

GENASIS System Architecture

On the way from environmental data repository towards research infrastructure

<u>Richard Hůlek</u>, Jiří Jarkovský, Miroslav Kubásek, Jana Klánová, Jakub Gregor, Kateřina Šebková, Jana Borůvková, Jiří Hřebíček, Ivan Holoubek, Ladislav Dušek Institute of Biostatistics and Analyses & Research Centre for Toxic Compounds in the Environment, Masaryk University

10.10.2013

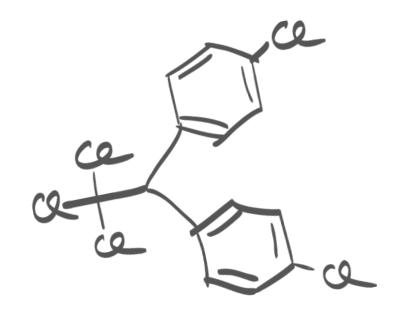






Content

- 1. Introduction
- 2. GENASIS overall system architecture
- 3. System components
- 4. Conclusions





-

Research Centre for Toxic Compounds in the Environment

I. I.

Learn, discover, prove and apply





GENASIS – Project Introduction

Global ENvironmental ASsessment Information System

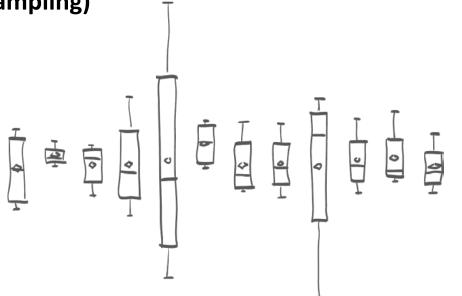
- Provides a comprehensive information on contamination of the environment by chemicals.
- Focused on **persistent organic pollutants** (POPs).
- Provides up-to-date information on spatial and temporal trends in POPs concentrations in the environment.

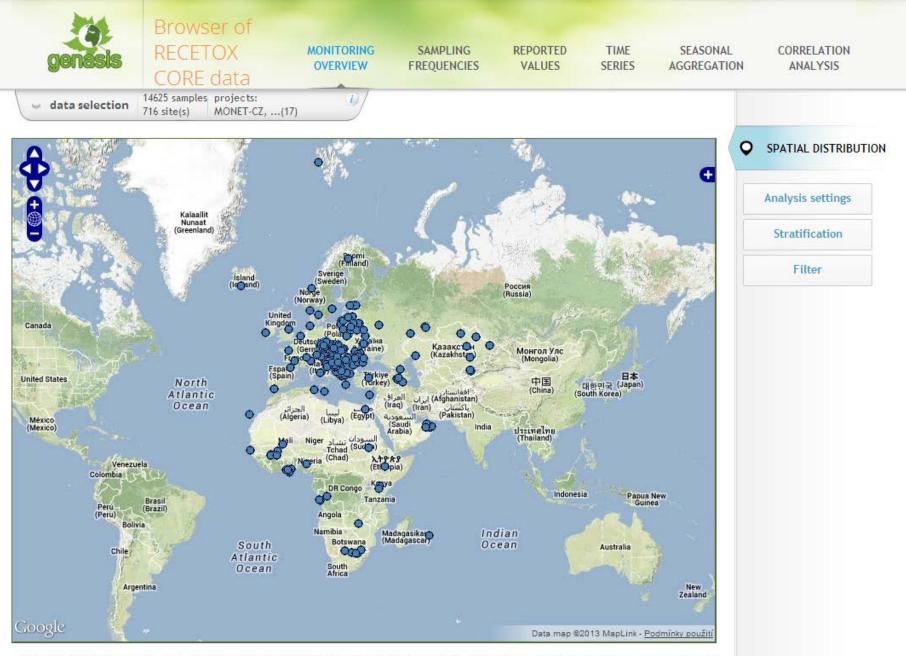
GENASIS – Project Introduction

- Combines expertise and validated data from:
 - Regular environmental monitoring programs
 - Partner institutions
- Provides :
 - Data repository
 - Analytical tools
 - Data management
- GENASIS handles :
 - Primary data
 - No data aggregations or modifications

Data Content

- Data repository for arbitrary data about **concentration of chemical pollutants** in the environment.
- Various environmental matrices:
 - Air (both active and passive sampling)
 - Water
 - Soil
 - Biota
 - Human tissues

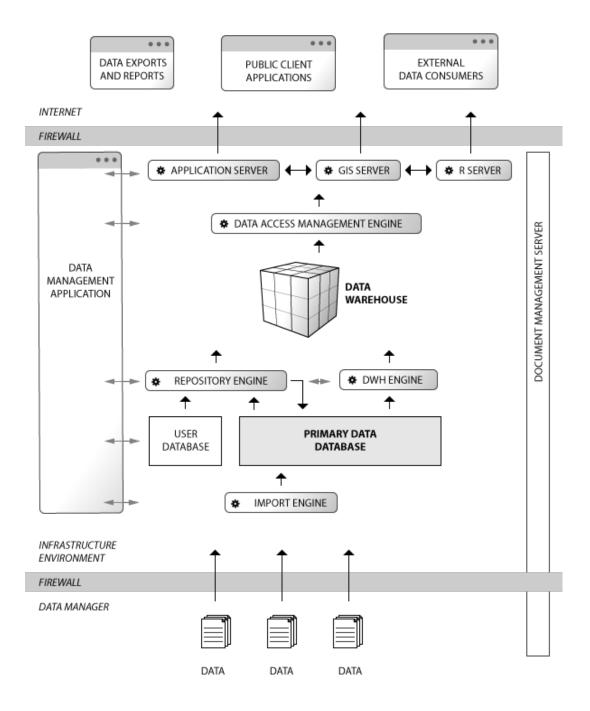




MONET-CZ KOSETICE-INTEGR ZI IN MONET-EU MOKRA MONET-CEEC MONET-AERICA APOPSBAL IMK BEROUN SPOLANA MONET-EU LIBEREC SCK ZU

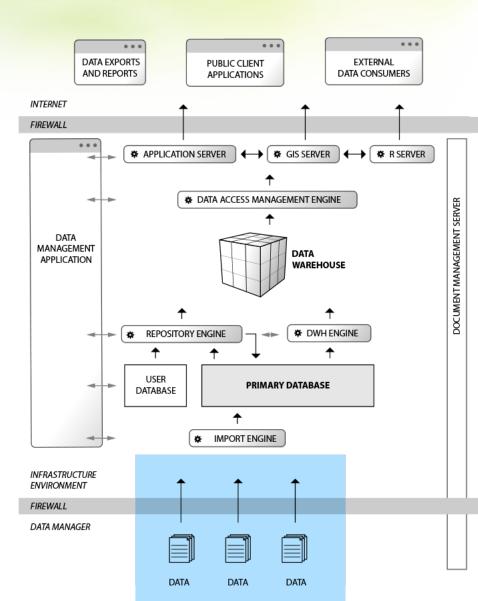
System Architecture Overview

- Data repository
- Controllers and engines
- Services
- Client applications
- Admin panel
- Supporting tools



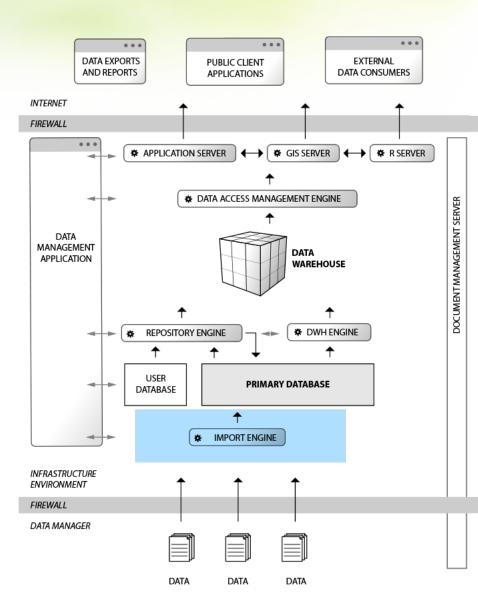
Data Providers and Data Validation

- Gathering data from various environmental monitoring networks from different data providers.
- Data validation
- Harmonization
- Transformation to common data format



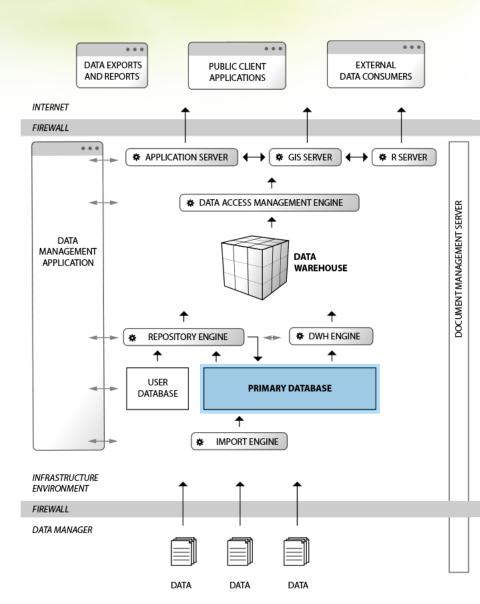
Import Engine

- System component which ensures proper import of data to the primary database.
- Transfer data to database
- Format validation and error detection
- Transactional processing

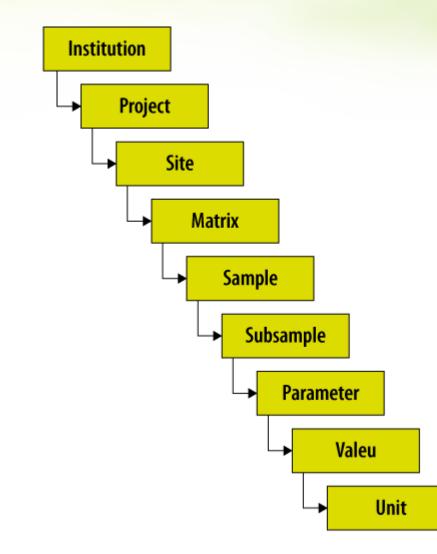


Primary Data Repository

- Storage capacity designed to handle heterogeneous data from environmental monitoring networks.
- Organized data structures
- Classification, categorization, detailed description
- Data standardization = comparability

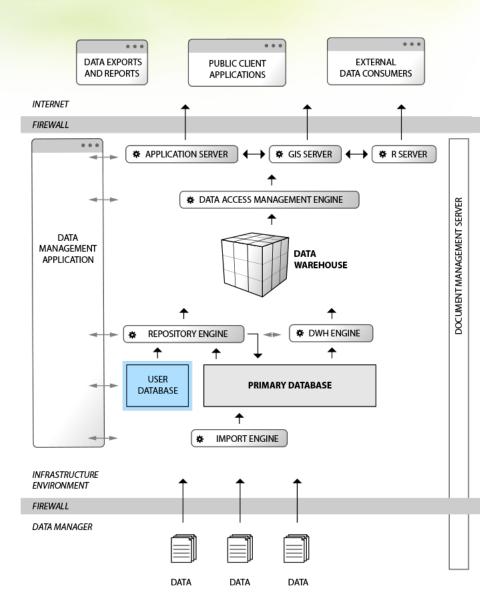


Data structure



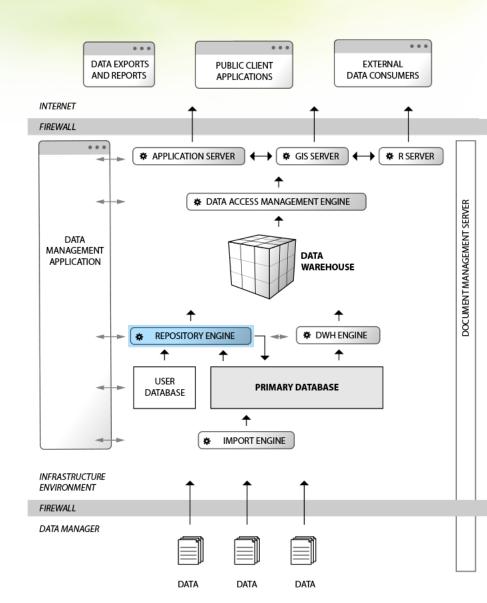
User Database

- User database serves as helper database of the system.
- User accounts
- Access rights



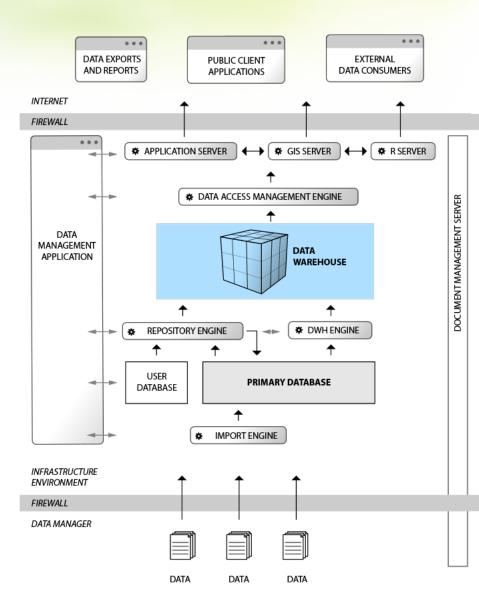
Repository Engine

- Repository engine expands functionality of primary database.
- Quality assurance heuristics
- Transformation procedures
- Procedures run in batches



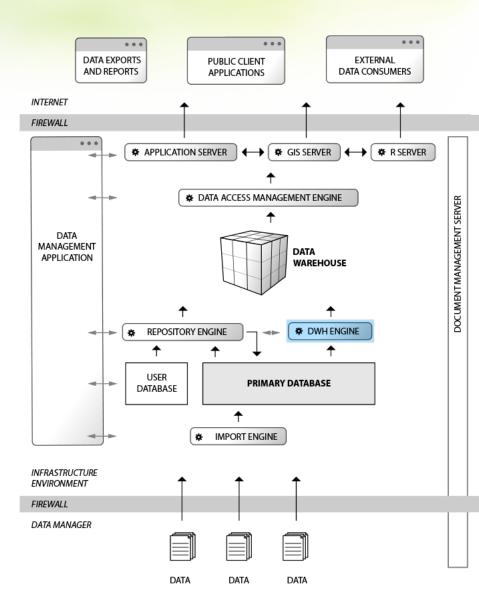
Data Warehouse

- Data warehouse database contains transformed and pre-processed data prepared to be delivered to end users.
- Integration of data from various sources/providers
- Preprocessing of complex computations and aggregations.



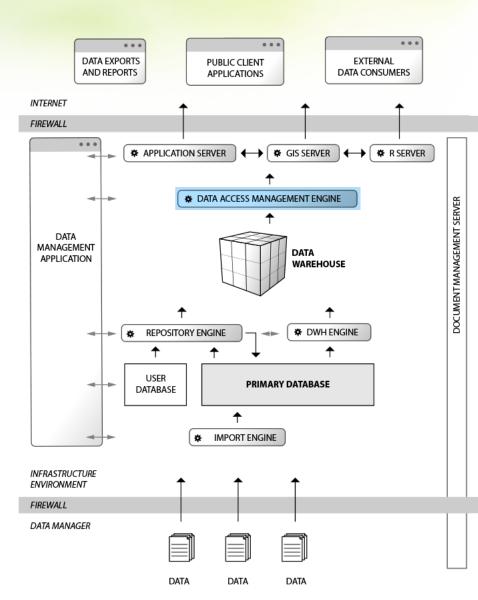
Data Warehouse Engine

- Data warehouse engine connects primary data database and data warehouse.
- Creates and performs updates data warehouse based on changes in the primary database.
- Implements all algorithms of calculations, transformations and aggregations.



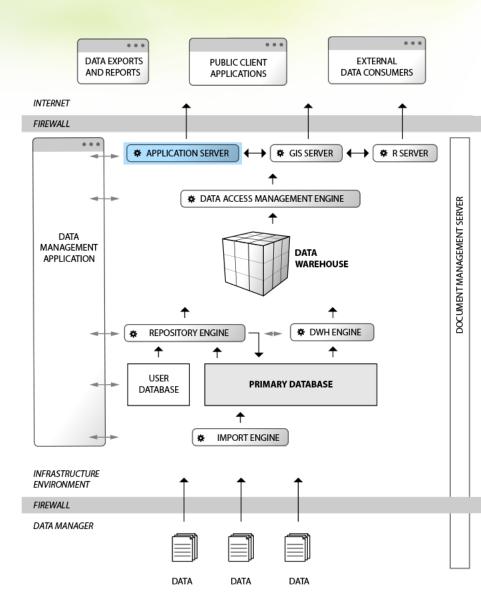
Data Access Management Engine

- Management of access rights and access scopes
- Various data-providers
- Various outputs



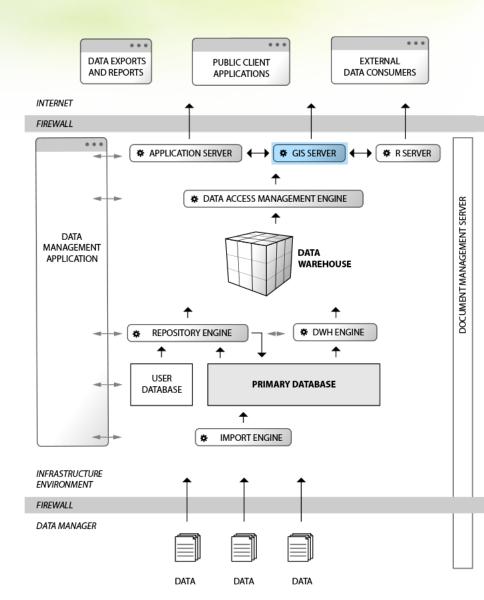
Application Server and Data Services

- Application server provides environment for various applications which are build on top of the GENASIS infrastructure.
- Connectivity to other compo-nents of the system
 - Data warehouse database
 - GIS server
 - R server

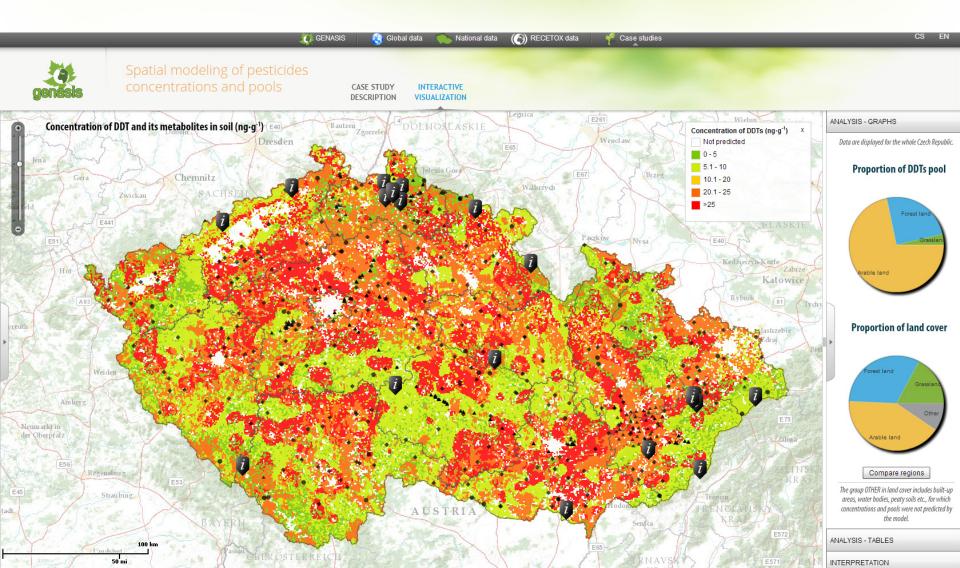


GIS Server – Spatial Services

- GIS server extends functions of the GENASIS system.
- Spatial data visualizations
- Map distribution
- Performing spatial analysis in real-time

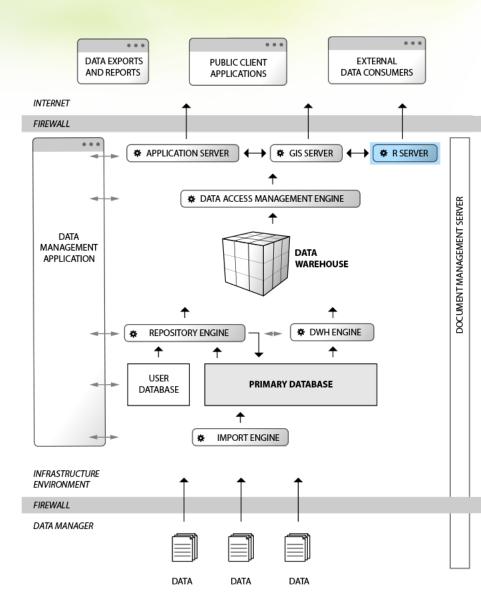


GIS Services in practice



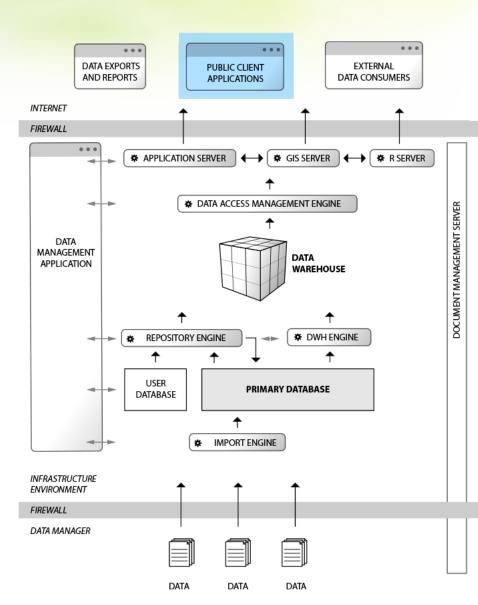
R Server – Computation Services

- *R statistical software was integrated as a server component.*
- Implemented as a service
- Make use of available libraries and software packages
- Sharing hardware resources



Publicly Available Applications

- Web based apps brings results to broad range of users without a need of specific skills.
- Analytical and interpretation tools are connected directly to the database.
- GENASIS on-line data browser.



Public Applications



MONITORING OVERVIEW

An interactive world map displaying sampling sites selected for analysis.



 \sum

 \sum

 \sum

SAMPLING FREQUENCIES

The charts show sampling frequency of individual compounds in selected data.

 \sum

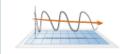
 \sum

>>>



REPORTED VALUES

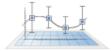
The charts are designed for direct presentation of reported values and concentrations.



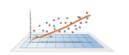
TIME SERIES

The charts show time trends of compounds concentrations together with trend analysis.





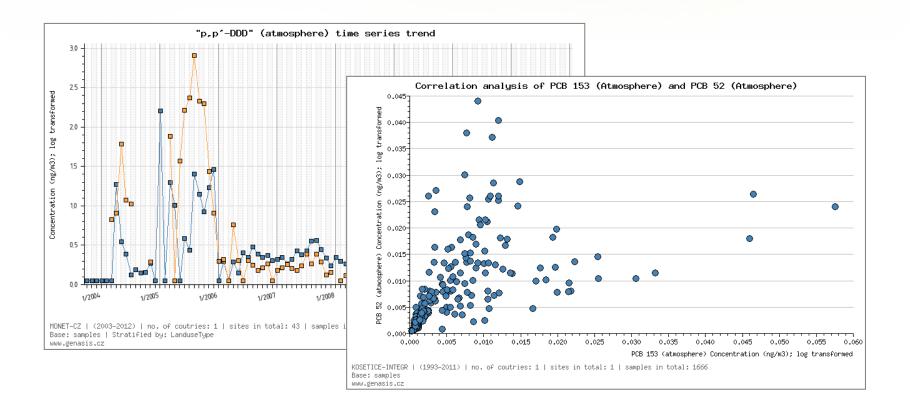
Seasonal changes of compounds concentrations in selected data are analysed in set of charts.



CORRELATION ANALYSIS

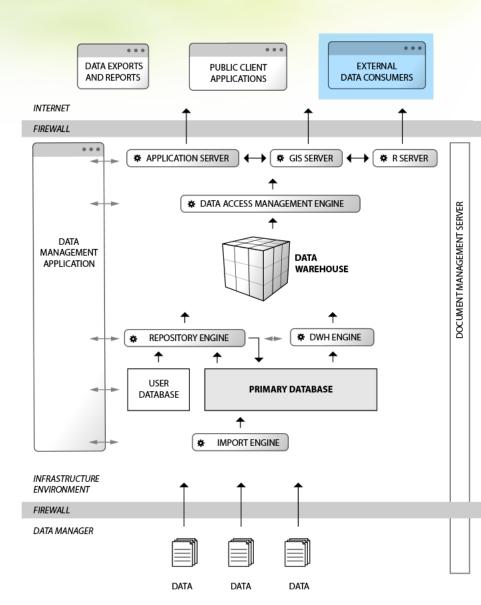
Correlations between compounds within or between matrices are provided for selected data.

Analytical tools



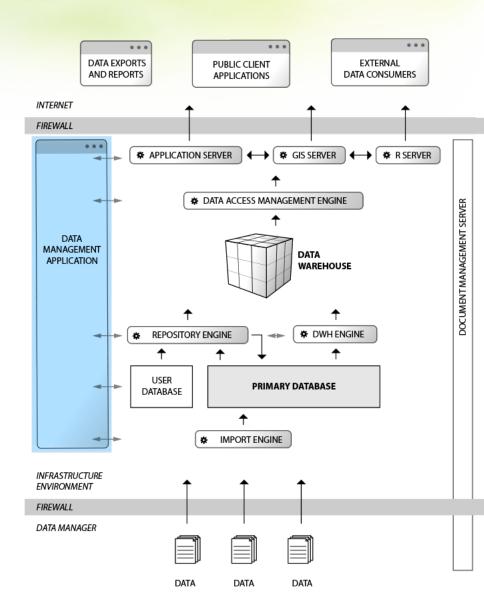
Interoperability

- Service-oriented and modular architecture
- Each system component can offer its services over dedicated interface
- Publication of spatial data (WMS)
- Calculation services
- Automated data exports for GEOSS



Data Management Application

- Management and control
 - Data imports
 - User database
 - Repository engine
 - Data warehouse (engine)
- Data exports of primary data

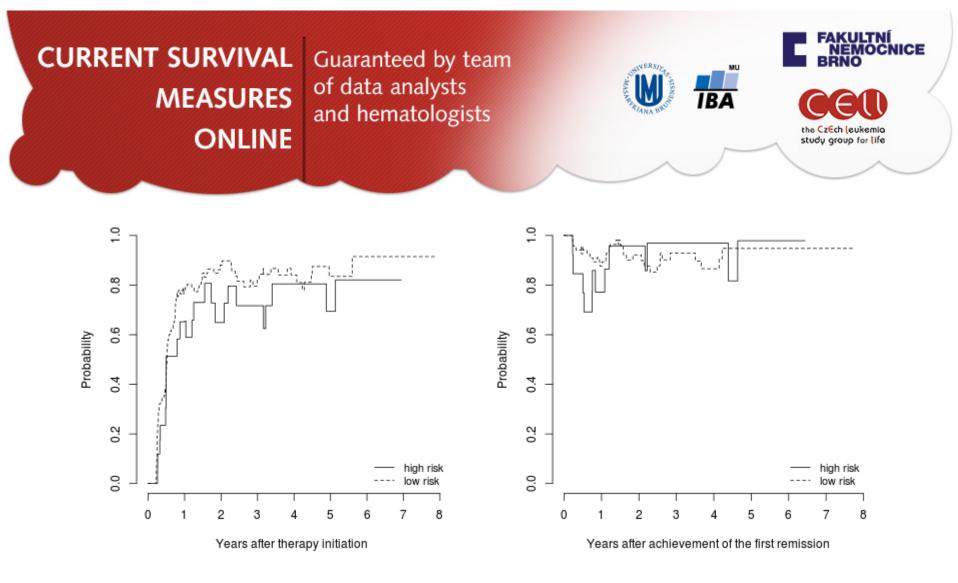


Conclusions

- GENASIS system and ICT tool represents important part of scientific infrastructure
 - Scientific background
 - Laboratories, monitoring capacities
 - Software tools
 - Data analytics
- Infrastructure is offered and used by third parties
 - MONET EU, MONET Africa
 - Environment of Canada (GAPS)
- Infrastructure is used in another projects
 - UNEP (Global Monitoring Plan)
 - Clinical assessment tools (Current survival measures)



Example of usage #1: Current Survival Measures Online



http://www.iba.muni.cz/data-analysis-tools/currentSurvival/

Example of usage #2: Global Monitoring Reports Visualizations



http://www.pops-gmp.org/visualization/

Thank you for your attention



www.genasis.cz

Richard Hůlek hulek@iba.muni.cz