Development and Support of Platforms for Research into Rare Diseases

Prof. Richard O. Sinnott
Technical Director, National e-Science Centre
University of Glasgow, Scotland
r.sinnott@nesc.gla.ac.uk

(Due to start as eResearch Director, University of Melbourne, July 2010)
NeSC Background

E-Science Hub

Externally

- Glasgow end of NeSC
  - Involved in numerous UK wide activities/projects

Internally

- Focal point for e-Science research/activities at Glasgow
- Work closely with foundation departments
  - Department of Computing Science
    » Established first UK Grid Computing course
  - Department of Physics & Astronomy
- Also working with other groups including
  - Bioinformatics Research Centre,
  - Biostatistics
  - Electronics and Electrical Engineering
  - Dept of Public Health, Dept. of Pathology,
  - Dept. of English, Arts & Humanities,
  - University Services,
  - Clinicians & numerous hospitals across Scotland,
    » Yorkhill, Royal Infirmary, Western General, Southern General ...

NeSC GU now part of University IT Services
# NeSC Glasgow Projects

## Completed
- National e-Science Centre (NeSC-I, NeSC-II, NeSC-III)
- Dynamic Virtual Organisations for e-Science Education (DyVOSE)
- Biomedical Research Informatics Delivered by Grid Enabled Services (BRIDGES)
- Grid Enabled Microarray Expression Profile Search (GEMEPS)
- GridNet
- Glasgow early adoption of Shibboleth (GLASS)
- Joint Data Standards Survey (JDSS)
- ESP-Grid
- GridNet-2
- HPC Compute cluster award
- Sun industrial sponsorship
- OGC Collision
- OMII-Security Portlets
- OMII-RAVE
- Integrating VOMS and PERMIS for Superior Grid Authorization (VPman)
- NCeSS Technical Management
- CESSDA PPP
- Pharming of Therapeutic RNA
- Grid Enabled Occupational Data Environment (GEODE)
- Towards an e-Infrastructure for e-Science Digital Repositories
- Grid enabled Biochemical Pathway Simulator
- Virtual Organisations for Trials and Epidemiological Studies (VOTES)
- Towards a European e-Infrastructure for e-Science Repositories
- Modelling, Inference and Analysis for Biological Systems up to the Cellular Level
- Drug Discovery Portal
- Advanced Grid Authorisation through Semantic Technologies (AGAST)
- ShinTau (Supporting Multiple Shibboleth Attribute Authorities)
- Grid-enabled Virtual Safe Settings - Security & the State of the Nation

## Running
- Scottish Bioinformatics Research Network (SBRN)
- Generation Scotland Scottish Family Health Study
- Meeting the Design Challenges of nanoCMOS Electronics (nanoCMOS)
- **EU FW7 EuroDSD**
- **EU FW7 AvertIT**
- Breast Cancer Tissue Biobank
- Management through e-Social Science (DAMES)
- NeSC Research Platform (NRP)
- NeSC Information Network (NIN)
- **ESF Network for Study of Adrenal Tumors**
- Scottish Health Informatics Platform for Research (SHIP)
- National E-Infrastructure for Social Simulation (NeISS)
- Enhancing Repositories for Language and Literature Researchers (ENROLLER)
- Proxy Credential Auditing Infrastructure for the NGS
- EU FW7 European Network for Study of Adrenal Tumors Cancer Research Platform
- **EU R4SME Diagnosis of Parkinsons Disease (DiPAR)**
Rare Diseases

- Data sharing especially important
  - Urgent need for
    - Critical mass of data (phenotypic/genotypic/treatment/…)
    - Understanding of best treatments & outcomes
    - Common platform for collaboration
      - Data harmonisation
      - Information Governance, Ethics, Security
        » Is/should be at the heart of all systems in this space!!!
      - Data disclosure issues

- Example rare diseases
  - Adrenal tumours
  - Disorders of sex development
European Network for Study of Adrenal Tumours (ENSAT)

- Adrenal gland
  - Small gland near kidneys
  - plays an essential role in adaptation to major stress (infection/trauma/shock)
  - in response to stress, adrenal gland secretes enhanced cortisol and catecholamine
  - adrenal impairment - stress response severely diminished and can lead to irreversible shock/death

- European Science Foundation-funded ENSAT consortium ([www.ensat.org](http://www.ensat.org)) especially interested in four major types and causes of adrenal tumours
  - **Adrenocortical carcinoma (ACC)**
    - postoperative disease free survival of patients diagnosed with ACC is below 50% over a 5-year period
  - **Aldosterone Producing Adenoma (APA)**
    - a form of cancer which results in secondary hypertension accounting for up to 5-10% of all hypertensive patients
  - **Non-aldosterone cortical adrenal adenomas (NAPACA)**
    - < 100 cases worldwide reported where adenoma secretes only androgens
  - **Pheochromocytomas and related paragangliomas (Pheo)**
    - Prevalence is about 0.1% in patients with hypertension and 4% in patients with a discovered adrenal mass.
    - ~10% of these tumours are malignant either at first operation or during follow-up
### ENSAT Data Models

- Data models (ACC, APA, NAPACA, Pheo) defined by clinicians over extended time period
  - Initial model of data sharing based on software/data on CDs

<table>
<thead>
<tr>
<th>Identification</th>
<th>Filled-in once</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patient History</td>
<td>Filled-in once</td>
</tr>
<tr>
<td>ENSAT Workup</td>
<td>Filled-in at diagnosis</td>
</tr>
<tr>
<td>Tumor Form</td>
<td>Filled-in after operation</td>
</tr>
<tr>
<td>Sample Form</td>
<td>Filled-in by teams with a Biological Research Centre</td>
</tr>
<tr>
<td>Follow-up Form</td>
<td>Filled-in at (yearly) follow up</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Year of diagnosis of ACC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Previous malignancy, if yes, specify (free text)</td>
</tr>
<tr>
<td>Cushing’s syndrome, if yes, tick one</td>
</tr>
<tr>
<td>Virilisation, if yes, tick one</td>
</tr>
<tr>
<td>Menstrual disorders, if yes, tick one</td>
</tr>
<tr>
<td>Feminisation, if yes, tick one</td>
</tr>
<tr>
<td>Hypertension, if yes, tick one</td>
</tr>
<tr>
<td>Diabetes, if yes, tick one</td>
</tr>
<tr>
<td>Hypokalemia, if yes, tick one</td>
</tr>
<tr>
<td>Abdominal pain, if yes, tick one</td>
</tr>
<tr>
<td>Palpable abdominal mass, if yes, tick one</td>
</tr>
<tr>
<td>Venous thrombosis, if yes, tick one</td>
</tr>
<tr>
<td>Incidentaloma, if yes, tick one</td>
</tr>
<tr>
<td>Others</td>
</tr>
</tbody>
</table>

- Internet-based solution highly desirable
ENSAT Collaboratory

- Development of ENSAT Data Sharing and Collaboration Infrastructure
  - Add patient data
  - Update patient data
  - Decide on granularity of sharing of patient data
  - Ethics of data sharing
  - Initial work for all ENSAT data resources completed in 2 months
ENSAT Notes

- No patient identifying information
  - Unique identifier auto-generated and used by clinician for local linkage/tracking
  - Only clinician identified
- Access and usage aligned with EuroDSD (more later)
- Long term follow up needed
  - Multiple edits to same patient records
    - (can make searches challenging)
- Match making service for clinicians, research and bio-research laboratory centres
  - Prototyping work has now lead to recently funded 6m€ ENSAT-CANCER project
    - due to start January 2011
  - Data sharing and bio-banking oriented collaboration
    - Includes major clinical trials to test new drugs/treatments on ACC and Pheo
Disorders of Sex Development

- Investigation of the molecular pathogenesis and pathophysiology of Disorders of Sex Development (DSD) - EuroDSD
  - Extreme Data Sensitivity
    - Typically 10-12 cases in Scotland per year
    - Family issues, child issues, legal issues, ...
  - 3-year project started May 2008
    - Built on initial software prototypes funded by ESPE by NeSC (in 1 month)
  - Clinical Contributors
    - UK, Germany, Netherlands, Italy, Sweden, France
  - Research partners
    - UK, France, Germany
  - Systems have gone through international (independent ethical review) and are now used for range of research areas into DSD
E-Security

- E-Security
  - AAA ... and the clubcard!

- Users like usernames/passwords
  - Provide them (once!)

- Users don’t like/understand X.509 based PKI
  - Forget training, education for most users!
    - `openssl pkcs12 -in cert.p12 -clcerts -nokeys -out usert cert.pem`
  - Should all be transparent to end users and aligned with the way that they want to work/access resources
    - Access Management Federation (Shibboleth) + authZ technologies
Shibboleth Federated VO-based Approach

1. User points browser at resource/portal
2. Shibboleth redirects user to W.A.Y.F. service
3. User selects their home institution
4. Home site authenticates user and pushes attributes to the service provider
5. Pass authentication info and attributes to authZ function
6. Make final AuthZ decision

Identity Provider

Service provider

Shib Frontend

Grid Portal

User points browser at resource/portal

Home Institution

User

W.A.Y.F.

Federation

 LDAP

AuthZ

AuthN

uid

LDAP

LDAP

Shibboleth

LDAP

AuthZ

Identity Provider

Service provider

Identity Provider

Home Institution

User

Home Institution

User
EuroDSD Work

- WP1 - aims to support complete Virtual Research Environment (VRE) for research into DSD
  - More later
- WP2 - Identification of novel genetic markers for DSD
  - Design and validate a DSD GeneChip and identify and confirm novel genetic markers of DSD
  - Identify deletion/duplication events in the human genome that are associated with DSD
- WP3 - Functional assessment of androgen action: AR mutant analysis
  - Europe-wide study of patients with partial androgen insensitivity syndrome (PAIS) to correlate phenotype, androgen receptor (AR) mutation and function in vitro with pubertal outcome
  - Define the molecular function of the AR and associated PAIS mutations during early genital development
  - Facilitate production of vas deferens cell lines stably expressing mutant AR for functional analysis
EuroDSD Work...ctd

- **WP4** - Characterization of the “androgen-memory” in cultured labioscrotal fibroblasts of XY-DSD individuals due to AIS by high throughput methylation analyses related to phenotype, genotype, molecular androgen receptor (AR) function and puberty outcome.

- **WP5** - Steroid Metabolomics
  - To employ steroid profiling by gas chromatography/mass spectrometry and liquid chromatography/tandem mass spectrometry as a discovery tool in patients with 46,XY DSD to identify hitherto unknown steroidogenic disorders as underlying biochemical and genetic causes of 46,XY DSD.

- **WP6** - DSD e-learning webportal
  - Interactive learning environment for an up to date program on DSD including normal development, patho-physiological mechanisms and current views on diagnostic and therapeutic interventions, psychological counselling and outcome.

- **WP7** - Project Management
At the heart of the VRE is the Registry for registering cases

“...we expect that around 1800 primary cases will fulfil the requirements for data entry, approximately 300 from each centre”

Other tools

Collaborative tools, e.g. wikis, Bioinformatics tools
## Registry Data Model

### CORE DATASET – ESPE DSD REGISTRY

<table>
<thead>
<tr>
<th>Field</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Year of Birth</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Local Hospital Identification</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Date of First Notification</strong></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Field</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Centre</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Reporting Clinician</strong></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Field</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Clinical Presentation</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Sex assigned</strong></td>
<td>Male o Female o NK o</td>
</tr>
<tr>
<td><strong>Phallus Size</strong></td>
<td>Normal Male o Small Male o Large Male o Normal Female o Large Female o Chordee o NK o</td>
</tr>
<tr>
<td><strong>Phallus Length</strong></td>
<td>Data o Length o Resection o NK o</td>
</tr>
<tr>
<td><strong>Urinary Meatus</strong></td>
<td>Normal Male o Hypospadias o Dipscis o Mid o Pruneumlus o Normal Female o NK o</td>
</tr>
<tr>
<td><strong>Labioscrotal Fusion</strong></td>
<td>Yes o No o NK o</td>
</tr>
<tr>
<td><strong>Right Gonad</strong></td>
<td>o</td>
</tr>
<tr>
<td><strong>Left Gonad</strong></td>
<td>o</td>
</tr>
<tr>
<td><strong>Extrernal Masculinisation Score</strong></td>
<td>(0 - 12) o NK o</td>
</tr>
<tr>
<td><strong>Tanner Stage</strong></td>
<td>(1 - 5) o NK o</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Field</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Internal Sex Organs</strong></td>
<td>Male o Female o NK o</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Field</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Karyotype</strong></td>
<td>46XY o 46XX o Presumed XY o Presumed XX o 46XX/46XY o 46X/46XY o Other o</td>
</tr>
<tr>
<td><strong>Disorder Type</strong></td>
<td>Disorder of Gonadal Development o Disorder of Androgen Synthesis o Disorder of Androgen Excess o Nonspecific Undermasculinisation Disorder o Other o</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Field</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Actual Diagnosis</strong></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Field</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Certainty of Diagnosis Based On</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Clinical Features</strong></td>
<td>Certain o Uncertain o NA o</td>
</tr>
<tr>
<td><strong>Biochemistry</strong></td>
<td>Certain o Uncertain o NA o</td>
</tr>
<tr>
<td><strong>DNA Analysis</strong></td>
<td>Certain o Uncertain o NA o</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Field</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Availability of Further Information</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Case Notes</strong></td>
<td>Yes o No o NK o</td>
</tr>
<tr>
<td><strong>Growth Data</strong></td>
<td>Yes o No o NK o</td>
</tr>
<tr>
<td><strong>Puberty Data</strong></td>
<td>Yes o No o NK o</td>
</tr>
<tr>
<td><strong>DNA</strong></td>
<td>Yes o No o NK o</td>
</tr>
<tr>
<td><strong>Thyrox</strong></td>
<td>Yes o No o NK o</td>
</tr>
<tr>
<td><strong>Cell Line</strong></td>
<td>Yes o No o NK o</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Field</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Family</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Parental Consanguinity</strong></td>
<td>Yes o No o NK o</td>
</tr>
<tr>
<td><strong>History of DSD</strong></td>
<td>Yes o No o NK o</td>
</tr>
<tr>
<td><strong>History of Infertility</strong></td>
<td>Yes o No o NK o</td>
</tr>
<tr>
<td><strong>Availability of Samples</strong></td>
<td>Yes o No o NK o</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Field</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Free Text</strong></td>
<td></td>
</tr>
</tbody>
</table>

---

**Notes:**
- Please note that any changes are subject to peer review.
- Further copies & Guidance are available from Register Coordinator.

---

**Contact:**
- Please fill all relevant copies together in case notes.

---

**UNIVERSITY of GLASGOW**

**CCGrid 2010, Melbourne,**

**17th May 2010**
EuroDSD Notes

- Data sharing
  - At discretion of clinicians/collaborators
    - Records only editable by person who added them

- Ethics and Standard Operating Procedures
  - Numerous documents/guidelines produced
    - The ESPE DSD Registry SOP [PDF]
    - App1. Data Flow within ESPE DSD Registry [PDF]
    - App2. ESPE DSD Registry information leaflet [PDF]
    - App3. ESPE DSD Register information leaflet children [PDF]
    - App4. Application to REC for approval of ESPE DSD Registry [PDF]
    - App5. Confidentiality Policy [PDF]
    - App6. Corporate Information Security Policy [PDF]
    - App7. System Level Security Policy (SLSP) [PDF]
    - App8. Caldicott Recommendation [PDF]
    - UG Data Protection Register details [PDF]
    - UK ethics approval for ESPE DSD Registry [PDF]
EuroDSD Roles

- **EuroDSD-investigator**
  - Responsible for clinical cases / ethics from a particular contributing site.
  - Contact person responsible for cases and charged with ensuring that the site adheres to the SOPs.

- **EuroDSD_contributor**
  - Person who is charged with adding cases to the registry
  - Typically this is delegated to a research nurse at a given site

- **EuroDSD_researcher**
  - Person who wishes to search across the EuroDSD data

- **EuroDSD_local**
  - A non-EuroDSD partner who wishes to use the VRE for their own personal use.
  - This role does not give access to data sets and their data sets are not added to EuroDSD.
Demonstration
VRE Statistics

As of this morning

- 654 cases and moving in the right direction!

- Numerous major papers published
  - NEJM, Cell, ...

- Many other groups now adopting this solution for own DSD data mgt
  - Argentina, Czech Rep, Estonia, Jordan, Morocco, Sudan, Turkey, UK, USA, ...

- Support 24/7 access / usage for over 2 years
Conclusions

- Systems driven by Information Governance/Ethics/Consent
  - MREC, LREC, PAC, PIAG, Caldicott Guardians
    - Research needs - not what middleware/IT can do!!!
  - Strict adherence to Standard Operating Procedures
  - NHS contracts, trust relationships
- Then “and only then” once defined we have the tools and techniques to roll-out e-Infrastructures
  - to support researchers
  - to improve patient/population health
    - (not described but also work in pull-oriented mode, e.g. where firewalls cannot be opened for incoming traffic)
    - (variety of clinical anonymisation routines supported)
- Scottish Health Informatics Platform
  - £3.5m funded by Wellcome Trust, EPSRC, ESRC, MRC
Questions …?