

Cubist: Structure

[illegible]

```
<w:LsdException Locked="false" Priority="19" SemiHidden="false" UnhideWhenUsed="false"
QFormat="true" Name="Subtle Emphasis" /> <w:LsdException Locked="false" Priority="21"
SemiHidden="false" UnhideWhenUsed="false" QFormat="true" Name="Intense Emphasis" />
<w:LsdException Locked="false" Priority="31" SemiHidden="false" UnhideWhenUsed="false"
QFormat="true" Name="Subtle Reference" /> <w:LsdException Locked="false" Priority="32"
SemiHidden="false" UnhideWhenUsed="false" QFormat="true" Name="Intense Reference" />
<w:LsdException Locked="false" Priority="33" SemiHidden="false" UnhideWhenUsed="false"
QFormat="true" Name="Book Title" /> <w:LsdException Locked="false" Priority="37"
Name="Bibliography" /> <w:LsdException Locked="false" Priority="39" QFormat="true" Name="TOC
Heading" /> </w:LatentStyles> </xml>
```

Workplan and Workpackages

The composition of WPs in CUBIST reflects the logical and chronological structure of the envisioned CUBIST architecture and follows well designed project logic. We first give an overview of the WPs and then describe the project logic, i.e. the dependencies between WPs, tasks, deliverables, and milestones:

WP1 - *Technological Architecture* is a spanning WP which lays down the technological foundations and requirements for the overall CUBIST architecture. It gathers input from all other technological WPs as well as from the use cases, and will provide directives for the implementation and evaluation of the system to be developed.

WP2 - *Semantic ETL and Data Integration* covers tasks related to harvesting unstructured and structured data from a variety of heterogeneous sources. This includes software components like wrappers or crawlers, and covers research directions like data fusion: unifying and enriching the data acquired from a variety of sources in order to make it accessible for BI operations.

WP3 - *Semantic Data Warehouse* deals with extending RDF triple stores in order to make them applicable in the domain of BI. The two major outcomes from this workpackage will be 1) extending the standard RDF query language, SPARQL, with OLAP query functionality for aggregates, reporting and cube/rollup analysis; 2) improving the performance of RDF triple stores in typical query scenarios for BI by adapting advanced indexing and materialisation techniques commonly applied in state-of-the-art data warehouse systems.

WP4 - *Analyzing and Visualizing Data* deals with all aspects which are related to the user interaction with the envisioned CUBIST system. Due to the general approach to provide visualization and analytical features based on FCA, WP4 particularly covers tasks which deal with selecting and computing appropriate contexts based on the overall available data of the CUBIST DW, or in scrutinizing different approaches for visualizing subsets of huge concept lattices. Moreover, various innovative user paradigms such as faceted navigation, visual querying of data, or the aggregated representation of data in form of dashboards are investigated. Feedback about these approaches from the use case scenarios are taken into account.

WP5 - *Dissemination, Exploitation and Standardisation* deals with dissemination, exploitation and standardisation issues.

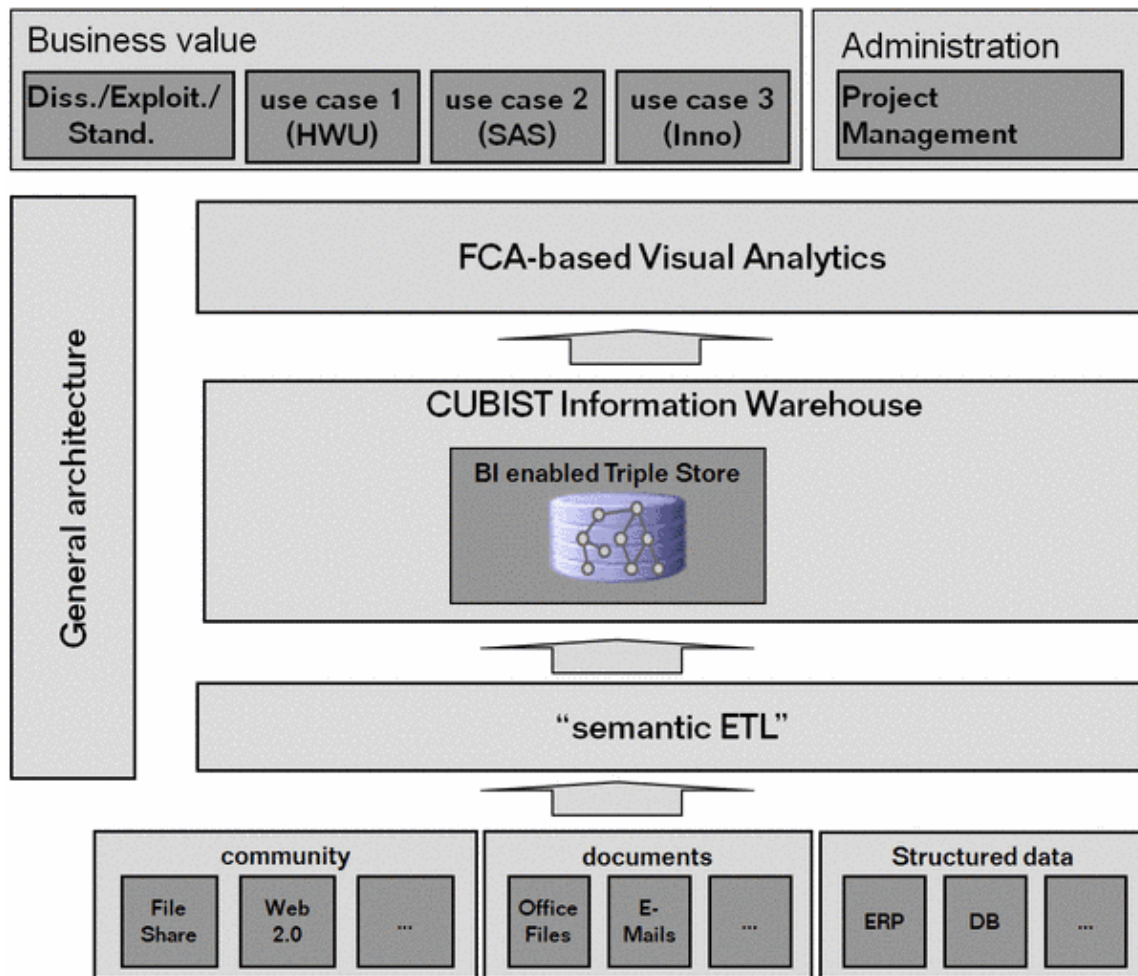
WP6 is dedicated to the project management of the CUBIST project.

WP7 - *Use Case: Biomedical Atlases* is a use case in the field of biomedical research. In this use case, two spatio-temporal biomedical atlases which consist of annotated 3D image reconstructions of mouse embryos and mouse brains, respectively, are brought together with three gene expression databases.

Cubist: Structure

WP8 - Use Case: Semantic Business Intelligence for Space Control Centres is a use case involving space control centres. Mission control rooms in space control centres use heterogeneous sources of information, including structured and unstructured data, for decision making and information tracking. Very large volumes of data are obtained, especially with telemetry data that are generated every second over large periods of time. This use case attempts to aggregate various information sources available to operators in mission control rooms using the technology provided by CUBIST. Aggregated data, ready for the BI processing, is expected to provide online support for taking better decisions, reveal hitherto undiscovered information and provide supportive evidence in debriefing and decision making processes related to the organisation of space control centre operations.

WP9 - Use Case: Semantic Business Intelligence for Recruitment is a use case for BI for the UK jobs market. It includes Market Intelligence (who is recruiting, where and when, how do they they recruit, what is the trend of jobs and specific sectors in the market, who are the biggest advertisers) and Competitive Intelligence (provisioning of data to UK employers to help track and better understand the recruitment activity of their competitors in addition to the talent pool). The major outcomes expected by participating in CUBIST are to: 1) harvest more data sources than currently, e.g. unstructured data sources such as blogs, forums, corporate wikis, tweets, etc.; 2) improve the reliability and performance of the current system, so that the negative impact of traditionally incomplete and inconsistent data in unstructured data sources is reduced; 3) generate and keep up-to-date a skills taxonomy for the UK jobs market; 4) use sentiment mining for UK employers.



Architecture and Workpackages