## UTAS “Big Data” Workshop 3-4 September

Tuesday September 3rd – Physics Lecture Theatre 1

<table>
<thead>
<tr>
<th>Time</th>
<th>Event</th>
</tr>
</thead>
<tbody>
<tr>
<td>8:45 AM</td>
<td>Registration/coffee</td>
</tr>
<tr>
<td>9:15 AM</td>
<td>Opening – Prof. Paddy Nixon, DVC (Research)</td>
</tr>
<tr>
<td>9:30 AM</td>
<td>David G. Barnes <em>Image visualisation and analysis beyond the data cliff: strategies for understanding large data</em></td>
</tr>
<tr>
<td>10:00 AM</td>
<td>Stephen Hardy <em>Data analytics at NICTA: quantifying uncertainty for decision making</em></td>
</tr>
<tr>
<td>10:30 AM</td>
<td>Edward King <em>TBA</em></td>
</tr>
<tr>
<td>11:00 AM</td>
<td>Coffee break</td>
</tr>
<tr>
<td>11:30 AM</td>
<td>Karen Meusemann <em>The 1000 Transcriptome Evolution project 1KITE: An example of handling &quot;Big Data&quot; from Evolutionary Biology</em></td>
</tr>
<tr>
<td>12:00 PM</td>
<td>Ben Simons <em>The Visualisation of Large Datasets with Houdini</em></td>
</tr>
<tr>
<td>12:30 PM</td>
<td>Lesley Wyborn</td>
</tr>
<tr>
<td>1:00 PM</td>
<td>Lunch</td>
</tr>
<tr>
<td>2:30 PM</td>
<td>Hands-on session 1 (Computer labs in Physics Rm 329 and Mathematics Rm 254)</td>
</tr>
</tbody>
</table>

Wednesday September 4th

<table>
<thead>
<tr>
<th>Time</th>
<th>Event</th>
</tr>
</thead>
<tbody>
<tr>
<td>10:00 AM</td>
<td>Nathan Bindoff (TBC) <em>Big Data at UTAS</em></td>
</tr>
<tr>
<td>10:30 AM</td>
<td>Group discussion</td>
</tr>
<tr>
<td>11:00 AM</td>
<td>Coffee break</td>
</tr>
<tr>
<td>11:30 AM</td>
<td>Hands-on session 2 (Computer labs in Physics Rm 329 and Mathematics Rm 254)</td>
</tr>
<tr>
<td>1:00 PM</td>
<td>Lunch</td>
</tr>
<tr>
<td>2:00 PM</td>
<td>Hands-on session 3 (Computer labs in Physics Rm 329 and Mathematics Rm 254)</td>
</tr>
</tbody>
</table>
Tuesday September 3rd : 9:30 AM

Image visualisation and analysis beyond the data cliff: strategies for understanding large data

David G. Barnes
Monash Biomedical Imaging, Monash University
Monash e-Research Centre, Monash University
Faculty of Information Technology, Monash University
VLSCI Life Sciences Computation Centre, Parkville

As biomedical imaging instrumentation is advanced, total data volumes increase in size, acquisition rate, and complexity. Spectacular examples include the Human Connectome Project and the Allen Human Brain Atlas. New informatics approaches are solving the challenges of storing, finding and retrieving large images and image collections, and instantiating sophisticated post-processing workflows. But what of traditional analysis and visualisation, and understanding and comprehending the data at hand? Driven by the data cliff that radio astronomy faces now and that biomedical imaging will encounter shortly, we have been exploring new strategies for enabling accelerated data analysis and visualisation in the terascale science era. In this talk I will describe the work we have been doing, and place it in the broader context of large data e-science.

Tuesday September 3rd : 10:00 AM

Data analytics at NICTA: quantifying uncertainty for decision making

Stephen Hardy
Technology Director, Computational Analytics, NICTA

The amount and variety of data available for analysis is expanding at an extraordinary rate and is driving major changes across science, health, business and government. In this talk I’ll give an overview of NICTA’s activities in the application of machine learning and optimisation to real world problems, focusing on data driven decision making and the role of machine learning in quantifying uncertainty. I’ll deep dive into a few specific areas - non-parametric statistics with application to infrastructure monitoring and geothermal exploration - and also outline a new project in machine learning for the natural sciences.

Tuesday September 3rd : 10:30 AM

TBA

Edward King
CSIRO Earth Observation Centre

Tuesday September 3rd : 11:30 AM
The 1000 Transcriptome Evolution project 1KITE: An example of handling "Big Data" from Evolutionary Biology

Karen Meusemann  
(on the behalf of 1KITE consortium), Postdoctoral Fellow, Australian National Insect Collection, CSIRO Ecosystem Sciences, Acton, ACT 2601

The 1KITE (1K Insect Transcriptome Evolution), started during the second half of 2011, will elucidate the phylogenetic relationships of more than 1,000 insects species based on Next Generation Sequencing of entire transcriptomes and the development of a new bioinformatics pipeline for analysing such a large dataset. Molecular datasets currently under construction will reach up to several billion million base pairs of aligned and homologised DNA sequence data. This data will be combined with data from morphology, paleontology, and embryology, and involves the collaboration of 70 researchers in these and related bioinformatics fields. As an example of empirical science, the general structure of the 1KITE project will be introduced on different levels (data collecting, data management, and data exchange including a short description of kinds of data the project deals with, e.g., primary sequence data, meta-data, and secondary processed and analyzed data). The IKITE database solution will also be addressed, touching on the aims and challenges. A brief glimpse will be given as well to 1KITE web-presence (website and internal WIKI) and how data is shared and accessed currently and prospects for future developments.

Tuesday September 3rd: 12:00 AM

The Visualisation of Large Datasets with Houdini

Ben Simons  
Lead Developer of the UTS Data Arena

In 2011 Dr D Studios in Sydney generated 2 Petabytes of data via 18000 cpu-cores during the creation of the Feature Animation "Happy Feet 2". This talk will cover some of the techniques used to tackle the big data involved in Film VFX, and how they may be applied in the generation of Computer Graphics for Big Data Visualisation.

3D Software typically used to create Film VFX, such as Houdini, is equally applicable to the visualisation of Large Data sets.

Specific approaches which operate in large data domains will be described, including Houdini's Channel Operators (CHOPS), Geometry Shaders (Vex Nodes & VOPS) and Instancing. Strategies which apply Level of Detail, Adaptive Meshes, Heirachical Data Formats, Caching & Filtering, and Replicated Data Stores will be outlined.
Tuesday September 3rd: 12:30 AM

TBA

Lesley Wyborn
   Geoscience Australia