Legal, ethical, and governance challenges & potential solutions

Dr Esther van Sluijs
MRC Programme Leader
MRC Epidemiology Unit & Centre for Diet and Activity Research
@EvanSluijs

Data harmonisation relies on data sharing

- Opportunities for data sharing need to be fostered and supported throughout the research process

Facilitating da

Early stage planning

Box 1. EAGDA recommendations on the governance of data access

1. All project proposals should include data sharing and management plans in funding applications.
2. Funders should support the development of data and metadata standards.
3. Data access processes should be discoverable and transparent for potential data users.
4. Studies should establish proportionate governance mechanisms for data access.
5. Collaboration should not be a necessary condition for data access.
6. Consent should, as far as possible, include provision for further data use beyond the original study.
7. Clear policies should be developed on how depletable resources will be managed.
8. Funders should establish clear penalties and sanctions for breaches of data-sharing rules.
9. Principles of data access should be harmonized as far as possible across studies. Study leaders should also consider whether harmonization of processes is appropriate.
10. Funders should seek to establish the short- and long-term costs of data access, and work to determine when cost-recovery is an appropriate model for studies.
11. Funders should jointly consider how best to sustainably support data-sharing infrastructures.

Burton et al, IJE 2017

#ispahDataCam18

Facilitating da

Grant writing/ funding

Box 1. EAGDA recommendations on the governance of data access

1. All project proposals should include data sharing and management plans in funding applications.
2. Funders should support the development of data and metadata standards.
3. Data access processes should be discoverable and transparent for potential data users.
4. Studies should establish proportionate governance mechanisms for data access.
5. Collaboration should not be a necessary condition for data access.
6. Consent should, as far as possible, include provision for further data use beyond the original study.
7. Clear policies should be developed on how depletable resources will be managed.
8. Funders should establish clear penalties and sanctions for breaches of data-sharing rules.
9. Principles of data access should be harmonized as far as possible across studies. Study leaders should also consider whether harmonization of processes is appropriate.
10. Funders should seek to establish the short- and long-term costs of data access, and work to determine when cost-recovery is an appropriate model for studies.
11. Funders should jointly consider how best to sustainably support data-sharing infrastructures.

Burton et al, IJE 2017

#ispahDataCam18

Data harmonisation relies on data sharing

- Opportunities for data sharing need to be fostered and supported throughout the research process

- Supporting data discovery through good data management: the FAIR Guiding Principles.
  - **F**indable,
  - **A**ccessible,
  - **I**nteroperable and
  - **R**eusable

Wilkinson et al, Sci Data 2016

#ispahDataCam18

Box 2 | The FAIR Guiding Principles

To be Findable:
F1. (meta)data are assigned a globally unique and persistent identifier
F2. data are described with rich metadata (defined by R1 below)
F3. metadata clearly and explicitly include the identifier of the data it describes
F4. (meta)data are registered or indexed in a searchable resource

To be Accessible:
A1. (meta)data are retrievable by their identifier using a standardized communications protocol
A1.1 the protocol is open, free, and universally implementable
A1.2 the protocol allows for an authentication and authorization procedure, where necessary
A2. metadata are accessible, even when the data are no longer available

To be Interoperable:
I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
I2. (meta)data use vocabularies that follow FAIR principles
I3. (meta)data include qualified references to other (meta)data

To be Reusable:
R1. meta(data) are richly described with a plurality of accurate and relevant attributes
R1.1. (meta)data are released with a clear and accessible data usage license
R1.2. (meta)data are associated with detailed provenance
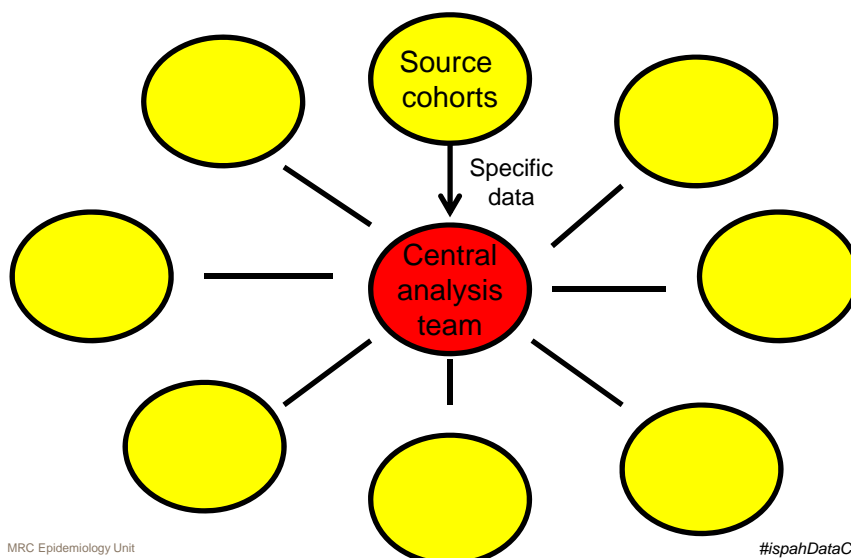R1.3. (meta)data meet domain-relevant community standards

Protocols

Metadata

Wilkinson et al, Sci Data 2016

#ispahDataCam18

Data harmonisation relies on data sharing

- Sharing data comes with legal, ethical, and governance challenges

- Different ways to share data
  - Presentation of four potential models

- Consider different perspectives (funder, researcher, users, participant, ethics committee)

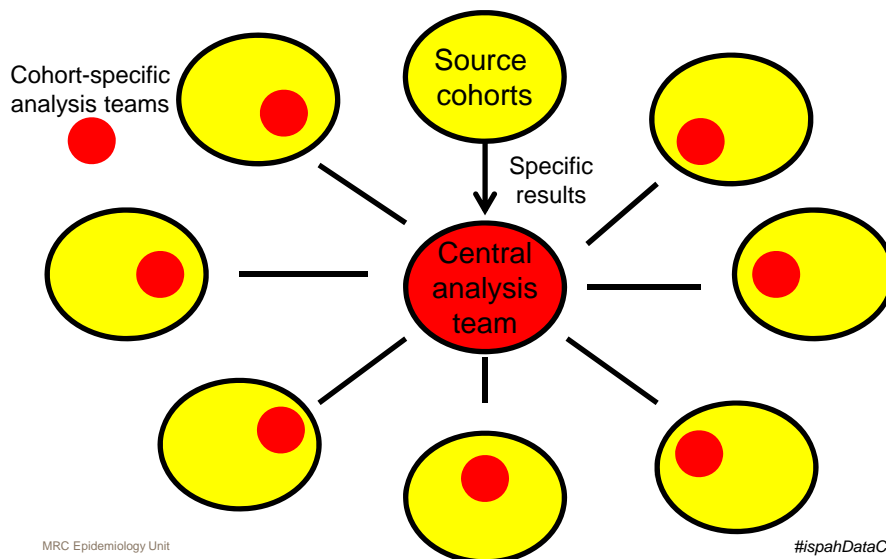# Sharing of data between cohorts using traditional collaboration/consortia agreements

Possible issues

- Considerable transactional burden

- Burden will increase exponentially as number of partners in consortia increases

- Difficult to control passage of data and use beyond the original intention

- Requires well-established collaborative networks

- One-off effort (funding, researchers)

#ispahDataCam18

Ad hoc consortia - sharing of results

Cohort-specific analysis teams

Source cohorts

Specific results

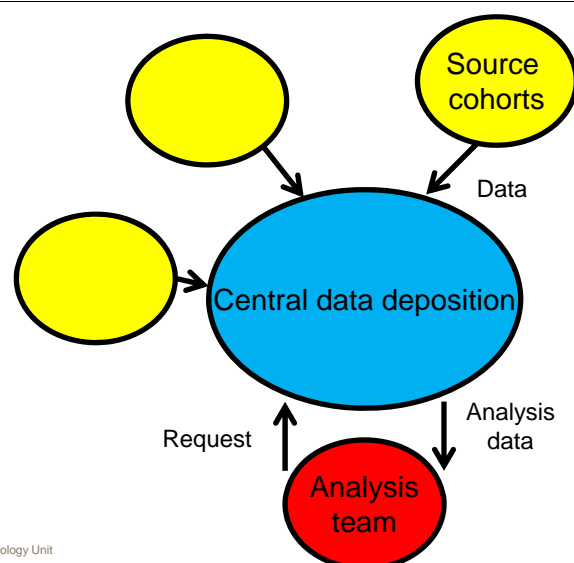Central analysis team

#ispahDataCam18

Possible issues

- Ad hoc consortia allow sharing of RESULTS without administrative or organisational complexity

- Analytical capacity required for cohort-specific analyses – tension with core activities

- Difficulties with data harmonisation given lack of control

- Analysis is potentially missing major between-cohort variation

#ispahDataCam18

Central deposition of data

Possible issues

- Approach works within some countries for some forms of data

- Likelihood of success for between-country collaboration low

- Major governance, ethical and legal challenges (sharing of data for multiple, unspecified analyses)

- Difficult to mandate for historical data

- Generally lack of control for users over data harmonisation decisions

- Resource intensive

MRC Epidemiology Unit

Federated meta-analysis

Non-identifying summary parameters allowed to pass between computers. Individual level data retained on data computer of origin

- Data stays within governance structure of source cohort

- Cohorts focus efforts on preparation of data and IT infrastructure for sharing

- Analytical effort more focused on the scientific–led questions

- Avoids some organisational complexity

MRC Epidemiology Unit

Concluding remarks

- Data need to be Findable, Accessible, Interoperable, Reusable
  - Availability of detailed protocols and metadata key

- Data-sharing considered from early stage research development (where possible)

- Different models of sharing bring different advantages and challenges

- Mix of data sharing models may be needed for optimal use

Suggested reading: literature

- **Atkin** et al. Harmonising data on the correlates of physical activity and sedentary behaviour in young people: Methods and lessons learnt from the international Children's Accelerometry database (ICAD). IJBNPA 2017;14:174
- **Burton** et al. Policies and strategies to facilitate secondary use of research data in the health sciencies. IJE 2017;1729.
- **Gaye** et al. DataSHIELD: taking the analysis to the data, not the data to the analysis. IJE 2014; 1929.
- **Doiron** et al. Data harmonization and federated analysis of population-based studies: the BioSHaRE project. Emerg Themes in Epi, 2013;10:12.
- **Lakerveld** et al. Identifying and sharing data for secondary data analysis of physical activity, sedentary behaviour and their determinants across the life course in Europe: general principles and an example from DEDIPAC. BMJ Open 2017;7.
- **Murtagh** et al. Better governance, better access: practising responsible data sharing in the METADAC governance structure. Hum Gen 2018;12:25.
- **Rumbold & Pierscionek**. A critique of regulation of data science in healthcare research in the European Union. BMC Med Eth 2017;18:27
- **Townend**. Conclusion: harmonisation in genomic and health data sharing for research: and impossible dream? Hum Gen 2018;137:657.
- **Wilkinson** et al. The FAIR Guiding Principles for scientific data management and stewardship. Scient Data 2016;3:160018.

Suggested reading: examples of current initiatives

- BioSHaRE: http://www.bioshare.eu/
- CLOSER: https://www.closer.ac.uk/
- DEDIPAC: https://www.dedipac.eu/
- ICAD: http://www.mrc-epid.cam.ac.uk/research/studies/icad/
- Interconnect: http://www.interconnect-diabetes.eu/
- Maelstrom: https://www.maelstrom-research.org/
- ProPASS: @ProPASSProject

#ispahDataCam18