

# MediGRID - Medical Grid Computing

U. Sax<sup>1</sup>, F. Viezens<sup>1</sup>, Y. Mohammed<sup>1</sup>, T. Lingner<sup>2</sup>, B. Morgenstern<sup>2</sup>, M. Vossberg<sup>3</sup>, D. Krefting<sup>3</sup>, O. Rienhoff<sup>1</sup>

<sup>1</sup> Department of Medical Informatics, University of Göttingen, Germany

<sup>2</sup> Department of Bioinformatics, University of Göttingen, Germany

<sup>3</sup> Department of Medical Informatics, Charité, Humboldt University Berlin, Germany

## Introduction

The MediGRID project [1] within the German D-Grid Initiative combines research from various institutes from the areas of Medicine, Biomedical Informatics, and other Life Sciences. The main goal of MediGRID is the development of a grid-middleware-platform as a basis for e-Science services for the Biomedical community.

## Materials and Methods

Unlike common Grids, data processing in the community biomedicine has to face particular challenges.

Beyond the problem of finding the relevant data sets via metadata description using the adequate searching tools, the access control enforcement is of paramount importance, as the owner of the data are foremost patients. Due to the heterogeneity of the data we need an additional ontology process to homogenize the data. Fig. 1 shows how data flow in MediGRID differs from classical grids by the need to a homogenizing step.

In contrast to data in classical grid applications, high dimensional data are typical for medical applications. Given semantic interoperability the researcher can correlate and analyze the data using suitable biomedical informatics methods. Biomedical data are not only heterogeneous, rather they contain numerous types of information and have different levels of privacy. They vary from aggregated data describing population and diseases (epidemiology, clinical practice, clinical trials), going more granular to patient data and pathological descriptions (health record, clinical history, physical exams), ending with cellular and molecular data (histology, genetic test results and genomic data) [2]. Having these data online with the suitable tools to connect, merge and analyze creates new challenges for data protection and privacy (s. Fig.2) [3].

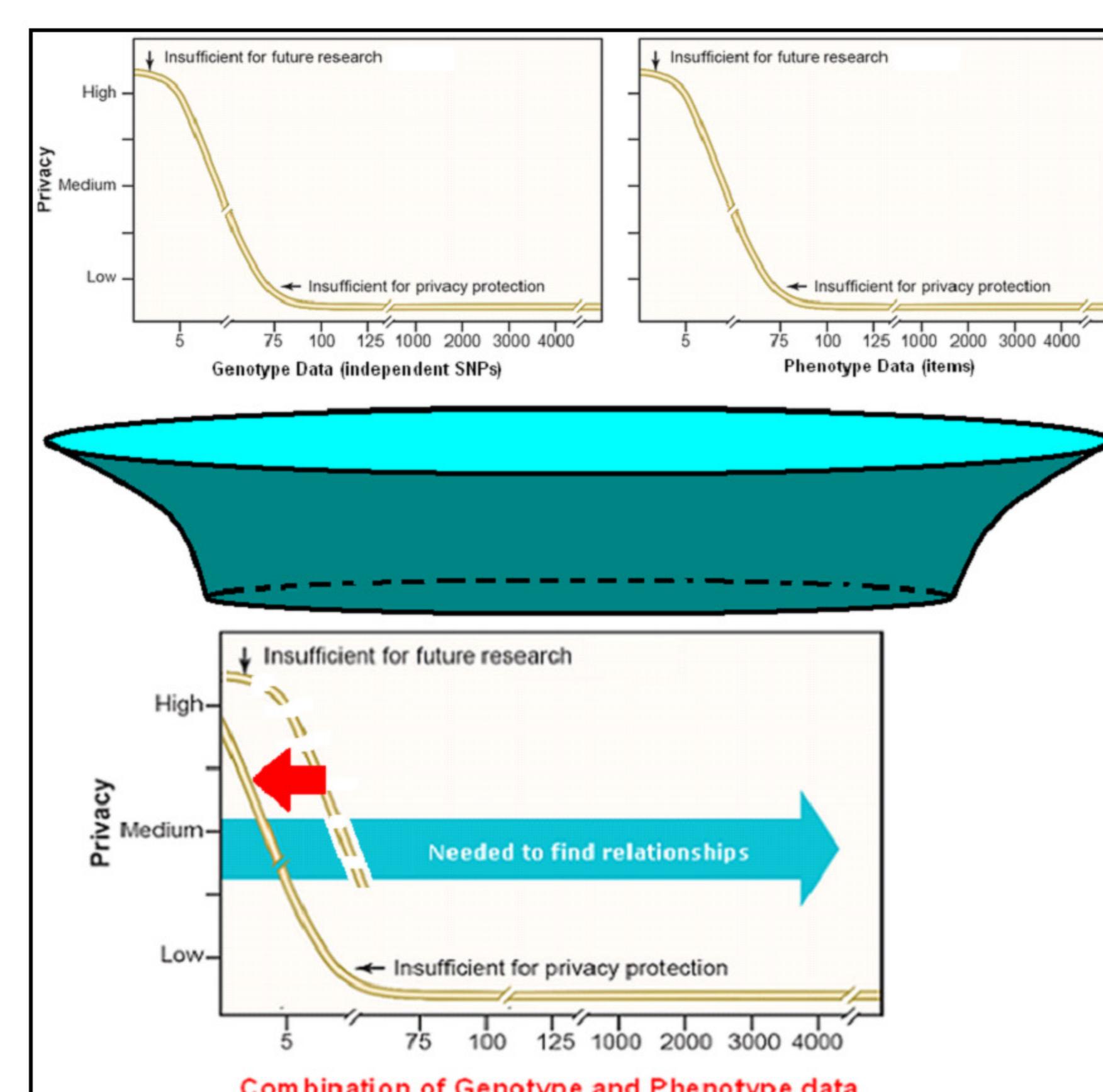


Fig. 2 Privacy and data protection problems in associated studies [3, 4]: combining different data sets, even if the data are anonymized, will increase the re-identification risk. The less items a data set has, the less the re-identification risk is (the higher the privacy is). Presented by the blue cone in the figure a correlation process between two data sets, genotype and phenotype, will shift privacy curve to left. In this case the risk of no privacy in the merged data set will appear resulting that otherwise less items are allowed to be released or the correlation itself should not be done.

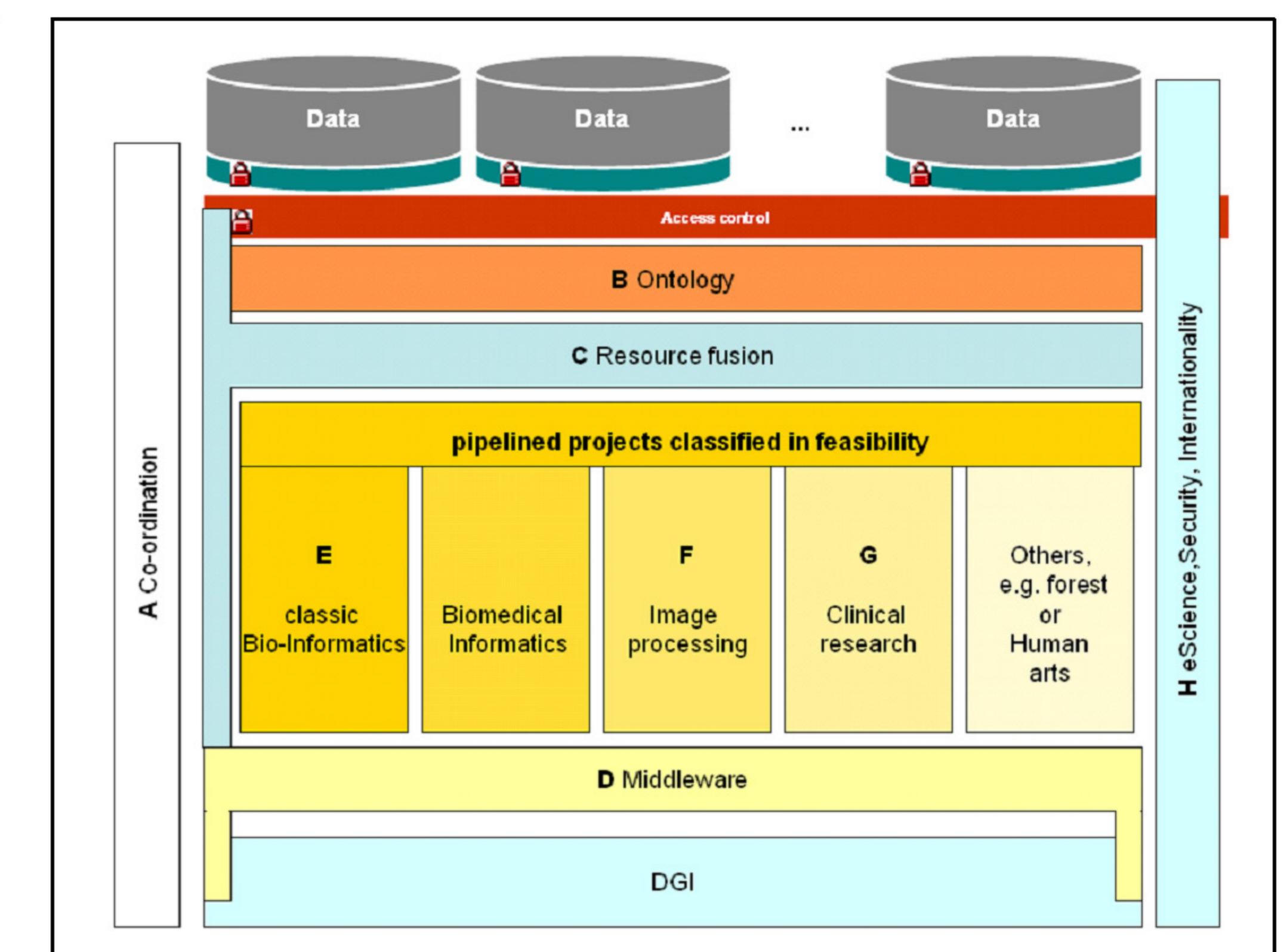


Fig. 3 The MediGRID structure: four methodological modules are responsible to construct a suitable infrastructure: ontology, resource fusion, middleware and e-Science. Three research modules take the initiative to use this infrastructure to assist their work: Biomedical Informatics, image processing and clinical research. The MediGRID consortium develops a middleware component as a connector for the medical community to resources of the integration project in D-Grid [5]. The community modules use this middleware to "gridify" their applications. A pipeline process for the gridifying process is defined. The first cases will cover highly grid enabled projects from biomedical informatics in order to develop best practices for grid enabling other biomedical projects (E, F then G).

## Privacy

Privacy is still the "black hole" [6] in grid computing standards. In medical applications the re-identification risk, the disclosure risk, patient data protection and privacy have to be addressed and assessed. This is rather difficult when talking about correlation of different data sets from different sites, which is of increasing interest for researchers. For genome wide association studies there are severe concerns relating privacy [4,7-9]. Hence, the development of standards for data protection in grids is not only necessary, rather is crucial for the success of grid computing (s. Fig.4) [10-12].

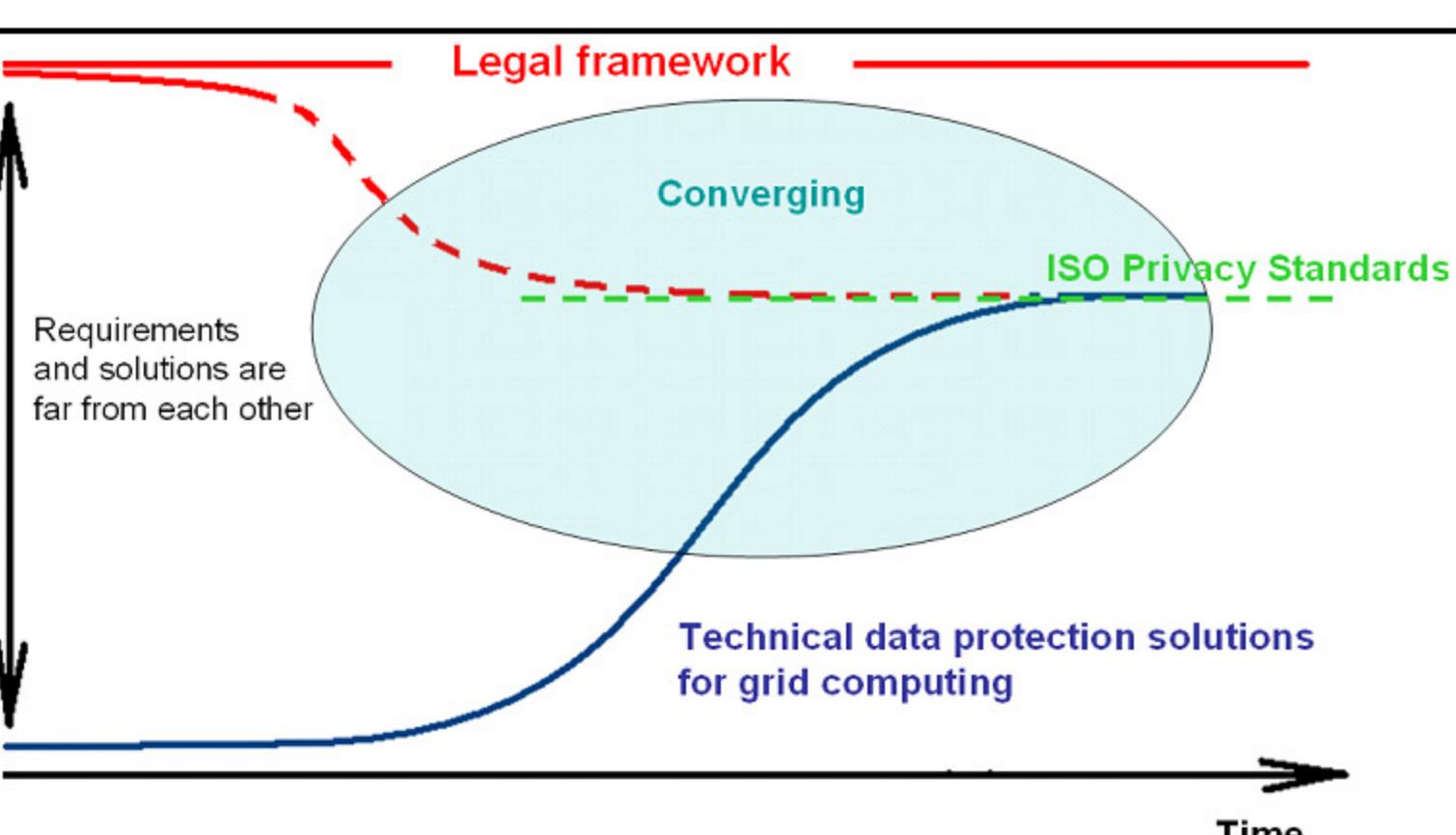


Fig. 4 Converging between the (legal) requirements and the (technical) solutions is necessary. The common level should be the ISO Standards [10,11]

## Results

### Privacy in MediGRID

In the first phase non person-related data is used until the first generation of privacy rules are set up. In the next step – dealing with person related data – additionally necessary advanced security and data protection methods should be defined (Enhanced Security, s. Fig. 5).

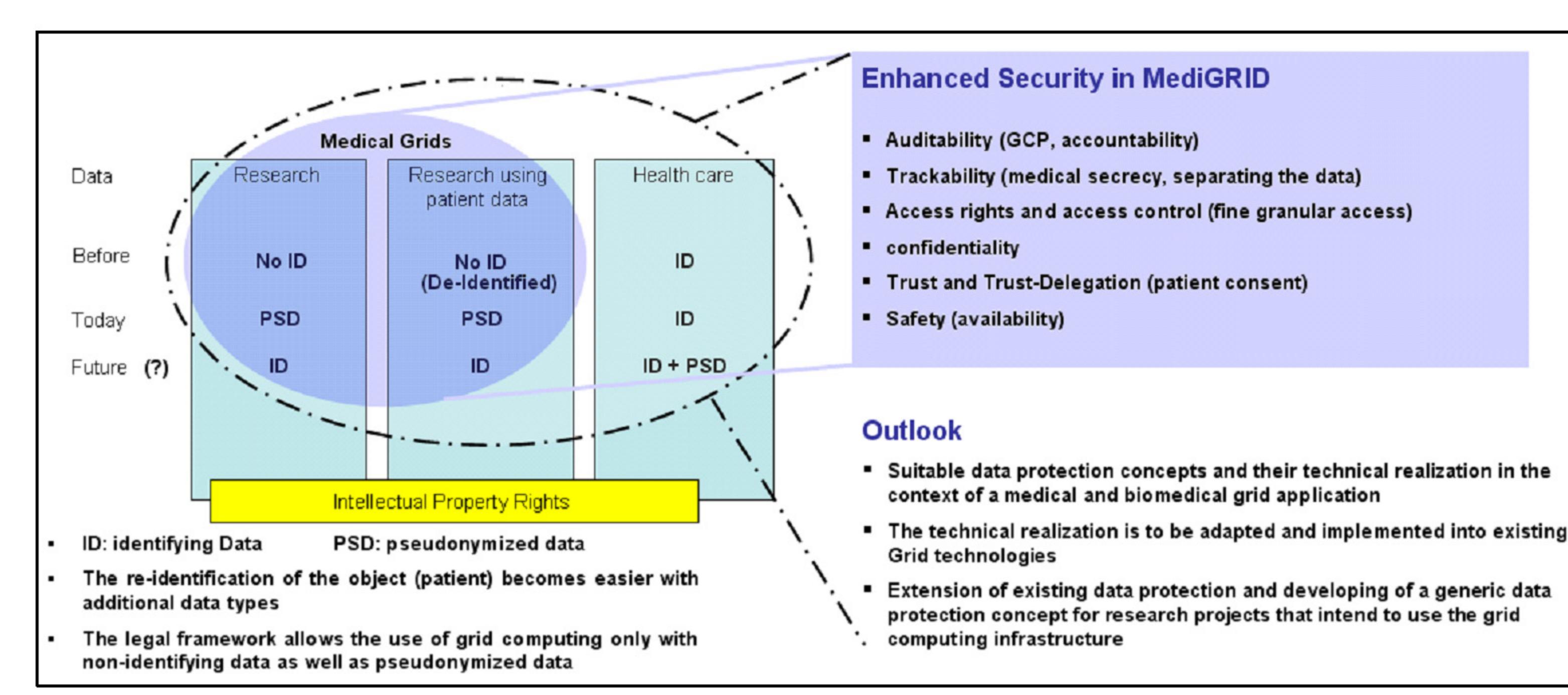


Fig. 5 The limits of using grid computing for medical applications

## Ontology

Using OGSA-DAI for Data Access and Integration in grids, the ontology module has successfully developed and implemented an ontology tool as a GridSphere-Portlet in the MediGRID portal, which is available for all project partners.

## Bioinformatics

Within the MediGRID portal a gridified version of *Dialign* (multiple alignments of nucleic acid and protein sequences) is implemented allowing the user to obtain high-quality alignments of bigger databases and longer sequences.

*Augustus* (s. Fig.6) predicts gene structures in eukaryotic genomic sequences with high accuracy. Users benefit from distributed computing to run several instances of the program aiming at faster computation. Further more grid resources allow regular update huge Expressed Sequence Tag (EST) databases [13-15].

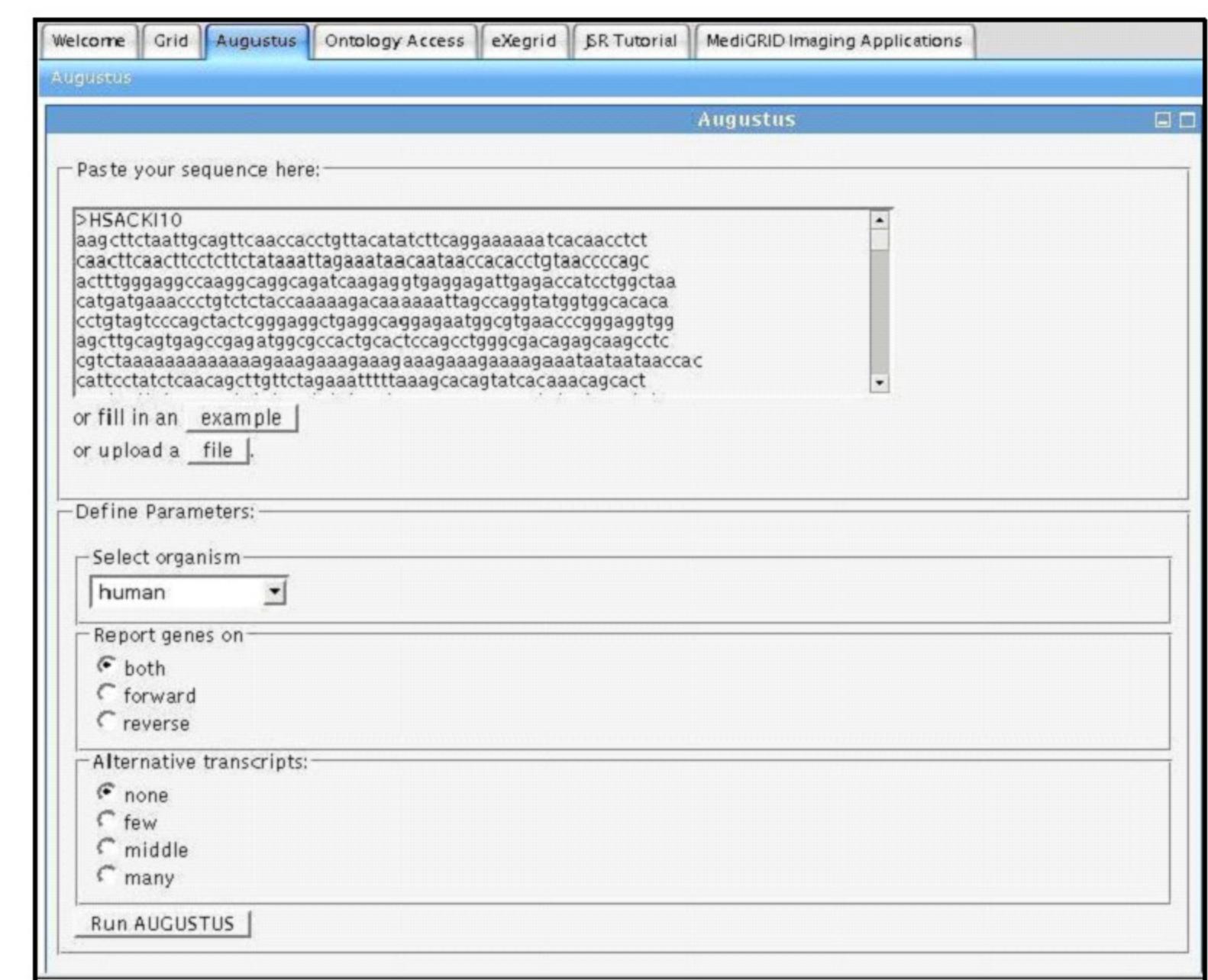


Fig. 6 Augustus running in the MediGRID portal [12-14]

## Image Processing

A 3D image reconstruction for prostate biopsy will register the different ultrasonic scans helping the physician to have a new view of the prostate (s. Fig.7).

The virtual vascular surgery helps to calculate and present the animated 3D blood flow field in the brain vessels, which will be used to anticipate the pressure on the walls of the vessels in order to predict a bleeding risk. The Brain 4D MRI image processing

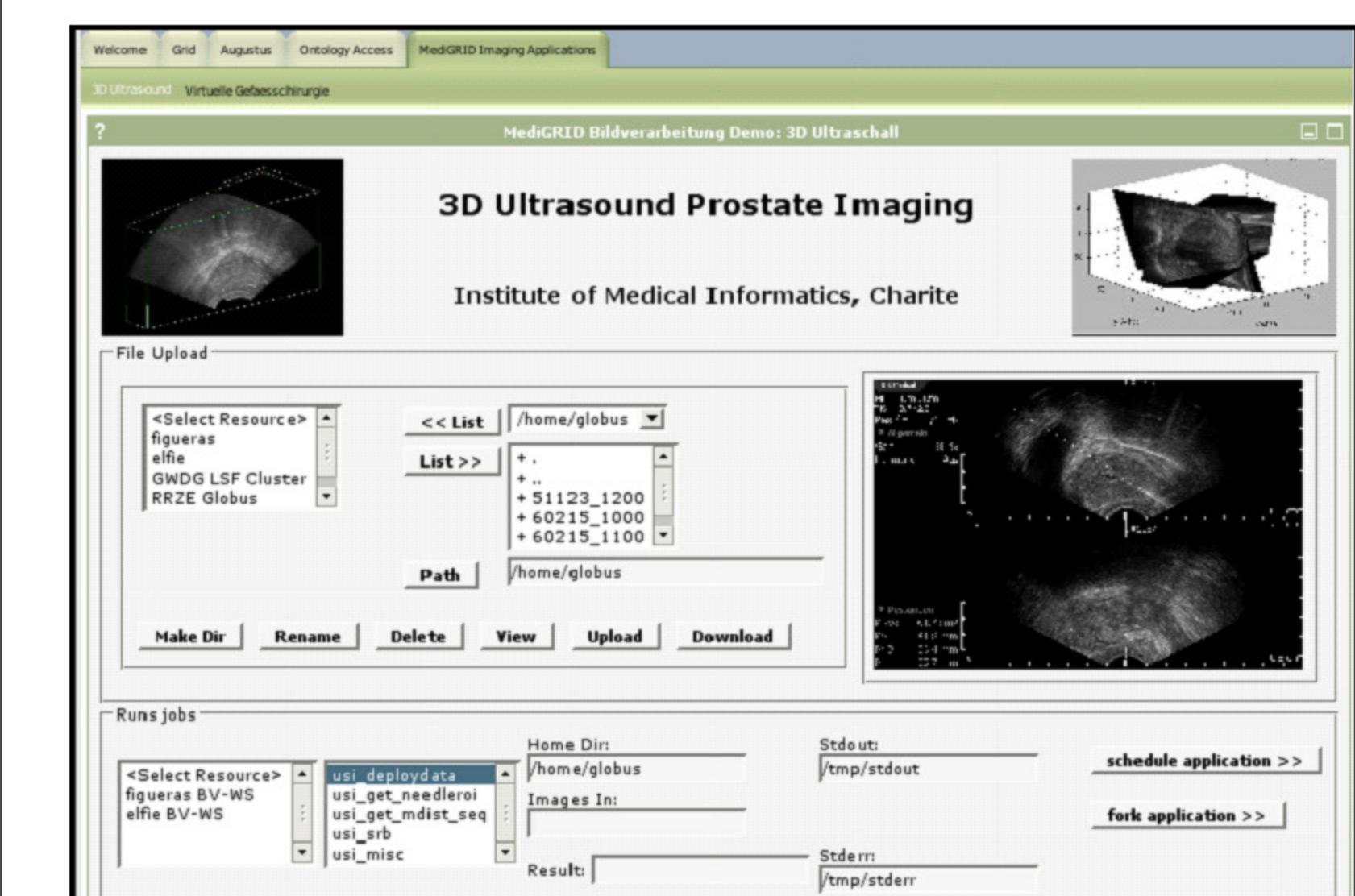


Fig. 7 3D Ultrasound Prostate Imaging running in the MediGRID Portal

application in MediGRID works on the identification of brain areas. Because of the constant increase of the data volume coherent with the development in the imaging techniques, a dynamic extendable computing and storage infrastructure is needed, which ideally will be a grid environment.

## Next steps

Beside defining the security policy for the second phase bearing in mind the different aspects of person related data, standardized privacy and disclosure risk, MediGRID will also gridify and implement new applications in the MediGRID portal and make sure that the infrastructure will be available for other project from the community biomedicine.

## References

- 1. MediGRID. [www.mediagrid.de](http://www.mediagrid.de), 2005.
- 2. M. Mohammadi, J. Viezens, and G. Lopez-Campos, *Integrating genomics into health information systems*, Methods Inf Med, 2002, 41(1): p. 25-30.
- 3. Ulrich Sax, et al., *Project D-Grid: Workshop Security and Privacy*, Georg-August-University Göttingen, Dept. of Medical Informatics, Göttingen, March 23, 2005.
- 4. Lin, Z., A.B. Owen, and R.B. Altman, *Genetics, Genomic research and human subject privacy*, Science, 2004, 305(5681): p. 183.
- 5. D-Grid. [www.d-grid.de](http://www.d-grid.de), 2005.
- 6. Klemann, T. *Standardisation Future Trends and Outlook*, in Gridka School, 2006, Karlsruhe, Germany.
- 7. Butte, A.J. and I.S. Kohane, *Creation and implications of a genome-wide network*, Nat Biotechnol, 2006, 24(1): p. 55-62.
- 8. Kohane, I.S. and Altman, R.B. *Health-Information Altruists – A Potentially Critical Resource*, NEJM, 2005, 353 (19): p. 2074-2077.
- 9. Sax, U. and S. Schmid, *Integration of genomic data in Electronic Health Records – opportunities and dilemmas*, Methods Inf Med, 2005, 44(4): p. 546-550.
- 10. Rienhoff, O., *A legal framework for security in European health care telematics*, Studies in health technology and informatics, v. 74, 2000, Amsterdam ; Washington, DC: IOS Press, x, 192 p.
- 11. Mohammadi, Y., et al., *Rechtliche Aspekte von Grid-Computing in der Medizin*, in *Rechtliche Aspekte der Telemedizin*, W. Niederlag, C. Diers, and H.U. Lemke, Eds., 2005, Springer, Berlin.
- 12. Willenborg, L.C.R.J. and T.d. Waal, *Elements of statistical disclosure control*, 2001, New York: Springer, xv, 261 p.
- 13. Lingner, T. and P. Meinicke, *Remote homology detection based on oligomer distances*, Bioinformatics, 2006, 22(18): p. 2224-231.
- 14. Morgenstern, B., et al., *Multiple sequence alignment with user-defined constraints at GOBICS*, Bioinformatics, 2005, 21(7): p. 1271-3.
- 15. Stanke, M., A. Tzvetkova, and B. Morgenstern, *AUGUSTUS at EGASP: using EST, protein and genomic alignments for improved gene prediction in the human genome*, Genome Biol, 2006, 7 Suppl 1: p. S11 1-8.

This work was supported by the D-Grid Project MediGRID, funded by the Federal Ministry of Education and Research (BMBF), FKZ 01AK03A-H.