Rosemary

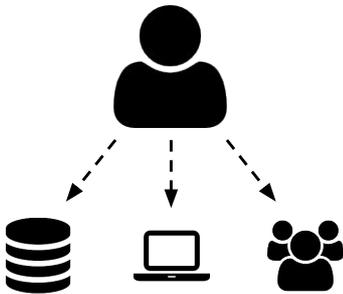*Make Complex Collaborative Big Data Analysis Easy*

The Problem

Data-driven Era

Scientists and decision makers have to:
- Tap into complex, big, and heterogeneous datasets
- Use various tools to process and analyze such datasets
- Collaborate closely with others for complementary expertise

Extracting Information From (Big) Data

Big data infrastructures are generally composed of several complex data and computing services, resulting in:
- Heterogeneous, distributed, and evolving software components and services
- Interfaces often exposing low-level technological details
- Required configuration and customization for specific applications
- Fragmented services, requiring manual and complex logistics

The Solution

Web-based customized interfaces have been successfully adopted to provide easy access to potentially complex systems. Here we exploit the particular concept of "Science Gateways". Science Gateways integrate and customize infrastructure services into one web-based system to provide workflow automation, increase usability and scalability, and guarantee security and traceability.

The AMC e-Science research group has six years of experience in the research, design, development, and operation of five generations of Science Gateways for biomedical research. These gateways are deployed and used by researchers in real use-cases. The gateways have been proven beneficial for researchers. With the gateways, researchers can perform complex collaborative big data analysis on infrastructures easier and quicker. As a result, gateway users could perform better high-quality research, which is shown by higher number of publications in prestigious international journals.

The latest generation is coined "Rosemary", which is a software platform that can be customized programmatically to develop Science Gateways for various domains and applications.

Rosemary Functions

Data Management
Integrate heterogeneous datasets, organize data in well-defined structures, browse datasets, search and filter data based on metadata, import and export datasets to/from various data sources using various formats and protocols

Computing Management
Provide a repository of tools (apps) with version control, ability to process datasets with tools on remote high-performance infrastructure, automate logistics such as transporting data and piping them through various tools, search and filter performed data processes

Collaboration Management
Fine-grained role-based access to datasets and tools for teams, notification center for latest news within teams, integrated message threads with reference to datasets and data processes

Traceability Management
Tracks data manipulations in a well-defined structure, enables traceability (provenance) and reproducibility

Implemented Big Data Solutions

Computational Neuroscience	In Vitro Fertilization	Genomics
• complex and multi-site • neuroscience data management • computing management on distributed computing platforms (grid) • collaboration management among scientists • traceability management	• IVF data management • privacy sensitive data sharing management • multi-site sharing authorization management • traceability management	• complex wet-lab and dry-lab data management • collaboration management among scientists • traceability management

Acknowledgement

Contact Us

AMC e-Science

website: **www.ebioscience.amc.nl**
e-mail: **contact@ebioscience.amc.nl**