Data Integration in eHealth: A Domain/Disease Specific Roadmap

Jenny URE[a b 1], Rob PROCTER[l], Maryann MARTONE[d], David PORTEOUS[a], Sharon LLOYD[b] , Stephen LAWRIE[b]., Dominic JOB[b], , Richard BALDOCK[g], Alistair PHILP [a], Dave LIEWALD [a], Frank RAKEBRANDT[b], Alan BLAIKIE [a], Clare McKAY[b], Stuart ANDERSON, John AINSWORTH[c], Jano van HEMERT[k], Ignacio BLANQUER[e], Richard SINNOTT [a], Christian BARILLOT, Frank Bernard GIBAUD, Alan WILLIAMS[c], Mark HARTSWOOD, Paul WATSON[m], Leslie SMITH[m], Albert BURGER, Jessie KENNEDY[k], Horacio GONZALEZ-VELEZ[n], Robert STEVENS[o], Oscar CORCHO[i], Robin MORTON [a], Pamela LINKSTED [a], Mylene DESCHENES[h], Mark McGILCHRIST [a], Paul JOHNSON [a], Alex VOSS[l], Renate GERTZ [a], Joanna WARDLAW[b].

[a]Generation Scotland, [b]NeuroGrid, [c]PsyGrid, [d]BIRN [e]HealthGrid Share, [f]NeuroBase, [g]MouseBIRN, [h]P3G Consortium, , [i]OntoGrid [j]CARO, [k]NESC, [l]NCESS, [m]CARMEN, [n]HealthAgents [o]Sealife

Abstract. The paper documents a series of data integration workshops held in 2006 at the UK National e-Science Centre, summarizing a range of the problem/solution scenarios in multi-site and multi-scale data integration with six HealthGrid projects using schizophrenia as a domain-specific test case. It outlines emerging strategies, recommendations and objectives for collaboration on shared ontology-building and harmonization of data for multi-site trials in this domain.

Keywords. Data integration, e-Health, psychosis, multi-scale imaging, ontology integration

1. Introduction

Grid technology has a key role in enabling the development of a European Research Area [1, 2] with the potential to allow querying across heterogeneous and distributed data sets if these can be integrated and represented in ways which are valid, usable and ethically and legally acceptable [3, 4]. In areas such as brain imaging, the opportunities and the challenges of integration have been particularly evident, requiring integration in multi-centre clinical studies of patients in early stages of psychiatric disorders, standardization of scanners and image processing techniques across mental health research networks as well as scalable integration of voxel-based image data at different levels of integration [5, 6], and the development of shared ontologies and spatial

[1] Corresponding Author: Jenny Ure, School of Informatics, University of Edinburgh, Jenny.Ure@ed.ac.uk

frameworks for reporting brain-related data, both for comparison across sites and to build up integrated views of the brain. The paper summarises a range of the problem/solution scenarios in multi-site data integration and multi-scale datasets from a series of eHealth workshops held in 2006 at the UK National e-Science Centre. The first workshop[2] looked at generic issues, and the second[3] brought together six HealthGrid projects[4] in the same disease domain to road map the issues in a single disease domain, using schizophrenia as a testbed. These include discussion of ontology. integration issues and a roadmap of short and medium term objectives for joint working This was co-hosted by the National e-Science Centre and the Generation Scotland national population genomics project[5] (a nationally funded collaboration of the Scottish University medical schools, the NHS and key research institutes), [7] together with the NeuroGrid[6] project [8] funded by the MRC to develop a Grid platform and toolkit for sharing imaging data for research on dementia, stroke and psychosis. This also builds on the ongoing road-mapping process to support collaboration in data sharing in particular disease domains initiated in the HealthGrid Share project[7]. One of the outcomes of the workshop is a wiki-based collaboration[8] between six HealthGrids from the UK, EU and US to develop shared measures of symptom in terms of imaging, genetic, clinical datasets, shared metadata and shared ontological representations of these.

2. Multi-site and multi-scale integration

The eHealth vision of large-scale, seamless data-sharing for research has to be tempered by acknowledgement of the very real barriers standing in the way of its realisation. The data 'supply chain' underpinning the concept of eHealth and translational medicine is a gradual conversion process where many types of error or bias can arise at different stages from sampling, collection, coding, aggregation, analysis or use, sometimes referred to as the 'social life of information' [9].

2.1. Differences in populations

Many of the datasets were drawn from regional or national studies on very different populations, raising the risk of confounding artifact with effect, whether across sites, or across projects. Data sets may be aggregated from very different populations, where there are physiological differences in brain shape and size that reflect ethnic differences rather than disease effects. In neuroscience, for example, in controlling for global brain volumes in high-risk subjects with and without psychotic symptoms, some measures (e.g. using height as a proxy for head size) and use of paternal social class as a proxy for environment may have variable effects across ethnic and cultural groups [5]. Indicating ethnicity as part of the core metadata would go some way towards ameliorating this, however the concept itself is hard to define, and obtaining this

[2] http://www.nesc.ac.uk/action/esi/contribution.cfm?Title=709
[3] http://www.nesc.ac.uk/action/esi/contribution.cfm?Title=684
[4] http://wikis.nesc.ac.uk/mod/Main_Page
[5] www.generationscotland.org
[6] www.neurogrid.ac.uk
[7] http://www.eu-share.org/
[8] http://wikis.nesc.ac.uk/mod/Main_Page

information is fraught with sensitivities, to the extent that some studies do not attempt to document this.

Participants from the consortium of national population genomics projects[9] also pointed out that recruitment, despite best efforts, is rarely truly representative. Participants are essentially volunteers, more often women than men, and many do so as a result of experience of particular illnesses in families. Recruitment itself is also necessarily opportunistic, and reflects the city and often the hospital setting where the clinical tests are carried out. This is also an issue with control groups in other studies, with some evidence of bias effects in small studies [10]. Epidemiologists here highlighted the importance of separating out differential errors that are likely to cancel out with large samples and non differential errors which are accentuated, inducing bias or confounding effects, and may not be easily identified.

Scanned images are assessed by statistical analysis of brain volumes or densities after registration and normalization, using normal scans as a benchmark. HealthGrids using imaging increasingly use sets of human 'phantoms' scanned at all the sites to provide a standard benchmark (although there are also variations in scans between the same individual at different times). In multi-site studies across more than one Grid project this would require further harmonization.

2.2. Differences in Collection, Coding and Collation

Known error rates of 30% were not uncommon when matching test data against patient data. Many of the problems related to the quality of the data originally filled in:

- Missing data
- Incomplete data
- Incorrect data, e.g., the patient's name being entered as "brain".
- Incorrectly formatted data, e.g., a patient name being specified so that the surname is "SmithJohn".
- Data in the wrong field
- Data in the wrong sequence
- Inconsistent data within a single file, e.g. If the patient's age is inconsistent with image date minus birth date.
- Inconsistent data for same patient on different visits e.g. are patients with the same issuer and patient ID but different names really the same patient?

Cleaning and error trapping software can only capture particular types of error and many anomalies would only be recognized by those with local knowledge of the population, the context and the method of data collection. For example, protocols were interpreted in different ways, or local events impacted on the administration of the test. In aggregated data, without this information, anomalies are hard to identify. Strategies for addressing this included wireless notepads or pens used at the data upload stage, so that data incorrectly loaded was automatically validated against the main database as it was stored. As indicated earlier, error trapping software was also used, together with metrics using probabilistic linkage. It was also seen as important to keep links to raw data (or data owners) where possible. Metadata on provenance was seen as very

[9] The P3G consortium www.p3gconsortium.org is working with over 11 Biobanks on harmonisation

important here [11], however some anomalies in data were often only evident to those familiar with the context, and prepared to follow these up on the ground. One study cited measures of heart rate captured in one site which were consistently higher in one site, and where this was consonant with known population differences. Data quality harmonization work between sites in the study highlighted the fact that participants in one site had to climb six flights of stairs before the ('resting') heart rate test while the lift in the hospital was out of order. This would not be evident from inspection of the data alone but required local input. Guidelines, checklists and toolkits were also used for enhancing data quality. The same communities who produce data were seen by many as particularly well placed to evaluate, quality assure and enhance it. In this context, one study proposed to have a panel responsible for ethics, linkage and data quality issues mediating requests for data as well as submission of data, building on their ability to interpret and also act on local processes for data collection [13], and in the extended healthcare team in primary care [14]. Other strategies included data audit tools, and process management tools to make the stages of the data collection and analysis process transparent to allow researchers and users to compare methods and see the strengths and weaknesses of the data [12].

2.3. Differences in Tests and Tools

In the context of schizophrenia, there were a range of tools, techniques and formats for capturing aspects of the same structure or process, making it hard to differentiate between real differences and artefacts. For example, structural magnetic resonance imaging (sMRI) was used by a majority of the participants researching reduced volumes of the medial temporal lobe and other limbic and paralimbic structures in schizophrenia [5]. There was a then a need for harmonization between different types of scanner, differences in the processes of registration or normalization, differences in the settings (even after servicing) etc. Similarly, in genetic analyses, results can be obtained by a range of different methods, such as microarray, in situ hybridization, and immunocytochemistry , thus raising the possibility that differences between datasets may be a function of the testing and analysis process itself, and this must be included in provenance metadata.

A range of strategies were evident here, including ongoing harmonisation/data quality testing between sites, use of shared 'phantoms' as controls and use of common tools. The BIRN project has been active in using all these approaches across US sites, and with collaborators in the UK such as PsyGrid and NeuroGrid, as well as with NeuroBase in France. In a number of other projects, early prototyping provided a vehicle for community engagement [8], and a number of harmonisation studies were underway within and between participating Grid projects. The use of shared tools as freeware is increasingly a strategy, with BIRN again providing a range of these to support (initially) their own work across multiple sites in the use, but increasingly also with collaborating nodes at eScience centres in the UK and EU. For many of the projects, the range of preferred local software and tests was not only an issue in mapping differences, but the fact that some of these are licensed, commercial or IP protected software, making their provision as tools for other unlicensed users particularly complex.

2.4. Differences in Requirements for Confidentiality of Patient Data

In the case of scans of patients at risk of early-onset psychosis in one study, direct access to the imaging data was regarded as too sensitive and the solution agreed was to provide access to derived statistical data on which algorithms could be run. This added some complexity to the workflows and the design as a whole, but aligned the competing requirements of the different stake-holding groups in a way which could be replicated elsewhere. Ethical and legal issues were seen as unresolved problem issues, reflecting a patchwork of disjoint technical ethical legal and administrative domains. The break-out session on ethical and legal issues highlighted the fact that current legal frameworks cannot provide clear answers for emerging new scenarios, and project teams were increasingly aware of the risks of legal challenge, and of delays in recruitment or use of data for ethical reasons or as a function of public perception. Collaborative stakeholder negotiation was increasingly a basis for agreeing a reasoned, and enforceable position within applicable legal and ethical frameworks, and some national initiatives have include the development of tool kits to support coherent approaches across extended health care communities.

2.5. Differences in Semantic, Ontological and Spatial Integration

Neuroscientists now have access to a vast array of large, heterogeneous and multi-dimensional data from multiple sources, and across multiple scales. As observers increasingly point out, the challenges are now more about integrating data and information, making sense of it (in machine and human terms) and representing it in ways that relate to (and evolve through) the aims and frames of reference of different user groups [3, 15, 6]. Integrating heterogeneous and distributed datasets is therefore a challenge for the e-Health and the e-Science vision, and a priority area for regional, national and international bodies supporting research in e-Health and e-Science.

Data at molecular level on synaptic proteins involved in human mental illness, such as schizophrenia, bipolar disorder and mental retardation [16] is even more valuable when integrated with scanning data, and genetic data yet this requires coordination in a spatial/anatomical frame of reference, using a shared data model, and ideally a human and machine readable format. Much as existing pieces in a jigsaw can support new insights about the structure as a whole, and the missing parts of it, the aggregation of disparate information within a shared model can support both interoperability and understanding if there is sufficient opportunity and motivation for joint working [17].

The workshop grew in part from the awareness that six Grid projects were developing different OWL[10] based ontologies to facilitate cross searching and knowledge discovery across multiple data sources, at different anatomical scales. Ontology integration approaches handle multiple different ontologies by identifying mappings between heterogeneous ontologies or by merging them into a single ontology [18], however this is both complex and variable in the results it achieves, given the semantic heterogeneity, and the variable perceptions of ontological relations across groups, and over time.

The approach adopted in the BIRN project has been to facilitate collective development of the underlying semantics, and conceptual relationship from which project ontologies

[10] The Semantic Web comprises the standards and tools of XML, XML Schema, RDF, RDF Schema and OWL. The OWL Web Ontology Language Overview describes the function and relationship of each of these

are then constructed using open source tools such as the BirnLex[11] and Firework Concept Browser tools. (The BIRNLex is a controlled vocabulary including common terms for neuroanatomy, molecular species, subject information, behavioral and cognitive processes, experimental practice and design, and the associated elements of primary data provenance required for large-scale data integration across disparate experimental studies. It also provides a core for the re-use and integration of existing community ontologies - e.g. OBI, CARO, BFO, and GO and some division of work between groups.). A variety of techniques producing data at different scales can now be superimposed onto the neuronal networks to create new models of the human brain [6, 17] in ways which are human and machine readable, and represented in ways which are visually intuitive. As De Roure [19] points out sense-making is increasingly the issue. Sharing anatomical correlates provides a basis for scalable spatial mapping and integration[12] that facilitates both human interpretation and also more extensive data mining and knowledge discovery.

This Open Source approach, sharing tools and resources, has been increasingly seen as a means of adding value, cutting costs, benchmarking approaches and sharing risk towards common ends in the development of sociotechnical systems [20,21]. The aim here has been to develop a dynamic knowledge infrastructure to support integration and analysis, and to identify and assess existing ontologies and terminologies for summarizing, comparing, merging, and mining datasets that include clinical assessments, assays, demographics, cognitive task descriptions, neuroanatomy, imaging parameters/data provenance in general, and derived magnetic resonance imaging data.

Although less of a short term problem, there was also a perceived tension between the benefits of a fixed frame of reference as implied by an onotology, and the fluid nature of knowledge emerging from ongoing research. Ontologies are a tangible model of domain knowledge involving fairly persistent logical and conceptual relations between classes. This is in tension with the accepted model of knowledge implicit in research on an evolving, hypothetical set of relations which will change, and may involve many parallel disputed interpretations at any one time. In business systems, [22, 23] the approach to this problem has been to separate out the core areas that can most easily be standardized, and to allow a range of approaches to evolve ‘at the edge’ in an evolutionary manner.

2.6. Differences in Diagnosis and Treatment

Another significant difference between projects was in the diagnosis of complex diseases such as the psychoses. Not only are there different measures of symptoms, in relation to the data sets held, the same symptoms can be associated with different formal diagnoses and treatment recommendations. A key task recommended by the collaborating group from the workshop was to achieve agreement on measures of symptom severity.

11 http://www.nbirn.net/birnlex/index.htm

12 Similar principles are used in Google Earth, and in web-based self-organising maps (WebSOMS) facilitating sense-making by building on the behaviour of visual and cognitive systems – scanning, zooming, and organising by semantic or physical similarity for example).

3. Recommendations and Outcomes

The workshop provided an opportunity for a collective focus on shared problems and possible strategies with regard to data-sharing. There were a number of general recommendations and specific recommendations for collaborative working on a shared task.

3.1. Collaborative ontology development – building on work of the BIRN Human Brain project.

The six HealthGrid projects have a number of common aims in integrating a wide range of datasets in the domain, for particular purposes and the potential therefore to benefit from collaborative working. This was identified as a task that could be jointly pursued through joint Access Grid meetings and a project wiki in the first instance, with secure access to the BIRN ontology effort, developing including common terms for neuroanatomy, molecular species, subject information, behavioral and cognitive processes, experimental practice and design, and the associated elements of primary data provenance required for large-scale data integration across disparate experimental studies. Given small local sample sizes, and a collective interest in large-scale multi-site clinical trials, this was seen as the best approach for a range of reasons.

3.2. Agree on measures of symptoms in relation to the data sets held

Another recommendation was to work towards shared measures of symptoms and symptom severity in datasets. A wiki has been set up for this on http://wikis.nesc.ac.uk/mod/Main_Page , and the Access Grid nodes at the eScience centre will also allow for collaboration in joint development meetings. This complements existing UK and EU collaborations with BIRN supporting the use and evaluation of BIRN applications in the UK eScience community, and the development of ontologies.[13]

3.3. Share and re-use tools and strategies

There were seen to be advantages in re-use of existing work where possible and finding synergies with the work of other groups towards common ends. In harmonisation of the scanning process for example, strategies for accommodating or harmonising inter site differences are important in disentangling disease effects from other factors. Individual projects had invested time and effort in a range of measures that were often duplicated. Sharing techniques and software development effort for general adoption was therefore perceived as offering multiple benefits. Re-use was also seen as possible from other distributed networked systems in e-business and e-learning where similar problem-solution scenarios have been addressed in transferable ways, in the business supply chain for example [22]. The wiki and the Access Grid provide a medium for developing this, as well as opportunities provided by future workshops.

[13] http://www.nesc.ac.uk/action/esi/contribution.cfm?Title=706

3.4. Coordinated Approach to the Development of Spatio-temporal Ontologies

A shared interest for many of the participants was the potential to explore approaches to the development of structural mappings of objects and relations in space and also over time. The diagnosis development and prognosis of disease, and the evaluation of treatment regimes were perceived as potentially crucial in generating benefits for patients. Work from the EMAP/EMAGE[14]and DGEMap[15] projects demonstrated at the workshops and at a prior UK-BIRN workshop[16] were perceived as areas of real interest for future development.

3.5. Data Quality and Situated Local Action

Although not a recommendation, there was a perception from the discussions in both workshops that distributed communities had a particular role in enhancing the quality of integrated resources, and that this is likely to be a key factor in usability [24] and also sustainability. The pattern in more established distributed business contexts has been to move towards models that allow greater leverage of the situated knowledge and agency of local communities to greater advantage [23] and this may also be the case in the context of eHealth.

3.6. Building Technology Around the Cognitive and the Social Process

Much of the discussion has been on the alignment and harmonization of distributed data, but an emerging area of interest has also been the representation of distributed data in spatial and temporal contexts that leverage existing intuitive cognitive and visual architectures to the analysis of complex data (semantic clustering of similar information, layering etc). This has now become a major factor in the usability of the semantic web on the scale implied by HealthGrids. Large-scale initiatives such as BIRN, GIS systems (think Google Earth[17]), and the work of Berriman et al in astronomy [25] highlight the potential of approaches that actively leverage cognitive and also social processes in the design of usable eScience systems.

4. Conclusions

The multiple Grid projects participating at the event[18] were all at different stages of developing data sets of the same type. Collaborating development on this scale can create de facto standards, share cost and risk, and support proposals for changes in the frameworks adopted at a national and international level [26]. Grid-based eHealth projects imply data-sharing across stake-holding groups and governance frameworks. This is a process which requires formal and informal opportunities for collaboration to take place, and perceived rewards for so doing. In the UK the e-Science and e-Social Science centres have been instrumental in providing support for this through the Data

[14] http://genex.hgu.mrc.ac.uk
[15] http://www.dgemap.com/
[16] http://www.nesc.ac.uk/action/esi/contribution.cfm?Title=706
[17] http://earth.google.com/#utm_campaign=en
[18] Appendix of details on the collaborating Grid projects is on http://wikis.nesc.ac.uk/_mod/Main_Page

Integration Theme[19] and agenda setting workshops [27] as well as supporting wiki-based collaboration such as this and joint Access grid meetings. There is, however, a balance to be struck between the benefits of shared standards and ontological mappings on the one hand, and the risks of limiting the diversity of knowledge deriving from multiple models and local variants. e-Business approaches to this same problem increasingly separate core and local in ways which can provide useful synergies, while new approaches in informatics such as agent based decision support systems [28] may offer other less constraining models for data integration in some contexts.

5. Acknowledgements

We would like to acknowledge the support of Generation Scotland, NeuroGrid and the UK eScience Centre for funding the workshop and NeuroGrid and Generation Scotland for funding the research. Generation Scotland is funded by Scottish Funding Council Grant No. SRDG03006 NeuroGrid is funded through MRC Grant Ref no: GO600623 ID number 77729.

References

[1] Breton, V., Dean, K. and Solmonides, T. (2005). The HealthGrid White paper, In *Solomonides, T., McClatchy R., Breton V., Legre, Y. & Norager, S. (eds.) From Grid to HealthGrid*. IOS Press ISSN 0926-9630.

[2] Blanquer I., Hernandez V. and Mas F., 2005 A Peer-to-Peer Environment to Share Medical Images and Diagnoses Providing Context-Based Searching, *Proceedings of the 13th Euromicro Conference on Parallel, Distributed and Network-Based Processing (PDP'05)* - Volume 00 pp 42-48

[3] Goble C.A., Corcho O., Alper P., De Roure D. 2006)
e-Science and the Semantic Web: a Symbiotic Relationship, Discovery Science 2006, Barcelona, Spain, Springer-Verlag, Lecture Notes in Artificial Intelligence (LNAI) 4265

[4] Sinnott R.,(2006) Virtual Organisations for Trials and Epidemiological Studies (VOTES) – Experiences & Prototypes after 1 year, In *Re-Use or Reinvention Workshop*, Nat. eScience Centre, Edinburgh

[5] Lawrie SM, Johnstsone EC, Weinberger DR. (editors) (2004) Schizophrenia: From neuroimaging to neuroscience. Oxford University Press, Oxford.

[6] Martone M., (2004), e-Neuroscience: challenges and triumphs in integrating distributed data from molecules to brains, In *Nature Neuroscience Vol. 7, No.5*, May 2004.

[7] Smith, B. Campbell, H., Blackwood D., Connell J., Connor M., Deary I., Dominiczak A., Fitzpatrick B., Ford I, Jackson C., Haddow G., Kerr S, Lindsay R., Mcgilchrist M., Morton R., Murray G., Palmer C., Pell J., Ralston S., St Clair D., Sullivan F., Watt G., Wolf R., Wright A, Porteous D., Morris D., (2006). Generation Scotland: the Scottish Family Health Study; a new resource for researching genes and heritability. BMC Med Genet, pp.74.

[8] Geddes et al (2006) NeuroGrid: Using Grid Technology to Advance Neuroscience, In *18th IEEE Symposium on Computer-Based Medical Systems (CBMS'05)*

[9] Duguid, P. & Brown, J.S. (2000). *The Social Life of Information*, Boston, MA: Harvard Business School Press.

[10] Sterne J.A., Gavahan D., and Egger M., (2000), Publication and related bias in meta-analysis. Power of statistical tests and prevalence in the literature. *Journal of Clinical Epidemiology, Vol.53, Issue 11, Nov. 2000, pp. 1119-1129*

[11] Zhao J. Goble C., Stevens R., Jun Zhao, Carole Goble (2006), An identity crisis in the life sciences. In *Proc. of the 3rd International Provnenace and Annotation Workshop*, Chicago, USA, May 2006. LNCS. extended paper.

[19] Workshop presentations (1) http://www.nesc.ac.uk/action/esi/contribution.cfm?Title=684
(2) http://www.nesc.ac.uk/esi/events/709/

[12] Van Vlymen J., De Lusignan S., Hague N. Chan T., Dzregah B. (2005), Ensuring the Quality of Aggregated General Practice Data: Lessons from the Primary Care Data Quality Programme (PCDQ).
[13] McGilchrist (forthcoming)
[14] Smith B. H., Watt G.C.M., Campbell H., and Sheikh A. Genetic epidemiology and primary care, *Brit. Journal of General Practice*, March 2006.
[15] Rector A.L. Rogers J.E. (2006)Ontological issues in using a description logic to represent medical concepts: Experience from GALEN , In *Methods of Information in Medicine* , Springer Verlag, Jan. 2000
[16] Armstrong JD, Pocklington AJ, Cumiskey MA, Grant SGN. (2006) Reconstructing protein complexes: from proteomics to systems biology. *Proteomics* 6, 4724 - 4731.
[17] Nonaka, I. and Nishiguchi, T. (eds.) (2001). Knowledge Emergence: Social technical and Evolutionary Dimensions of Knowledge Creation, Oxford University Press, Oxford.
[18] Abels S., and Haak L. and Hahn A., (2005) Identification of Common Methods use for Ontology Integration Task, In Interoperability Of Heterogeneous Information Systems, *Proceedings of the first international workshop on Interoperability of heterogeneous information systems*, (2005) Bremen.
[19] De Roure D., Jennings N., and Shadbolt N., (2003) *Research Agenda for the Semantic Grid, In Grid Computing: Making the Global Infrastructure a Reality*, (Eds) Berman F., Hey AA.J.G. and Fox, G., John Wiley and Sons pp 437-470 2001 to 2003.
[20] Shneiderman B.(1998) Codex, memex, genex: The pursuit of transformational technologies. CHI98, Plenary Address. *Conference on Human Factors in Computing Systems*, Los Angeles, CA, April 1998.
[21] Kling R., McKim G., and King A. (2003) A Bit More To IT: Scholarly Communication Forums as Socio-Technical Interaction Networks. In *Journal of the American Society for Information Science and Technology 54(1)*, 47-67.
[22] Ure J. & Jaegersberg G., (2005) Invisible Architecture,: the benfits of aligning people, processes and technology- case studies for system designers an managers. BCS, ISBN 1-90250-5-59-X
[23] Sawhney, M. & Parikh, D. (2001). Where Value Lives in a Networked World, *Harvard Business Review*, January, pp175-198.
[24] Procter R., Borgmann C., Bowker G., Jirotka M., Olson G., Pancake C., Rodden T., and Schraefel M.C., (2006) Usability research challenges for cyberinfrastructure and tools, In *Conference on Human Factors in Computing Systems CHI '06 extended abstracts on Human factors in computing systems* pp: 1675 - 1678 ISBN:1-59593-298-4
[25] Berriman G.B., Deelman E., Good J., Jacob J. C. Katz D.S. Laity A.C., Prince T.A. Singh G., Su M. (2007) Generating Complex Astronomy Workflows. In Taylor, I.J., Deelman, E., Gannon, D.B. and Shields, M. (eds.) *Workflows for e-Science: Scientific Workflows for Grids.* Springer, 2007. pp. 19-38
[26] Wilson P., and Lessens V., (2006) Rising to the Challenge of e-Health Across Europe's Regions. *In eHealth2006*, May, 2006, Malaga
[27] Lin Y. Procter R., Randall D., Rooksby J., Sharrock W., Ure J., Voss A. (2007) Issues in Ontology Development and Use. (forthcoming)
[28] Arús, C., Celda, B., Dasmahaptra, S., Dupplaw, D., González-Vélez, H., van Huffel, S., Lewis, P., Lluch i Ariet, M., Mier, M., Peet, A., and Robles, M. On the design of a web-based decision support system for brain tumour diagnosis using distributed agents. In *WI-IATW'06: 2006 IEEE/WIC/ACM Int. Conf on Web Intelligence & Intelligent Agent Technology (Hong Kong, Dec. 2006), IEEE*, pp. 208-211