

IST-2004-027749

**H-e-C
Health-e-Child**

**D.1.3.a Annual Report
Periodic Management Report**

Instrument Integrated Project

Thematic Priority IST

Reporting period: from January to December 2006

Date of preparation: 15 February 2007

Start date: January 2006

Ending: December 2009

Project Coordinator Joerg Freund
Siemens Medical Solutions

Revision [draft, 4 ...]

PUBLISHABLE EXECUTIVE SUMMARY

The Health-e-Child (HeC) project aims at developing an integrated healthcare platform for European paediatrics, providing seamless integration of traditional and emerging sources of biomedical information. The long-term goal of the project is to provide uninhibited access to universal biomedical knowledge repositories for personalised and preventive healthcare, large-scale information-based biomedical research and training, and informed policy making.

The objectives of the HeC project are: i) to gain a comprehensive view of a child's health by vertically integrating biomedical data, information, and knowledge, that spans the entire spectrum from genetic to clinical to epidemiological; ii) to develop a biomedical information platform, supported by sophisticated and robust search, optimisation, and matching techniques for heterogeneous information, empowered by the Grid; iii) to build enabling tools and services on top of the HeC platform, that will lead to innovative and better healthcare solutions in Europe, namely:

- Integrated disease models exploiting all available information levels;
- Database-guided biomedical decision support systems provisioning novel clinical practices and personalised healthcare for children;
- Large-scale, cross-modality, and longitudinal information fusion and data mining for biomedical knowledge discovery.

The partners involved in HeC are: Siemens AG (Siemens), Lynkeus Srl (Lynkeus), I.R.C.C.S. Giannina Gaslini (IGG), University College London – Great Ormond Street Children's Hospital (UCL), Assistance Publique Hopitaux de Paris – Necker (APHP), European Organisation for Nuclear Research (CERN), University of the West of England (UWE), University of Athens (UoA), Università degli Studi di Genova (DISI), French National Institute for Research on Computer Science and Control (INRIA), European Genetics Foundation (EGF), Aktiaselts Asper Biotech (Asper), Gerolamo Gaslini Foundation (FGG), Maat G Knowledge (MAAT). The Project Coordinator is Dr. Joerg Freund (Siemens) (joerg.freund@siemens.com, tel.+49 (9131)843389, mob. +49 (162) 2923526). The Project Management Team Leader is Prof. Edwin Morley-Fletcher (Lynkeus), the Clinical Coordinator is Prof. Giacomo Pongiglione (IGG), the Technical Coordinator is Dr. Martin Huber (Siemens). The Chairman of the Scientific Committee is Dr. Dorin Comaniciu (Siemens), the Chairman of the Ethical and Legal Review Committee is Prof. Alberto Martini (IGG), the Chairman of the IPR Committee is Prof. Yannis Ioannidis (UoA), the Chairman of the Executive Board is the Project Coordinator, the Chairman of the Governing Board is Dr. Alok Gupta (Siemens).

The HeC project is organised in 16 Work Packages, which have achieved the following results in the first 12 months of activity:

WP1 Project Coordination and Management

Alongside the setting up of the initial project communication infrastructure and of the project management internet Project NetBoard ("PNB") specific training sessions were organised and ad hoc guidelines were produced.

Assistance by email and by telephone was provided in order to help HeC partners to respond to the various reporting requirements of the project, and monthly updating of the PNB with task progress was encouraged.

All reports were delivered on time, except for D 1.1 Quality Assurance Guideline, which was delivered with some delay by agreement with the Project Officer. The timely production of all other deliverables was regularly monitored as well as the fulfilment of all major milestones.

Great efforts were made to integrate clinical and technical involvement, and three Project Management meetings were organised at the premises of the three clinical partners in Genoa, London and Paris during the second quarter.

The day-by-day coordination of all PM activities was mainly operated through weekly telephone conferences, involving first the PC and the PMT leader, and then, after July 2006, extended to the Technical Coordinator and the Clinical Coordinator.

At month 7 it was decided that the PMT (previously restricted to WP leaders) would be extended to a representative from each partners.

A growing number of major institutions signalled their interest in joining the project and the potential advantage of having additional clinical partners was therefore analysed, but the whole issue was scheduled for discussion at the GB meeting in Paris in January 2007.

PMT and Clinical and Technical Review meetings were organised in Munich in July, in Toledo in October, and in Athens in December.

WP 2 User Requirements specification

System views and models for the essential features of the HeC system were developed and models were created to capture the information gathered via the elicitation sessions. The Users Requirements Specifications document was prepared and submitted as project deliverable D2.1 at month 6, as originally planned. A requirements management plan was created which governs the revision process and evolution of the users' requirements specifications. The requirements specifications will be reviewed on a regular basis, the first revision being planned for project months 22-24.

WP 3 Legal, Ethical and Regulatory issues

A rheumatology protocol was created, showing the design of the study, the objectives to aim at, and the kind of information to be collected at each stage. Detailed instructions about clinical assessment, and how to collect and store biological samples, were also prepared as well as case report forms to collect these data.

Data collection protocols for Paediatric Heart Diseases were established with regard to cardiac MRI, echocardiographic investigations, exercise-testing, familial and paraclinical investigations for CMPs and genetic investigation (proper diagnostic algorithms have been prepared).

A Brain Tumour–Gliomas protocol integrating objectives, rationale, eligibility criteria and methods of investigation for all research components: clinical, imaging and genetic studies, was designed. Criteria for access to data bank and informed consent were revisited in relation to the specific project and a specific data base of tumour tissues and blood samples was created.

Study protocols for informed consents of parents and patients were defined, illustrating design, purpose, risks and potential benefits of the project, and also containing information regarding the measures for preventing unauthorised access to medical information and ensuring the privacy of patient data. Biological samples will be stored in a biobank, and two other more informed consents have therefore been prepared.

All the protocols and the related informed consent forms were approved by the Ethical Committees of the three hospitals involved in the HeC project

WP4 Privacy & Security

A Certification Authority (CA) for the development process was created. The testbed for HeC also required a CA for delivering electronic identities useful to servers, services and users' authentication and authorisation, and the web-based management suite OpenCA (<http://www.openca.org>) was selected and customized. Web forms and scripts were reengineered, and new protocols for requesting, revoking and signing certificates developed.

Additional mandatory technological blocks were:

- 1st HeC Virtual Organisation (VO management tools provided by EGEE gLite) service and structure,
- Security Prototype for Authentication in the HeC VO,
- Firewalls Installation and Configuration on the Central Servers.

A java-based prototype client application and corresponding service have been designed and developed. The prototyped solution is cross-platform, makes use of the HeC certificates and uses state-of-the-art technologies from the grid community. A basic VO structure has been configured to allow users to store files within the grid testbed, to use grid middleware resources.

Authentications within the HeC platform now follow a complex and secure unique workflow. The Logging web portal and service facilities for Gateway services were also completed, allowing the efficient storage of logging information and enabling back traceability of every action undertaken in the system.

Additional security mechanisms were coded in the Gateway to complete the existing logic.

WP5 Grid Platform

The EGEE-based prototype was delivered and the HeC test-bed was deployed and configured. Thanks to the test-bed and to gLite training, a good understanding and overview of the middleware was achieved, and the testbed was used, tested and refined, ending up with more than 20 virtual servers, 2 VOs and 3 sub-groups.

Concerns of resource optimisation aspects and virtualisation resulted in the achievement and test of a virtualised grid access point, which can seamlessly be deployed in new sites. The gLite functionality required for the HeC project started to be listed and classified, developing a list of mandatory functionalities to be made programmatically available. The API was shaped and delivered, then tested and completely integrated with the other blocks of the HeC platform.

WP6 Medical Vertical Knowledge Representation

The key initial issues were: organisation of the state-of-the-art review on existing data modelling and integration techniques with respect to medical domain, coordination with the requirements specification WP and synchronisation with the related WP4. Meetings with the clinicians of the rheumatology and cardiology departments focused mainly on the usage of the data in clinical workflows (e.g. data acquisition, patient journey) and subsequently possible modelling directions were identified.

The analysis of the existing biomedical ontologies with respect to their suitability for the project as well as the review of the methodologies for ontology engineering is on-going. Clinical protocols and databases being used at the three hospitals and for the data collection were used as a basis for the modelling of the clinical data. The relevant entities were identified and the UML models for rheumatology and cardiology produced. External ontologies were analysed with respect to the coverage of the HeC domain and the relevant fragments were extracted.

WP7 Data Management Layer & Data Integration Mechanisms

After a preliminary evaluation of the HeC Gateway requirements based on the Users Requirements, a first architectural overview of the HeC platform was produced. A prototype Gateway was built, featuring all available contributions for testing, making it available in the shared source code repository.

A second draft design of the HeC Gateway defined accurately the main platform layers and components in a harmonised way, featuring advanced security and grid functionality. The prototype Gateway was augmented with a database management system exposed as a web service to allow query resolution tests.

WP 8 Medical Query Processing

The concept of Medical Query Processing was introduced as a structured way to handle the complex processing needs of the HeC applications, which include Knowledge Management, Decision Support, and Patient record Access, while vertical data integration, horizontal data integration, data distribution and scalability are the key requirements for the Medical Processing Engine (MPE). The latter will interact with Grid Infrastructure and Database Engines and Ontological Layer Processing. A preliminary architecture of MPE has been defined, including the definitions of the Medical Data Model, and the Medical Query Language. The relevant literature for Distributed Query Processing, Object Matching and Resource Management was surveyed. The initial prototype and associated documentation are being prepared for the deliverables scheduled for Month 18.

WP9 Data Collection

Three objectives were to be pursued: Diagnostic Coding System for Paediatric Heart Diseases, Data collection protocols for the three disease groups, Data collection at the three hospitals at month 12.

A diagnostic coding system ([http://www.aepc.org/aepc/nid/European Paediatric Cardiac Coding](http://www.aepc.org/aepc/nid/European_Paediatric_Cardiac_Coding)) for Paediatric Heart Diseases was selected; CRFs for collecting history and clinical findings were created for each of the three disease groups; data collection protocols were discussed between clinicians of the three hospitals for paediatric heart diseases and inflammatory diseases. Approval of the study protocols by the local Ethical Committees of the three hospitals was delayed, resulting in fewer than expected enrolled patients by the end of the reporting period:

- Inflammatory Diseases: approved on 16th May at IGG, on 25th September at APHP and on 21st December at UCL
- Paediatric Heart Diseases: approved on 29th September at IGG, on 25th September at APHP and on 21st December at UCL
- Brain Tumours: approved on 15th November at IGG.

WP10 Ground Truth and Clinical Knowledge Gathering

For all diseases, imaging data was studied jointly by clinician and IT specialists. For the Inflammatory Diseases study, existing scoring methods used for X-ray and ultrasound images were refined for the project. Wrist and hip MRI scans were jointly studied and it was decided to try to automatically measure the synovial volume. For the Brain Tumours study, the tumour cases stored in the IGG's tissue bank were classified and criteria for evaluation of tumour volume and extension were defined. For the Paediatric Heart Diseases study, echocardiographic features that could be used to classify/evaluate the severity of RVO were proposed and discussed, while MRI examinations were studied jointly at UCL. On the IT side, annotation tools for MR were selected (for Brain Tumours) or further developed (for Inflammatory Diseases and Paediatric Heart Diseases) but due to a lack of patient data (see WP9), only few image data sets were annotated by the end of the 4th quarter.

WP11 Integrated Disease Modelling

A common tool to register model and patient images has been designed for both cardiac diseases and brain tumours. It is now planned to connect it to “registration” web services on the HeC Grid. For cardiology, the WP investigated how to create biventricular models from MR image segmentations, to generate fibre orientations and standard anatomical segments for any anatomy, and to create disease specific models of cardiac anatomy in cases of right RVOs and Cardiomyopathies. First preliminary results were obtained on the measurement of the right ventricular volume from clinical MRIs. On the omics level, the protocols for dealing with RVO data have been designed. In rheumatology, the WP has primarily focused on the quantification of the volume of the inflamed synovia, as it is expected to be an early predictor of the disease severity. Most of the effort has been invested on the assessment of the best MRI modality for this segmentation.

For brain tumours, practical laboratory protocols data have been investigated and implemented for the genetic study and the data-management of affimetrix files has also been worked out (this is valid for the three diseases). On the macroscopic level, the computational model of brain tumours has been modified to describe the motion of the tumour front rather than the whole distribution of tumour cells, and to estimate the tumour invasion margin beyond this visual. This will allow the model to be fitted to patient specific MR data, and to estimate some patient specific parameters for diagnosis and planning.

WP12 Decision Support System

After a first round of elicitation meetings, with focus on user requirements, the structure of the state of the art survey on decision support systems was decided at a combined WP11, WP12 & WP13 workshop. Relevant literature and systems were subsequently reviewed and potential development environments (MeVisLab, Inspace) were evaluated, leading to the completion of the combined State-of-the-Art report by month 6. Various options for developing the DSS applications were evaluated and summarised in the design specification document. A mock up client based on XUL was developed. The decision processes for some of the diseases were analysed at Gaslini.

In the second part of the year, the WP started tool developments for the imaging tools to extract features from cardiac MRI and 2D echo (RV segmentation), hip and wrist MRI (synovial fluid), and brain MRI (brain tumour segmentation), but was hampered by the lack of patient data.

With regards to cardiac MRI, a simplified mesh to assess the localisation of the right ventricle in paediatric MR images was defined and trained. First results for the RV mesh were presented and discussed at Gosh, which will provide some cases with isotropic resolution and larger volumes that incorporate the apex and if possible the pulmonary valve.

With regards to wrist and hip MRI, first attempts started to automatically determine the volume of synovial fluid in wrist MRI, which will be the only automated feature extraction for images. The planned MR sequence had to be changed to achieve better resolution both for the radiologists as well as the segmentation tool.

A first version of a tool to automatically determine the tumour volume was realized. It is based on an open source brain extraction tool, K-means clustering and FGM-EM/HMRF-EM (Zhang et al., 2001) voxel classifier operating on multiple (up to 3)MR input sequences, and was tested with mixed success on 3 MRI data sets from Gaslini. Further improvements of the algorithm are required and/or some manual editing tool to correct for misclassified voxels. The hidden Markov random field based EM algorithm was refined by utilizing gradient and intensity information and the results were compared to a method based on mean field approximation.

WP13 Biomedical Knowledge Discovery

For the first 6 months of the project, all partners were involved in a survey and evaluation of all the promising state-of-the-art data mining and knowledge discovery methods for medical and biological applications. The combined work of WP11, 12 & 13 lead to the State-Of-the-Art (SOA) survey report: a comprehensive 435-page document. After this, the WP began working on the design of the first generation of novel classification and clustering data mining techniques suitable for integrating heterogeneous vertical data. The first results of this work are expected in the next reporting period; however, some preliminary results were presented at the Paris general meeting in January 2007.

WP14 Deployment of the Data Management System & Grid Gateway

The WP first monitored the deployment of a site mainly carried out under the WP5 coordination, in order to follow/support a first real deployment case and so better understand the complexity and logistics needs. In the course of this deployment, the WP14 team started working on the elicitation of a set of documents and tools to support deployment activities and ensure the knowledge related to the grid middleware has been well absorbed.

In the Third Quarter, the team engaged in deployment logistic configuration and its establishment. MAAT agreed to act as a hardware relay for medical centres, and will purchase the necessary servers, batch install them and send them back to the Hospitals for final onsite configuration.

In the Fourth Quarter, the WP14 team was mainly busy at acquiring more knowledge on gLite and maintaining the current infrastructure, and also started to plan the deployment roadmap together with CERN.


WP15 Training

A prototype model for testing advanced training modules was prepared, the training platform requirements have been developed, course topics finalised, and venue features identified. A check list was produced to enable the management of the scientific, technical and organisational aspects of the courses and this involves a requirement for web-casting modality. HeC courses will be offered to EGF Remote Training Centres (RTC) Network.

The first course on “Genetics of rheumatic and auto-inflammatory diseases” has been planned and will take place from 1 to 4 April 2007..

The opening of the PNB included training sessions for each partner - during a meeting at the CERN in May, and then to various partners by telephone and in person. In addition specific guidelines were produced by the PMT.

WP16 Dissemination

The HeC logo  was chosen and a website (www.health-e-child.org) was created and developed. The website contains a public area for the dissemination of results, and a private area devoted to the exchange of information between the members of the consortium.

Initial information about the project was posted on the public side of the site as well as information about forthcoming events and conferences of general interest. On the private side, most of the active partners started to share information about their progress, and all Deliverables and Minutes of Meetings were regularly uploaded.

The Dissemination Strategy Plan was delivered at month 6. The plan for the quarterly newsletter was finalised and scheduled for publication at the end of the 4th quarter. The first issue of HeC Newsletter was circulated among all Consortium partners in November and uploaded on HeC website in December.