

Call for Participation in a mini-workshop at AHM2004 on "e-Science and Data Mining" to launch the e-Science Data Mining Special Interest Group

<http://www.allhands.org.uk>

Introduction: e-Science and Data Mining

One of the principal motivations for the e-science programme is the data avalanche being experienced in many scientific disciplines. This will change the way that science is done, as many conventional analysis techniques will not scale to the exponentially increasing data volumes found in many scientific domains, nor do they typically take advantage of the multitude of distributed data sources that may be relevant to a given science problem. The data mining and visualization communities have made much progress in recent years in developing techniques for summarising, and extracting salient features from, large, multi-dimensional datasets, and these may greatly help the exploration of data within e-science.

The e-Science Data Mining Special Interest Group (*esdm-sig*)

In October 2002, NeSC hosted a workshop on *Scientific Data Mining, Integration and Visualization* (SDMIV[1]), which brought together fifty people, ranging from software engineers developing Grid infrastructure software, to computer scientists with expertise in data mining and visualization, to applications specialists from a wide range of disciplines. The SDMIV workshop identified a set of common problems related to data mining facing the e-science community, but few potential solutions. A year on, the situation is different, as a number of projects (notably DAME[2] and DiscoveryNet[3]) have data mining systems operating in an e-science environment, and the e-Science Data Mining SIG[4] is being set up to advance the interaction between the data mining and e-science communities and address the general topic of data mining for e-science; a related group (*evis-sig*[5]) fulfils the same role for visualization.

The *esdm-sig* mini-workshop at AHM2004

The first task of the *esdm-sig* is to conduct a thorough study of the data mining requirements and expertise within the e-science community, and the implications this has for the further development of e-science middleware. This study is starting now, and details of its format and some of its preliminary results will be presented at the AHM2004 mini-workshop. The remainder of the workshop will comprise four short (15min) talks addressing the following three themes:

- Science drivers for data mining in e-science
- Technical issues for data mining in e-science: scalability, distributed data sources, heterogeneity, etc
- Examples of the state of the art in e-science data mining.

Contributions are invited for talks and posters on any of these topics, and it is intended that the mini-workshop will be followed by an informal discussion session for those wishing to take an active role in the *esdm-sig*. Abstracts for the mini-workshop should be submitted via the AHM2004 WWW site [add URL here], while further information on the mini-workshop and on the *esdm-sig* can be obtained from the SIG convenor, Bob Mann (rgm@roe.ac.uk).

References

1. Mann R.G., Williams R., Atkinson M.P., Brodlied K.W., Storkey A.J., Williams C.K.I., *Scientific Data Mining, Integration and Visualization*, UKeS-2002-06: <http://www.nesc.ac.uk/talks/sdmiv/report.pdf>
2. Distributed Aircraft Maintenance Environment (DAME): <http://www.cs.york.ac.uk/dame/>
3. Discovery Net: An e-Science test Bed for High Throughput Informatics: <http://www.discovery-on-the.net>
4. e-Science Data Mining Special Interest Group (*esdm-sig*): <http://www.nesc.ac.uk/resources/sigs/esdm-sig>
5. e-Science Visualization Special Interest Group (*evis-sig*): <http://www.nesc.ac.uk/resources/sigs/esvis-sig>