



# Data Security and Trust

## BoF Session

Rebecca Wilson, Paul Burton, Rudolf Mayer  
17.09.2016

research data sharing without barriers  
[rd-alliance.org](http://rd-alliance.org)

# Agenda

- Welcome and introduction to the WGDST
- Aims for this meeting [5 mins]
- Key background [5 mins]
- Aims of the working group
- Scope [5 mins]
- First steps - current practices
  - Survey results [5 mins]
  - Conclusions from WUN workshop [10 mins]
- Discussion of next steps [30+ mins]
  - Aims, scope and results
  - WG structure, modus operandi and future meetings
- Wrap up [5 mins]

# Data Security and Trust WG

- BoF at RDA P6 (Paris)
- First iteration of case statement with use cases
  - Available in WG Wiki:
  - <https://rd-alliance.org/groups/working-group-data-security-and-trust.html>
- Feedback from TAB: EU-centric
  - Identify new members / co-chairs

# Specific aims for the meeting (11:00-12:30)

- Identify new members
- Agree aims and scope of WGDST
- Discuss preliminary findings of survey and WUN Workshop
- Discuss and agree next steps of work in the WG
- Discuss and agree structure and operation of WG
- Feed into refined case statement for the establishment of a **RDA Working Group on Data Security and Trust**
- Discuss links with other WGs and IGs
- Interaction and meetings going forwards

- Trust
  - Trust is essential for sharing data across geographical, institutional and technological boundaries
  - Trust is hard to establish, but easy to lose
  
- Mutual trust can be established when:
  - Research partners can agree on data protection standards.
  - The compliance to these standards can be ensured
  - A common understanding of the data security requirements exists
  
- Trust can be established when we have an additional (technical) layer of trustworthiness

# Trustworthiness by Standardisation and Agreement

- Trustworthiness can be established by
  - Defining what the privacy requirements for a research data set are
  - Showing that these requirements can be met by:
    - Technical standards
    - Best practices
    - Policies
  - Access can be granted and revoked based on
    - Valid criteria which we can agree upon
    - The processes are transparent and comprehensible
    - Fair for all stakeholders
  - Agreement needed on:
    - Standards (authenticity, integrity)
    - Best practices (data sharing)
    - Policies (privacy)
- We will focus on the development, application and evaluation of technology to further progress in the agreed context

# Benefits for all Stakeholders

- Agreeing on good practice and standards has benefits for several stakeholders.
- Researchers:
  - Assess the trustworthiness of research data beforehand
  - Rely on secure data and verify integrity of data sources
- Data providers:
  - Provide verifiable data sets
  - Establish and maintain trust in your organization and its data
  - Increased willingness to share data
- Regulators and funding agencies:
  - Security audit recommendations

# Aims of this WG

- Group focussed on **technical solutions**
- Gather the different aspects of research data security and trust
- Distill common questions
- Learn what works for most participants
  - Where are currently deficits?
  - Identify the current best practices and find commonalities
  - Agree on a set of minimum standards
- Goal: Explore opportunities & challenges and come up with a **set of guidelines** on how institutions define security requirements and how they can exchange their data in a secure way.



# The Scope of the WGDST

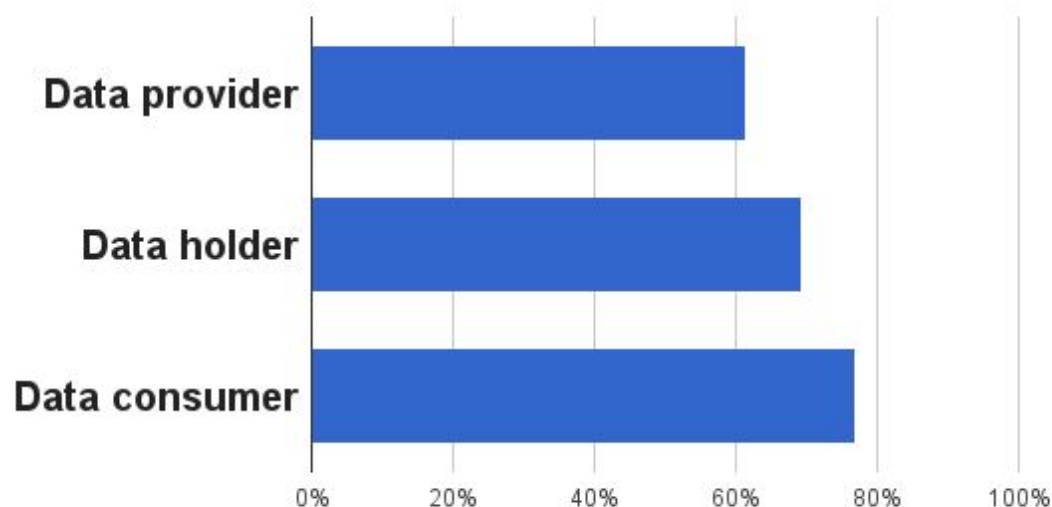
- Not all relevant questions can be addressed within this group
- Technical solutions & methods aimed at addressing specific challenges
  - Within a context founded on a data environment that is already appropriate and fit-for-purpose
- Focus on practical topics achievable over 18 months
  - Address high priority topics achievable in this time frame
  - Address the lowest hanging fruits first
  - Frame the program in term of new opportunities for data sharing if these issues were solved
- Possible disaggregation of topics
  - Means of data access/transfer
  - Single source vs multi source
  - Data partitioning: horizontally, vertically and complex

# First Steps: Current Practices Survey

- Survey to gather feedback from practitioners in the domain
  - Rather short, 10 questions
- Still open for responses: <http://bit.ly/rdadst>
  - Please give your input!
  - Only takes a few minutes to fill it in...
- Presentation of preliminary results

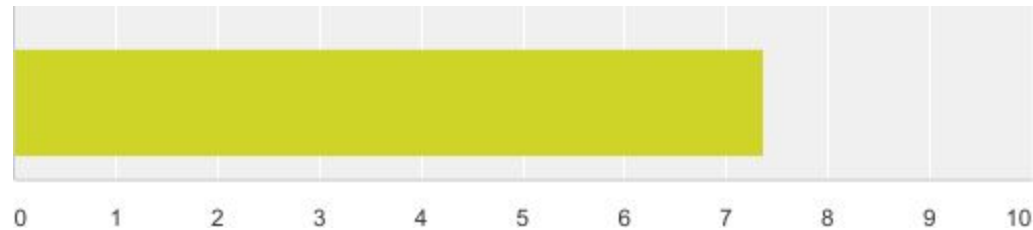
# First Steps: Current Practices Survey

- Participation from diverse domains
  - Bioinformatics, Biomedical sciences, Economics, Genomics, Public Health, Mathematics, Computer Science
- Data located in UK, other EU countries, USA/Japan
- Respondents have multiple roles

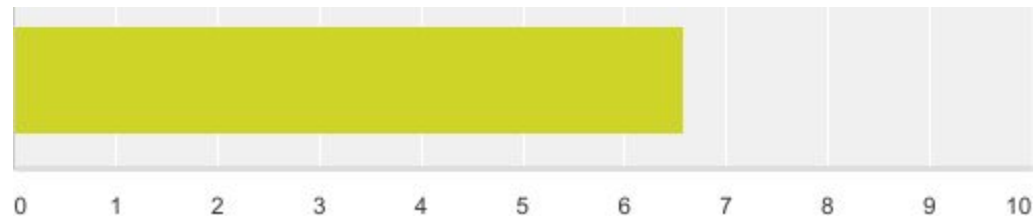


# First Steps: Current Practices Survey

- Concern about infrastructure security (scale 1-10)

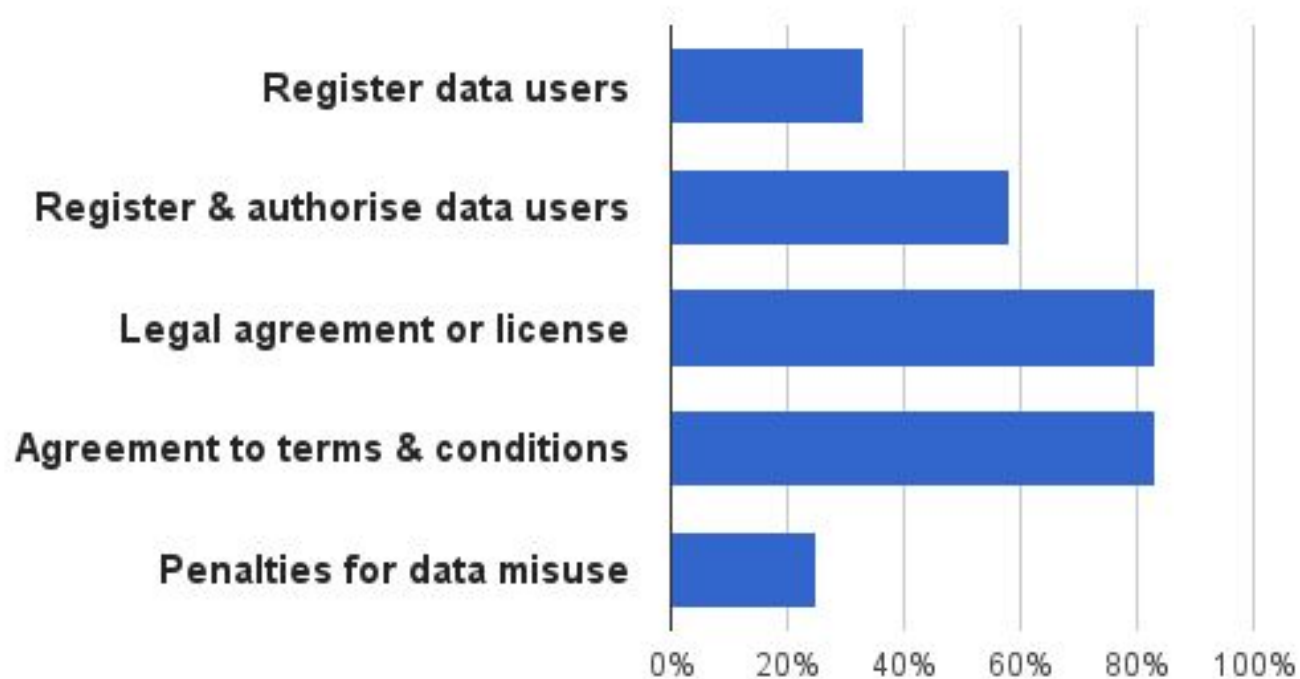


- Concern about illegitimate use/inference of data (1-10)



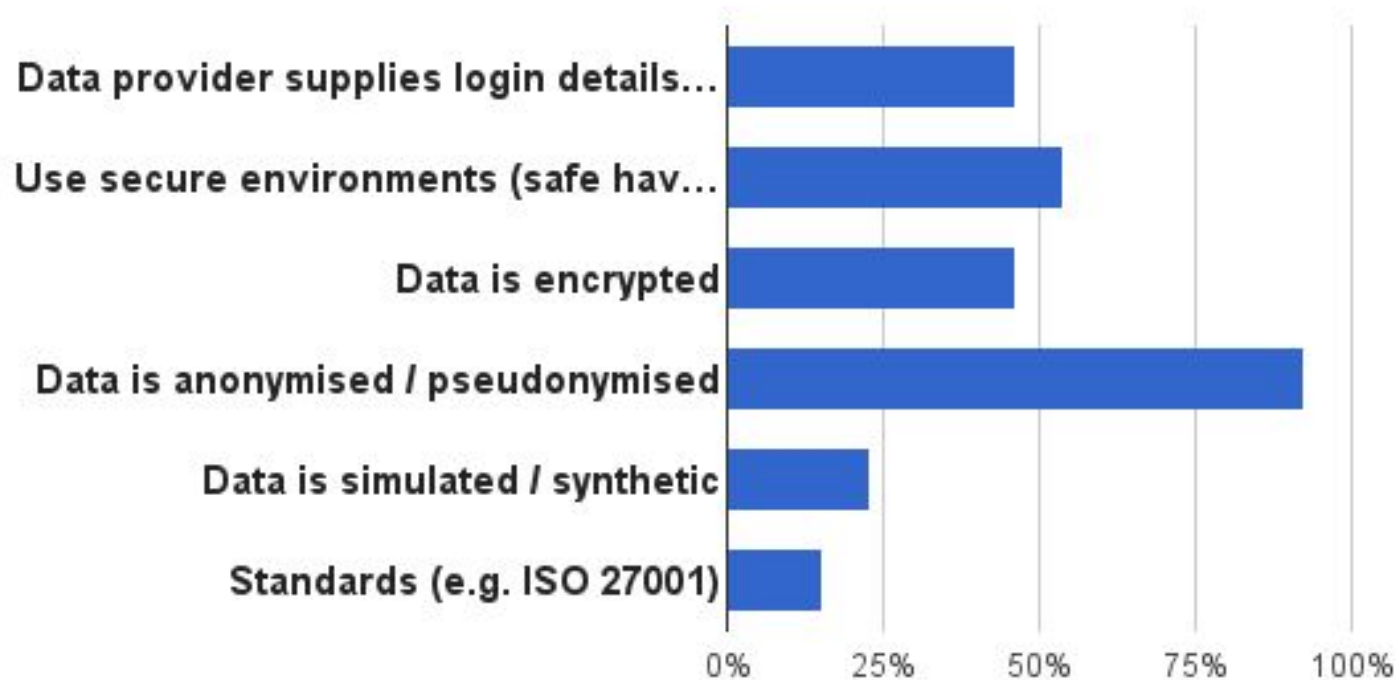
# First Steps: Current Practices Survey

- Currently used governance solutions



# First Steps: Current Practices Survey

- Technical solutions used to maintain security/privacy



# First Steps: Conclusions from the Data Analysis with Privacy Protection for Epidemiological Research (DAPPER) workshop

21-23rd August University of Bristol [www.bristol.ac.uk/dapper](http://www.bristol.ac.uk/dapper)  
Satellite meeting to IPDLN Conference 2016



- International examples:
  - Current methods & processes
  - Within and outside our domain
  - Governance - legal obligations
  - Software and tools demonstrations
  - Cultural differences

@Data2Knowledge  
@DrBeccaWilson





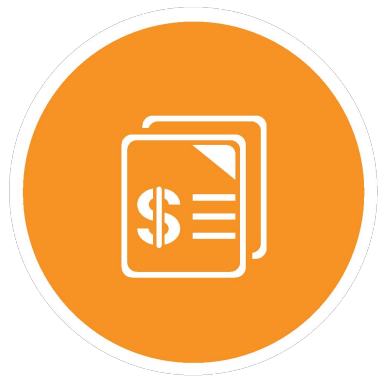
# SCIENCE AND TECHNOLOGY BARRIERS TO PROGRESS

- **ELS-governance-technical issues:** good awareness
- **Funding:** broken
  - Competition vs collaboration
  - Project based funding tied to answering research question.
  - No sustainability
- **Lack of coordination:** in community to drive change
  - ELS-governance-technical providing new processes/solutions
  - Data stakeholders e.g. funders, data owners, policy makers, users

@Data2Knowledge  
@DrBeccaWilson







## ENABLING RESEARCH

- **Wider dissemination:** best practice, exemplars, champions
  - Internationally & Interdisciplinary. Assistance in PR.
- **Funders need to be more involved:** take ownership of some issues
  - New funding models to build/sustain technical-social infrastructure
    - create generalisable, scalable sustainable solutions
    - Answer millions of research questions - not just a handful

@Data2Knowledge  
@DrBeccaWilson





# STRATEGIES FOR MOVING FORWARD

- **Tool / process integration:** many existing complementary solutions
  - Users can then adopt most appropriate components for them
- **Communication:** Advertising and PR of best practice
  - Additional work required in outreach / education
    - increase public trust in researchers & use of medical data
- **Organising the community:** take ownership and give a voice
  - lobby for sustainable investment, changes in policy & governance

@Data2Knowledge  
@DrBeccaWilson





# STRATEGIES FOR MOVING FORWARD

Creation of an international, professional stakeholder community with interest in developing strategies to facilitate health data sharing / reuse

Data Sharing for Health - INnovation Group (DASH-ING)

Feeding into and drawing from expertise in a number of RDA IGs/WGs

Subscribe to our mailing list to contribute to our community:  
[www.jiscmail.ac.uk/DASH-ING](http://www.jiscmail.ac.uk/DASH-ING)

@Data2Knowledge  
@DrBeccaWilson



# Topic-specific scoping reviews

- Identify priority topics
- Scope nature, challenges and opportunities associated with each topic
- Identify and disaggregate key tasks/subtasks and review current approaches to addressing these tasks
- Review good practices
- Identify gaps
- Consider ways that the gaps could be filled

# Structuring the discussion

- Possible disaggregation/identification of candidate topics
  - Domain: e.g. microdata, metadata, publication
  - Data partitioning: single-site, multisite: horizontal, vertically and complex
  - Infrastructural model: federated approaches, physical warehousing, secure enclaves, data-centric approaches
- Any comments on background, WGDST survey results and conclusions from WUN workshop?
- What is the research case (will the WG produce something useful)? And business case (will it be used)?
- Do we have critical mass (are the right people involved to do something useful)?
- Potential new members/co-chairs
- Links with other WGs and IGs
- Need to refine case statement to establish **WGDST**

# Closing thoughts / Next Steps

- Establish diverse chairs and core WG team
- Contribute: Join our working group  
<http://rd-alliance.org/groups/working-group-data-security-and-trust.html>
- Comment in our survey to gather current data security and trust practices <http://bit.ly/rdadst>
- Minutes of the meeting and slides on the WGDST wiki
  - Reshape the case statement
  
- Reaching out to other IG/WGs
  - IG Health Data: Health Data Privacy & Security issues
  - WG RDA/NISO Privacy Implications of Research Data Sets
- Interested in the domain of Health data:  
[www.jiscmail.ac.uk/DASH-ING](http://www.jiscmail.ac.uk/DASH-ING)



RESEARCH DATA **ALLIANCE**

research data sharing without barriers  
[rd-alliance.org](http://rd-alliance.org)