Semantic Technologies
  - Introduction into Formal Concept Analysis
  - Key Messages
- CUBIST Prototype
  - Architecture
  - Different Means to Access Information
  - Semantic Search
  - Query Generation
  - Explorative Search
  - Conceptual Scaling
  - Visual Analytics
- Outcome
  - User Evaluation
  - Our Take
  - Conclusions

CUBIST is a Prototype

Problems

- Some basic features missing
- Stability
- Performance
- Visual Analytics are cluttered, layout problems

Using Semantic Technologies for BI

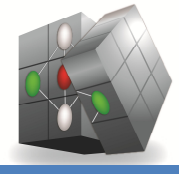
- **Good:** schema last (CUBIST would not work with RBDMS)
- **Good:** using ontologies, there is no separation between “data schema” and a “semantic layer” needed
- **Good:** graph-based schema good for graph exploration
- **Good:** Beyond SoA for ST
- TS: graph db is suited for *specific* use cases.
- **Challenge:** performance w.r.t. some BI-related queries
  - TS not good at operational queries
  - TS is essentially *transactional* repository.

FCA in CUBIST

- **Good:** Acting on “real data” and “real data repository”
- **Good:** Powerful generation of formal context on the fly
  - “FCA-BI as a self service”
- **Good:** Conceptual scaling on the fly
- **Good:** Powerful FCA visualizations
  - Highly interactive
  - Different visualizations
  - Combinations with graphs and traditional BI
- **Challenge:** Layout, usability

Agenda

- Project Setup and Key Technologies
  - First Introduction into CUBIST
  - Use Cases
  - Introduction into Semantic Technologies
  - Introduction into Formal Concept Analysis
  - Key Messages
- CUBIST Prototype
  - Architecture
  - Different Means to Access Information
  - Semantic Search
  - Query Generation
  - Explorative Search
  - Conceptual Scaling
  - Visual Analytics
- Outcome
  - User Evaluation
  - Our Take
  - Conclusions



CUBIST Lessons Learnt

CUBIST provides a glimpse at my FCA “dream system”

- Acting on real data
  - Adding data sources on the fly (e.g. connectors to linked data)
- Acting on large data / big data
  - Data preprocessing is needed before contexts are generated
  - Still high-performance concept mining needed
    - e.g. parallel processing (Hadoop, you name it …)
- Interaction in future BI systems and future FCA systems is key
- Visual transformations of lattices when context is changed
  - This requires mathematical investigations
- Combination of FCA and other analysis means (graphs, traditional charts)
  - Linking and brushing
- “Fuzzy” and “Fault-Tolerant BI”
- New kinds of diagrams / lattice visualizations

Final Recommendations from Evaluation

Proposed recommendation 1: Future BI tools should not only focus on the analysis (in the BI understanding) of data, but on the search in data and the exploration of data as well. Integrating different components which target different information needs is challenging and needs further investigations.

Proposed recommendation 2: It is very reasonable to have faceted search based frontend in future BI-solutions for searching and filtering the data. The evaluation gives clear hints on which filtering functionalities are requested by the users.

Proposed recommendation 3: Future BI solutions, which aim at providing means to explore the data, should incorporate functionalities which resemble the functionalities of the “Explore Selection” Component. Designing the interface for such exploration means deserves closer attention.

Proposed recommendation 4: Future BI-tools should comprise quite different Visual Analytics means, ranging from traditional to novel ones (e.g. graph-based). One should not hesitate to include unfamiliar, sophisticated visualizations into expert BI tools, even if those visualizations are not ease to digest from the very beginning.

Links

Links

- www.cubist-project.eu
- https://www.youtube.com/user/CUBISTFP7ICT

Open Source

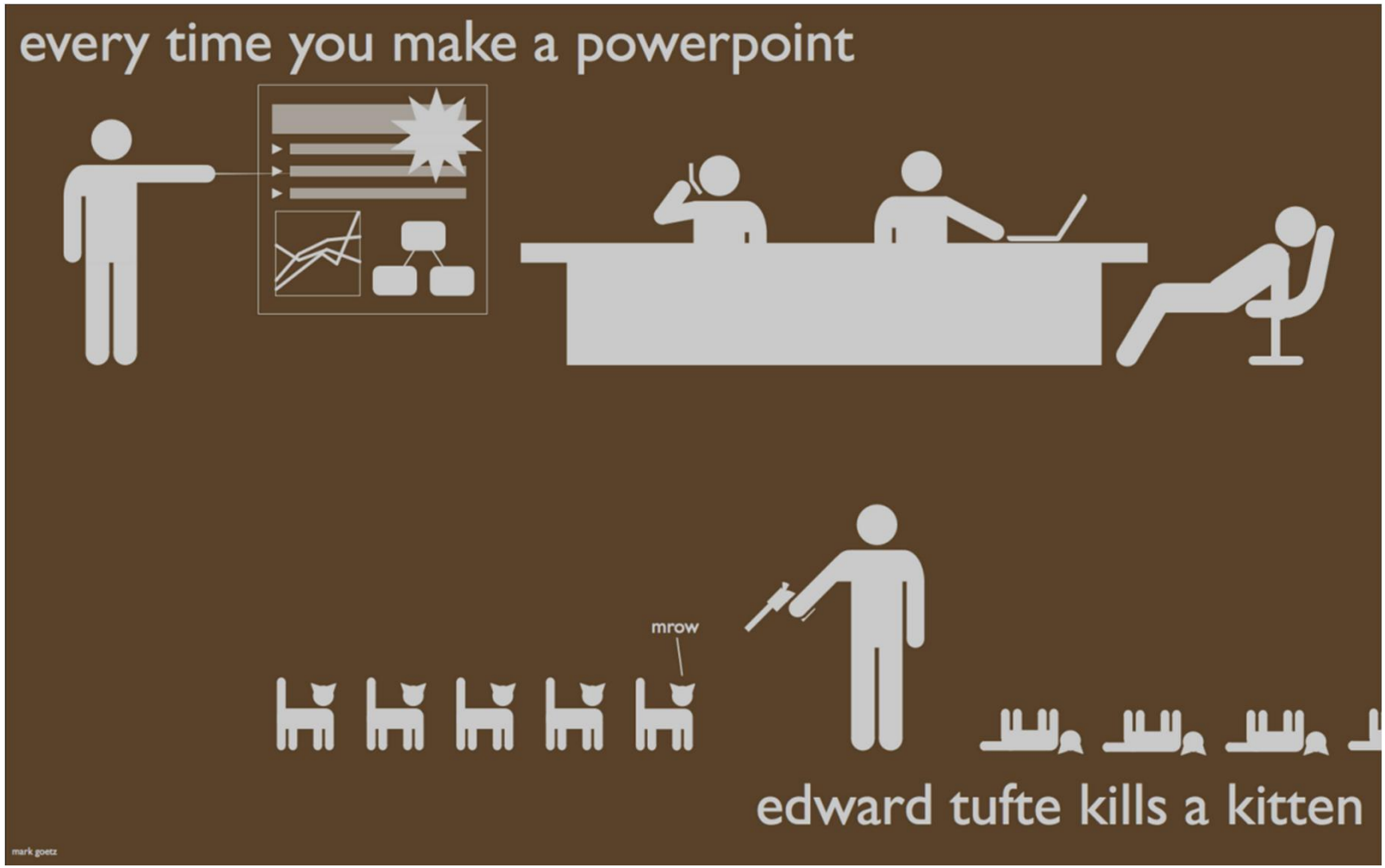
- FCAService:https://github.com/acesco1/rdf2fca-service
- CUBIX: https://github.com/ksiomelo/cubix

Scientific:

- Special CUBIST Edition of the International Journal of Intelligent Information Technologies (IJIIT
- Workshop
- Talks etc

Me ☺

- Frithjof.dau@sap.com


every time you make a powerpoint
mrow
edward tufte kills a kitten
mark goetz

Thank You!