RDMF 11th March 2010

Dealing with Sensitive Data

Summary

Graham Pryor
Legend

Tracing the human infrastructure thread is marked by white text

Actions for, or already planned by DCC/RIN shown by light blue text
DWP

• Stores/uses/accesses personal data
• Security upgraded following data losses
• Highly systematised, based on risk analysis (Generic Risk Profile):
  – Assurance for Research & Analysis (SARA)
  – Data Labs – limits on contractor access
• Data Access Ethics Committee
  – Independent/Govt. membership
  – Influential, wide-ranging brief to review and recommend
NDAD

• Data selected by National Archives
• Data preserved with its imperfections, requiring contextual information
• Complex and large datasets
• Data withheld on 30 year rule, DPA, FOI
• Regulation of sensitive content not systemic in source depts.
• NDAD always checks for sensitive data
• Source dept. has to sign off before release
• Active redaction of data/metadata
• Reasons for restrictions defined; history of restrictions preserved – responsibility to Public
• Work closely with data owners – need trust

GHSP/11-03-2010
UKDA SDS

- Promote researcher access, minimise (sensitive) data disclosure
- Apply access criteria
- Use data security model to ensure
  - safe project, safe people, safe data, safe setting, safe output, safe use
- Platform (Citrix) security – no data removal, no outbound traffic, separate work areas, final output checked by staff
- Users are approved and trained (no training no access!)
- Measures for robust response to misuse/breach of legislation
UKDA RDMS

• Enabling the sharing of sensitive data
• Research ethics guidelines – duty of confidentiality, duty to include participants in data decisions, duty to share with public
• Be clear what is sensitive and what is not
• Apply key principles – informed consent, managed access, protected identities, secure storage
• Guidance to researchers on ethical data management/sharing
• Cannot make anonymised data open – need to be seen to sustain trust of researchers (data producers)
CITL

- Institutional role in managing risk – esp. for ‘hot button’ topics?
- Researchers need to be prepared and be given relevant support
- Responsibility for curating research data?
  - Not practical to leave to research teams
  - Other effective data management?
- Need researchers to think about risks specific to their project
- Need a process of audit
  - What was produced and how is it being managed?
  - What has changed as a result of the research process?
  - When should the data be reviewed?
- Recognition of a data lifecycle – beyond the research lifecycle
Messages (_presentations_)

• Institutional repositories organised as a bundle of services, not passive destination
  – Advisory, Training, Data Policy and Compliance, Preservation/Curation, Audit
  – Relationship with research project lifecycle

• Appropriate and achievable sanctions?

• **DCC training programme**: data ethics, managing sensitive data
Messages (breakout 1: good practice checklist)

• Templates for controls at different levels
• Senior management enforcement
• Set of policies, identified responsibilities and processes, plus sanctions
• Central knowledge ‘store’
• Infrastructure of support services inc. training, with…
• Mechanism to engage with researchers from conceptual stage
• Means of identifying what is sensitive data
• Researcher awareness of data issues and support mechanisms
• Means for proper ‘closure’
• DCC – advocacy at high level, best practice models, targeted training, work through pilot
Messages (breakout 2: optimal technical approaches)

• Access policy (how to express), access control arrangements
• Pre-ingest/ingest – know what’s coming
• Capability of stakeholders
• Important to have policies for data handling – and buy-in
• Data controller
• Funding ends – whose responsibility for data?
Messages (breakout 3: training for data managers)

• Develop a national standard ‘driving licence’ – start as a pilot?
• Data mgt. in context of researcher devpt. – DCC ‘Train the Trainer’?
• Ethics committees and IT managers – need broader perspective (data lifecycle)
• Training needs in data mgt. plans?