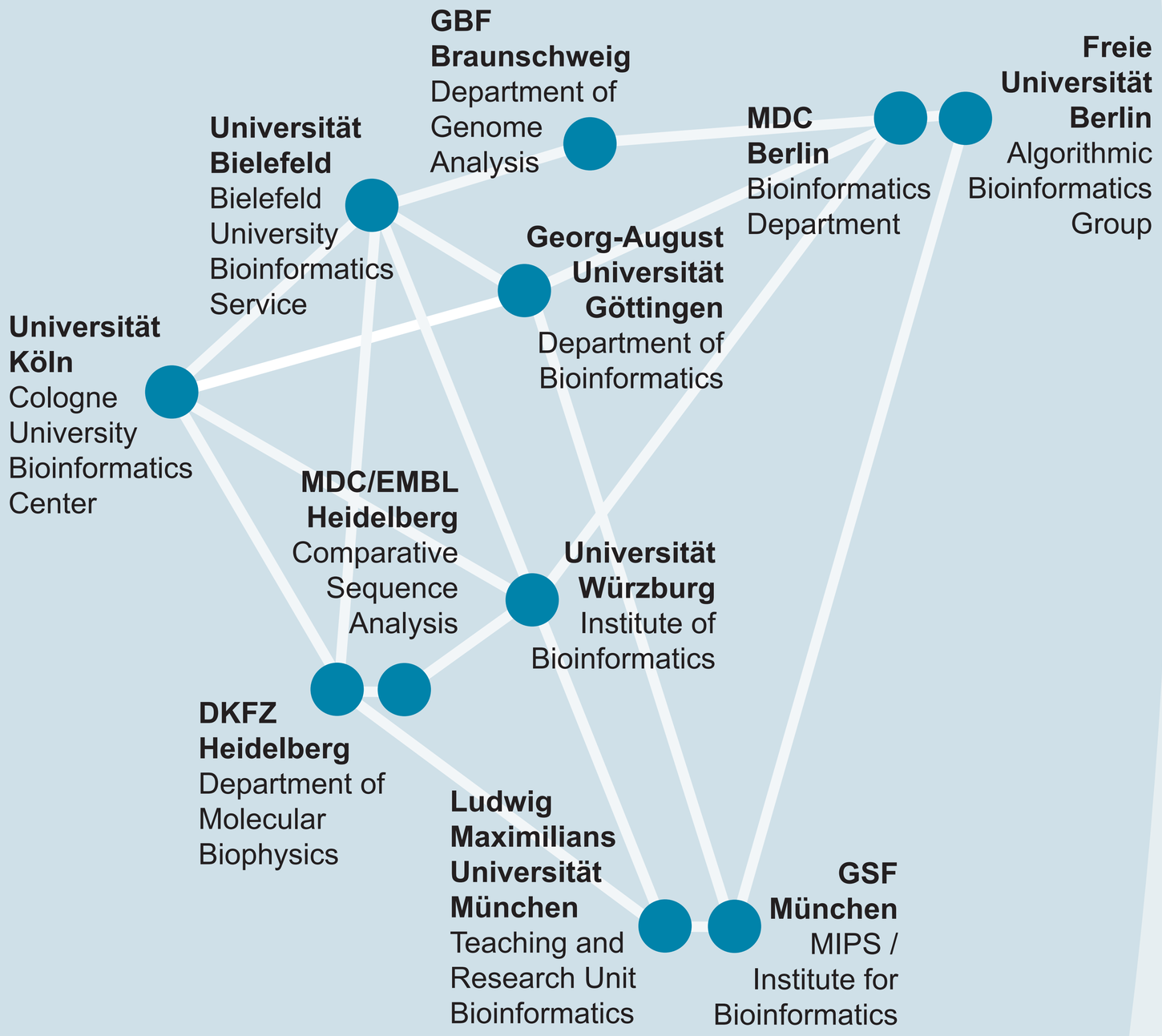
HOBIT: A Network for Integration of Data and Resources based on Industry-Standard Web Services

Volker Stümpflen[1], Philipp Seibel[2], Henning Mersch[3], Björn Bredohl[4], Peter Ernst[5], Andreas Kuntzag[6], Eva Lange[7], Jens Warfsmann[8], Knut Schwarzer[9], Martin Szugat[10]

MIPS/ Institute for Bioinformatics, GSF[1]; Institute of Bioinformatics, Universität Würzburg[2]; Bielefeld University Bioinformatics Service, Universität Bielefeld[3]; Department of Genome Analysis, GBF Braunschweig[4]; Department of Molecular Biophysics, DKFZ Heidelberg[5]; Bioinformatics Department, MDC Berlin[6]; Algorithmic Bioinformatics Group, Freie Universität Berlin[7]; Bioinformatics Center, Cologne University[8]; Department of Bioinformatics, Georg-August Universität Göttingen[9]; Teaching and Research Unit Bioinformatics, Ludwig Maximilians Universität München[10]

Introduction

The urgent need for the flexible integration of data resources and applications is a challenge and hurdle for the progress in comparative genomics and e-bioscience. Although several proposals such as DAS and BioMoby have been developed and implemented, they do not reveal the full potential of state-of-the-art distributed technologies. Drawbacks of these approaches are not only programming language dependencies but also missing standards for data exchange. The Web Service technology as proposed by the W3C consortium and the WS-I (Web Service Interoperability Organization) offers meanwhile an established and mature solution to develop components designed expressly to meet the challenges of linking heterogeneous distributed biological information systems.



Current situation

Commonly bioinformatics resources are highly localised and only accessible via interactive web pages. This causes several disadvantages. It complicates for example the possibility to integrate external resources in local applications. As a result this leads very often to redundant installations of external applications and databases. Consequential information may diverge especially in databases if additional information is generated. Beside this a significant administrative overhead is required. Other problems are for example no uniform access mechanisms for information and the requirement to deal with different technologies.

Requirements

For the interconnection of various heterogeneous bioinformatics resources we identified several demands:

- Open access of databases and applications
  - Platform and programming language independance
- Maintaining investment in legacy systems
- Open standards for data exchange (XML based)
  - Web wide access based on standard web technologies
- Open Source based

Webservices

- Open Source software for development ...
- ... language independent
- Transport via HTTP, FTP or SMTP protocol and the corresponding standard ports
- Transport of standard XML based data formats
- They allow location transparency and high pervasiveness
- Web Services are highly scalable and allow a huge number of nodes
- Technical requirements are very low. Even an Apache web server with Perl CGIs will fit the first needs

HOBIT ...

… an open project for Web Services in bioinformatics like

- Sequence Information and Annotation
- Sequence Matching
- Protein-Interactions
- Protein Structures
- Protein Descriptions
- Genetic Networks
- Textmining

Based on the definition of community agreements for interfaces and data structures

Information

Further information about HOBIT is available at
http://hobit.sourceforge.net

Conclusion

The HOBIT initiative is dedicated to form the core of a network linking bioinformatics centres together. It shall be understood as an initial organisational and technological platform for interconnection of bioinformatics activities. The aim of the network is to concatenate applications and resources in a uniform way so providing an efficient communication tier for bioinformatics resource access.

Funding

We greatfully acknowledge the funding from the "Impuls- und Vernetzungsfonds der Helmholtz-Gemeinschaft Deutscher Forschungszentren e.V."