Delivering e-Science to Clinicians

John Ainsworth, Robert Harper, Lucy Bridges, Pauline Whelan, William Vance, Iain Buchan
School of Medicine, The University of Manchester
john.ainsworth@manchester.ac.uk
http://www.psygrid.org
Technical Challenges

- Data quality and integrity
- Timeliness of response
- Privacy and confidentiality
- Knowledge representation
This paper is about how we developed and operate the PsyGrid system and how we organised the PsyGrid project.

PsyGrid does what our user community wants; problems are resolved and new features added continually.

There are some specific things about the clinical community that shaped the project.

... and some requirements that complicated the architecture and operational procedures ...

... but mostly its just about doing the basics well.
What is PsyGrid?

- e-Science project funded by the MRC, and the Department of Health

- **Aim 1 (Technological):** Develop applications and middleware to support studies and clinical trials in mental health
  - Software
  - A supported operational system

- **Aim 2 (Clinical):** Use outputs of ‘Aim 1’ to perform longitudinal study of schizophrenics from First Episode Psychosis over 2 years and study epidemiology of schizophrenia
E-Infrastructure for studies and trials
Project History

- Began in May 2005
- Electronic data collection began May 2006
- Now support 5 studies in mental health; 2 studies in diabetes imminent; over 100 users; over 1500 subjects
- Number of studies growing monthly
- Adopted by the UK MHRN for all studies and trials
Risk Analysis

- The first thing we did
- Can inform all aspects of the project ...
  - Development process
  - Technology decisions
  - Operations
  - Deployment
  - Support
Risk Identification (1)

- PsyGrid to be deployed on NHS
- Stakeholders: NHS Trust IT directors
  - Deployment and operation must not require any changes to a Trust’s existing network security policy
  - The system must not deny resources to existing clinical systems or require any support from NHS IT personnel
  - Assume the most restrictive infrastructure exists
    - It does! Only port 80 and 443 reliably open across the NHS. Some Trust’s web proxies use Microsoft proprietary NTLM authentication.
Risk Identification (2)

- Cost of operation prevents adoption and use
- Stakeholders: PsyGrid developers; clinical research networks

Three costs to consider:
- Cost of the software
  - eliminated as PsyGrid is open source GPLed
- Cost of supporting hardware and its maintenance
  - Support a range of hardware deployments; from single box solutions to fully redundant hardware architectures
  - Economies of scale apply.
- Costs of designing and implementing studies
  - Provide easy to use tools to design, configure and deploy new studies
  - Enable “round-trip study management” by clinicians
  - Integrate with governance systems
Risk Identification (3)

- Immature technology causes instability
- Stakeholders: PsyGrid developers
  - Conservative middleware selection
    - WS-I compliant plain web services, no Grid middleware.
  - Prefer stable and mature over latest and greatest
  - Focus on application, not innovation; its innovation for the scientists, not the technologists
Risk Identification (4)

- User Resistance
- Stakeholders: The clinical community

- CI & CPM: Nothing to lose!
- PIs; Clinical Academics: Mostly positive but sceptical
- RNs: Use the system the most; but benefit the least
Risk Identification (4 cont.)

- Very easy to alienate the Research Nurses
  - Make sure it works (robust, reliable, quick enough)
  - Make sure they are trained
  - ... but don’t expect them to acquire a new level of IT sophistication (X.509 anyone?)
  - User centred design
  - Immediate response 1st line support
  - Fix problems quickly
  - Implement user feedback => sense of ownership
Organization

- Single office base for the development team
  - Light weight, informal process
- One PM; 4 developers
- Shared ownership
  - Just as well - 60% staff turnover after 18 months
Development Process

- Users are clinical people
- Little knowledge of IT; do not know what is possible and what is not
  => Evolutionary Prototyping
- New ideas generated all the time
  => Agile development
Release Process

- Major functionality release every 6 months (approx)
  - Corresponding to work package deliverables

- Continuous development and release of minor features
  - Reflects need to respond positively to user feedback

- Each release soak tested on staging server before deployment to NHS
  - Approximately one release every two weeks

- Once deployed execute test suite on a test data set
Operations

- Developers do support (email & telephone)
  - More building relationships between users; no hiding when it’s not working!
  - Rotates throughout the team; one week shifts; improves knowledge of the system across the team

- Errors are reported automatically if possible
  - Some users do not report them because sending a log file was too difficult

- The majority of support requests (90%) are to do with either NHS connectivity problems or confusion about on-line/off-line operation
Operational Issues

- **Patching**
  - Modifying a live dataset
  - Mostly additions, but some deletions

- **Online/offline mode issues**
  - Causes confusion amongst users
  - Data synchronisation issues between local copy and repository ...
    especially when the dataset is patched
Conclusion

- Identify risks and mitigation
- Identify stakeholders and their needs
- Understand your users and engage with them
- Create positive feedback between developers and users
- Do the basics well!
- It's e-Science – not rocket-Science

- PsyGrid Demonstration in the MRC booth @ 2pm tomorrow