ECTU
Edinburgh Clinical Trials Unit

# Data Sharing in Clinical Trials

## *Practical guidance on anonymizing trial datasets*

Usher Institute
Population Health Sciences & Informatics

**C Tuck,** S Lewis, G Milne, C Keerie

THE UNIVERSITY *of* EDINBURGH

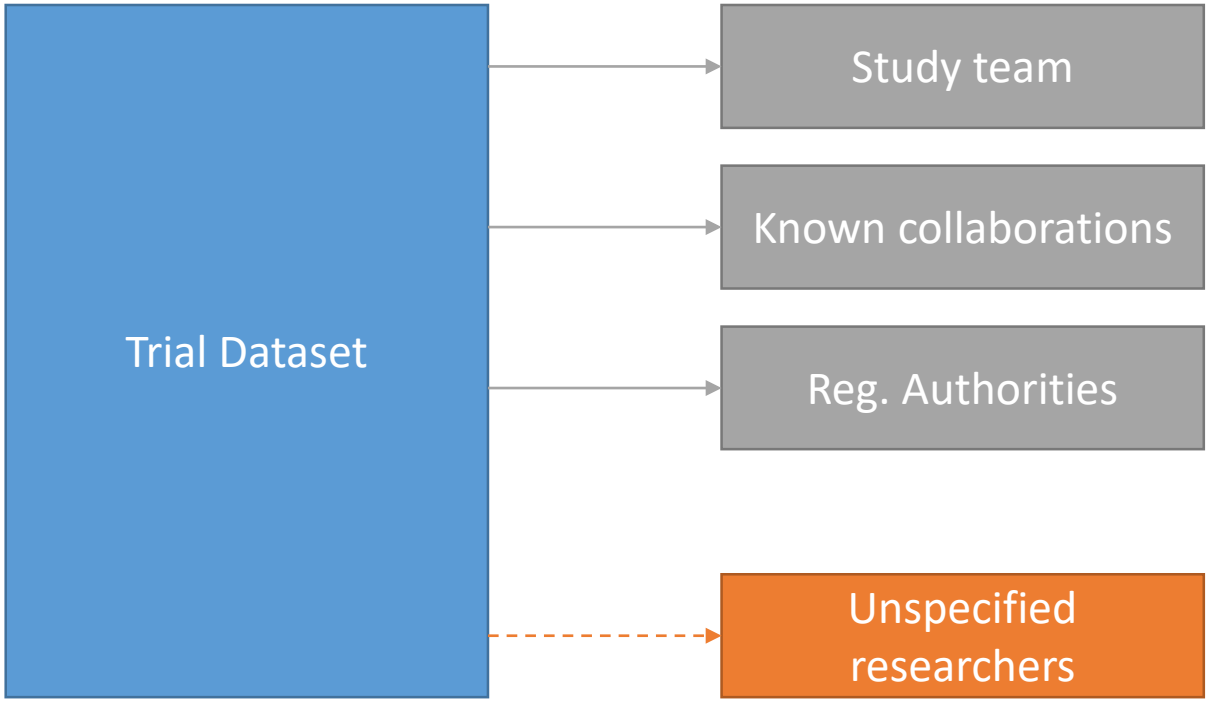
# Today's talk

Usher Institute
Population Health Sciences & Informatics
www.ed.ac.uk/usher
@EdinUniUsher

- Background to data sharing
- Current guidance
- Our work on anonymization
- How we use Datashare

# Data sharing (1)

Usher Institute
Population Health Sciences & Informatics
www.ed.ac.uk/usher
@EdinUniUsher



# Data sharing (2)

Usher Institute
Population Health Sciences & Informatics
www.ed.ac.uk/usher
@EdinUniUsher

- Usually at the end of the study
- Sharing data with unspecified secondary researchers
- Sharing data to the individual participant level

# Anonymization (1)

Usher Institute
Population Health Sciences & Informatics
www.ed.ac.uk/usher
@EdinUniUsher

- Anonymization and deidentification used interchangeably

- “information which does not relate to an identified or identifiable natural person or to personal data rendered anonymous in such a manner that the data subject is not or no longer identifiable” [Recital 26]

- Allows for wider use of information

# Anonymization (2)

Usher Institute
Population Health Sciences & Informatics
www.ed.ac.uk/usher
@EdinUniUsher

| Trial ID | Date of Enrolment | Initials | Age at Enrolment | Postcode | Gender |
|---|---|---|---|---|---|
| | | | 67 | EH1 | M |
| | | | 56 | OX4 | M |
| | | | 72 | IP2 | M |
| | | | 86 | SW1A | F |
| | | | 82 | KT6 | F |
| | | | 79 | EH4 | M |
| | | | 65 | IP3 | F |

The Daily G

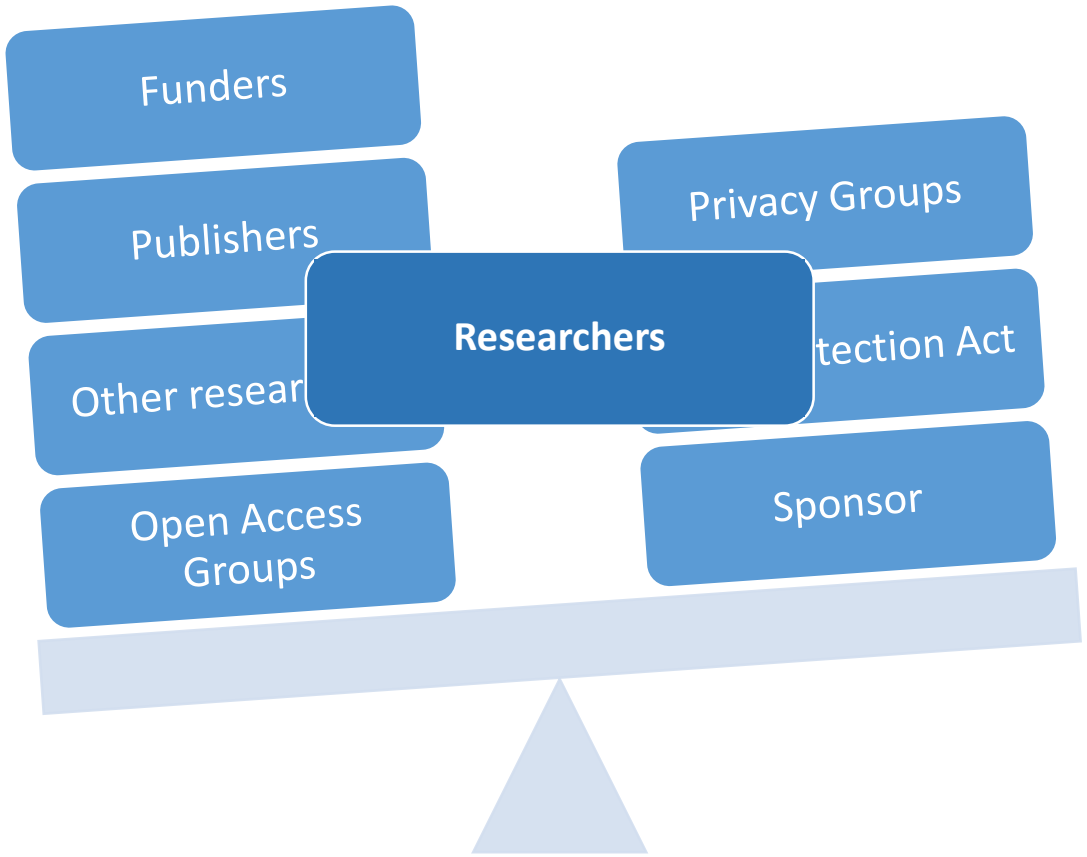
Friday, 18 May 2012

Queen enters clinical trial

Buckingham Palace officials have today confirmed that the Queen entered a clinical trial this week. The shock announcement came as the Royal Family moved to quash rumours of Her Majesty's ill-health

# The balancing act

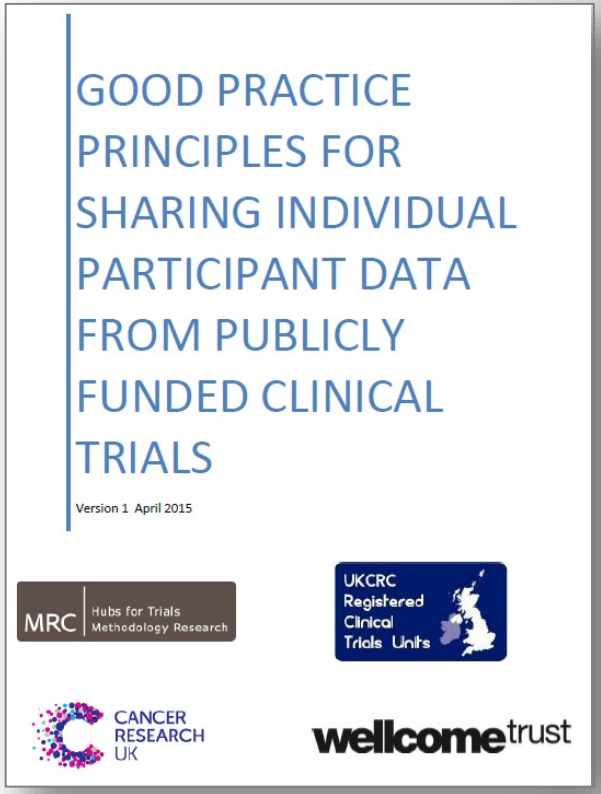
Usher Institute
Population Health Sciences & Informatics
www.ed.ac.uk/usher
@EdinUniUsher



# MRC guidance (1)

Usher Institute
Population Health Sciences & Informatics
www.ed.ac.uk/usher
@EdinUniUsher

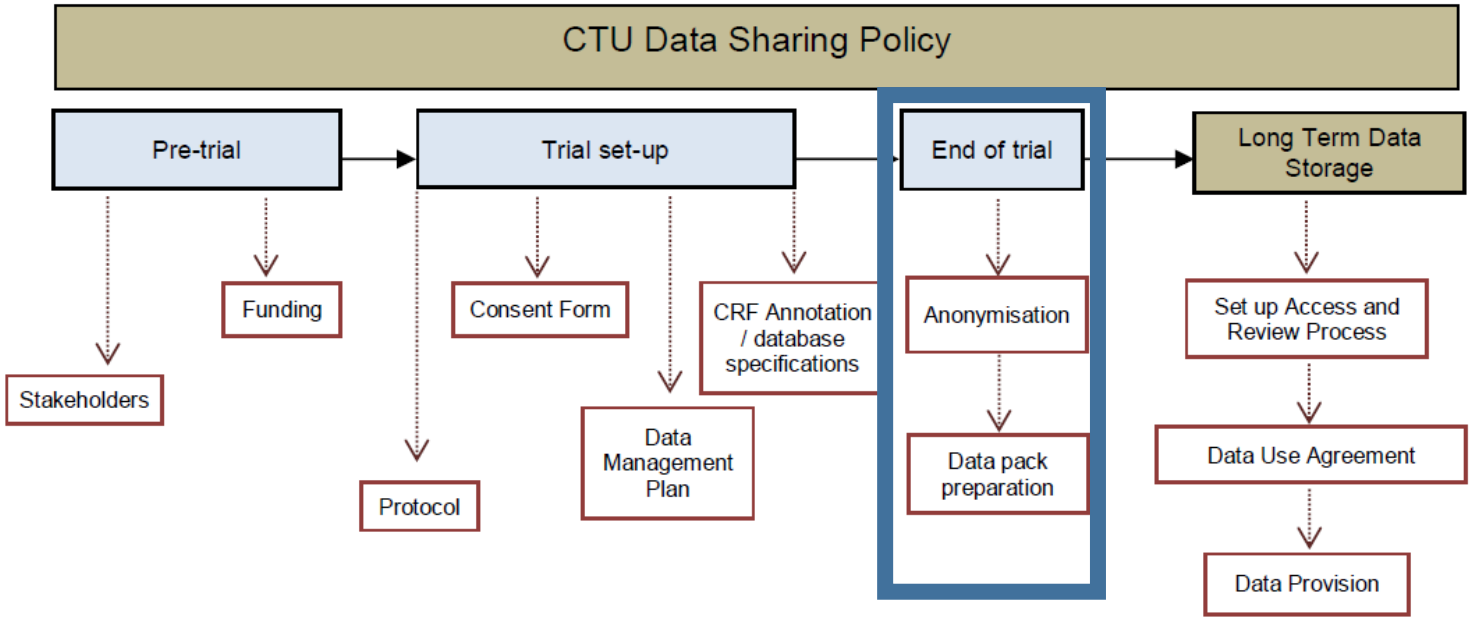
- Aimed at CTUs
- Recommends a controlled access model
- Published April 2015




Good Practice Principles for Sharing Individual Participant Data from Publicly Funded Clinical Trials. Tudur Smith C, Hopkins C, Sydes M, Woolfall K, Clarke M, Murray G, Williamson P. April 2015.

# MRC guidance (2)

Usher Institute
Population Health Sciences & Informatics
www.ed.ac.uk/usher
@EdinUniUsher



# The test case

Usher Institute
Population Health Sciences & Informatics
www.ed.ac.uk/usher
@EdinUniUsher

- TOPPIC trial
- 240 participants

## Mercaptopurine versus placebo to prevent recurrence of Crohn's disease after surgical resection (TOPPIC): a multicentre, double-blind, randomised controlled trial

*Craig Mowat, Ian Arnott, Aiden Cahill, Malcolm Smith, Tariq Ahmad, Sreedhar Subramanian, Simon Travis, John Morris, John Hamlin, Anjan Dhar, Chuka Nwokolo, Cathryn Edwards, Tom Creed, Stuart Bloom, Mohamed Yousif, Linzi Thomas, Simon Campbell, Stephen J Lewis, Shaji Sebastian, Sandip Sen, Simon Lal, Chris Hawkey, Charles Murray, Fraser Cummings, Jason Goh, James O Lindsay, Naila Arebi, Lindsay Potts, Aileen J McKinley, John M Thomson, John A Todd, Mhairi Collie, Malcolm G Dunlop, Ashley Mowat, Daniel R Gaya, Jack Winter, Graham D Naismith, Holly Ennis, Catriona Keerie, Steff Lewis, Robin J Prescott, Nicholas A Kennedy, Jack Satsangi, for the TOPPIC Study Group**

*Lancet Gastroenterol Hepatol* 2016; 1: 273–82

Published Online August 30, 2016 http://dx.doi.org/10.1016/S2468-1253(16)30078-4

### Summary

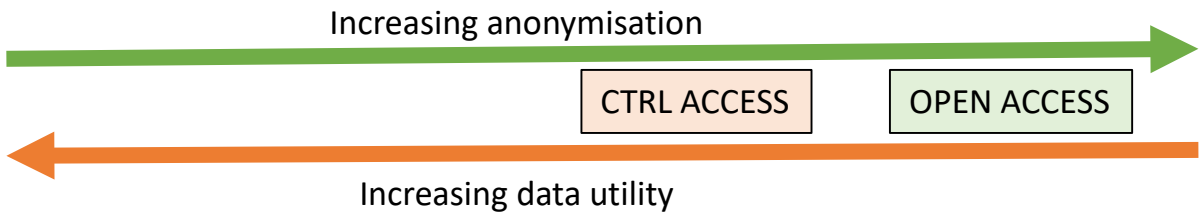
**Background** Up to 60% of patients with Crohn's disease need intestinal resection within the first 10 years of diagnosis, and postoperative recurrence is common. We investigated whether mercaptopurine can prevent or delay postoperative clinical recurrence of Crohn's disease.

**Methods** We did a randomised, placebo-controlled, double-blind trial at 29 UK secondary and tertiary hospitals of patients (aged >16 years in Scotland or >18 years in England and Wales) who had a confirmed diagnosis of Crohn's

# Preparing the dataset

Usher Institute
Population Health Sciences & Informatics
www.ed.ac.uk/usher
@EdinUniUsher

- What to remove?




- 28 participant identifiers (Hrynaszkiewicz & colleagues)
- Both direct and indirect identifiers to consider

# Potential identifiers examples

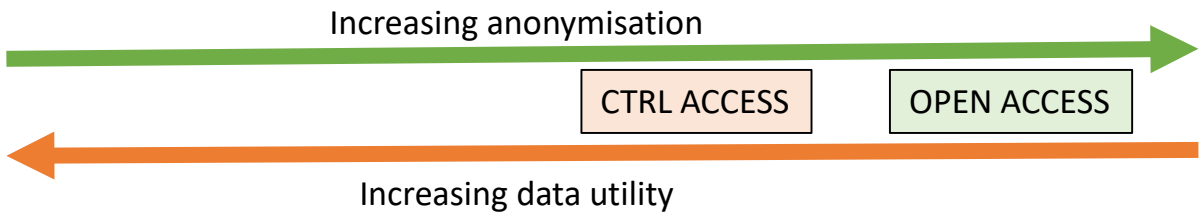
Usher Institute
Population Health Sciences & Informatics
www.ed.ac.uk/usher
@EdinUniUsher

| Direct Identifiers | Indirect Identifiers |
|---|---|
| Name | Place of treatment |
| Initials | Sex |
| Address, including full or partial postal code | Rare disease or treatment |
| Dates related to an individual (inc. date of birth) | Year of birth or age |
| Unique identifying numbers | Small denominators - population size of <100 |
| Medical device identifier | Very small numerators - event counts of <3 |

# Preparing the dataset

Usher Institute
Population Health Sciences & Informatics
www.ed.ac.uk/usher
@EdinUniUsher

- What to remove?




- 28 participant identifiers (Hrynaszkiewicz & colleagues)
- Both direct and indirect identifiers to consider
- Also – remove superfluous data (e.g audit)

# Anonymisation process

Usher Institute
Population Health Sciences & Informatics
www.ed.ac.uk/usher
@EdinUniUsher

**1st Pass**
- Assess data variables & assign value
- Decide on anonymisation method for direct identifiers

**2nd Pass**
- Determine indirect identifier risk
- Decide on anonymisation method for high risk indirect identifiers

**Test**
- Check dataset for successful anonymisation and utility

**Release**

# Anonymisation process 1st pass

Usher Institute
Population Health Sciences & Informatics
www.ed.ac.uk/usher
@EdinUniUsher

- Variables coded and assigned value
- Direct identifiers 01-14 (+15, superfluous)
- Indirect identifiers A-N
- 1st pass of data dictionary
  - Either direct/superfluous, indirect, or not requiring modification
  - Direct identifiers – assign method of anonymization
  - Indirect identifiers flagged

# 1st pass example

Usher Institute
Population Health Sciences & Informatics
www.ed.ac.uk/usher
@EdinUniUsher

| Variable name | Description | Data type | Require anon (Y/N/?) | Reason code | Method |
|---|---|---|---|---|---|
| SubjectNo | Subject Number. | int | Y | 06 | Recode |
| AEDescription | description of the AE | int | ? | C, N | |
| Surgery | | int | ? | C | |
| StartDD | Start Date (day) | varchar | Y | 14 | Study day |
| StartMMM | Start Date (month) | varchar | Y | 14 | Study day |
| StartYYYY | Start Date (year) | varchar | Y | 14 | Study day |
| AECategory ID | | int | N | | |

Reason code 06 = Unique identifying number
Reason code 14 = Dates related to an individual
Reason code C = Rare disease or treatment
Reason code N = Verbatim responses or transcripts

# Study day modification

Usher Institute
Population Health Sciences & Informatics
www.ed.ac.uk/usher
@EdinUniUsher

- Use date of randomisation as day 0
- All other dates relative to date of randomisation
- e.g. date of randomisation 15/01/2014, start date of AE 16/01/2014
- New study date of AE = 1

# Anonymisation process 2nd pass

Usher Institute
Population Health Sciences & Informatics
www.ed.ac.uk/usher
@EdinUniUsher

- Indirect identifiers – decision to anonymise or leave alone
- Currently use a consensus model
- Some summarized to determine risk (event counts)
- May need medical input

# 2nd pass example

Usher Institute
Population Health Sciences & Informatics
www.ed.ac.uk/usher
@EdinUniUsher

| Variable name | Description | Data type | Require anon (Y/N/?) | Reason code | Method |
|---|---|---|---|---|---|
| SubjectNo | Subject Number. | int | Y | 06 | Recode |
| AEDescription | description of the AE | int | ? **Y on 2nd pass** | C, N | |
| Surgery | | int | ? **N on 2nd pass** | C | |
| StartDD | Start Date (day) | varchar | Y | 14 | Study day |
| StartMMM | Start Date (month) | varchar | Y | 14 | Study day |
| StartYYYY | Start Date (year) | varchar | Y | 14 | Study day |
| AECategory ID | | int | N | | |

Reason code 06 = Unique identifying number
Reason code 14 = Dates related to an individual
Reason code C = Rare disease or treatment
Reason code N = Verbatim responses or transcripts

# Dataset release

Usher Institute
Population Health Sciences & Informatics
www.ed.ac.uk/usher
@EdinUniUsher

- Is it anonymous?
  - Motivated intruder test
  - May not be required for every dataset (risk based approach)

- Is it useful?
  - Re-run analysis with modified dataset
  - Further QC checking might be required

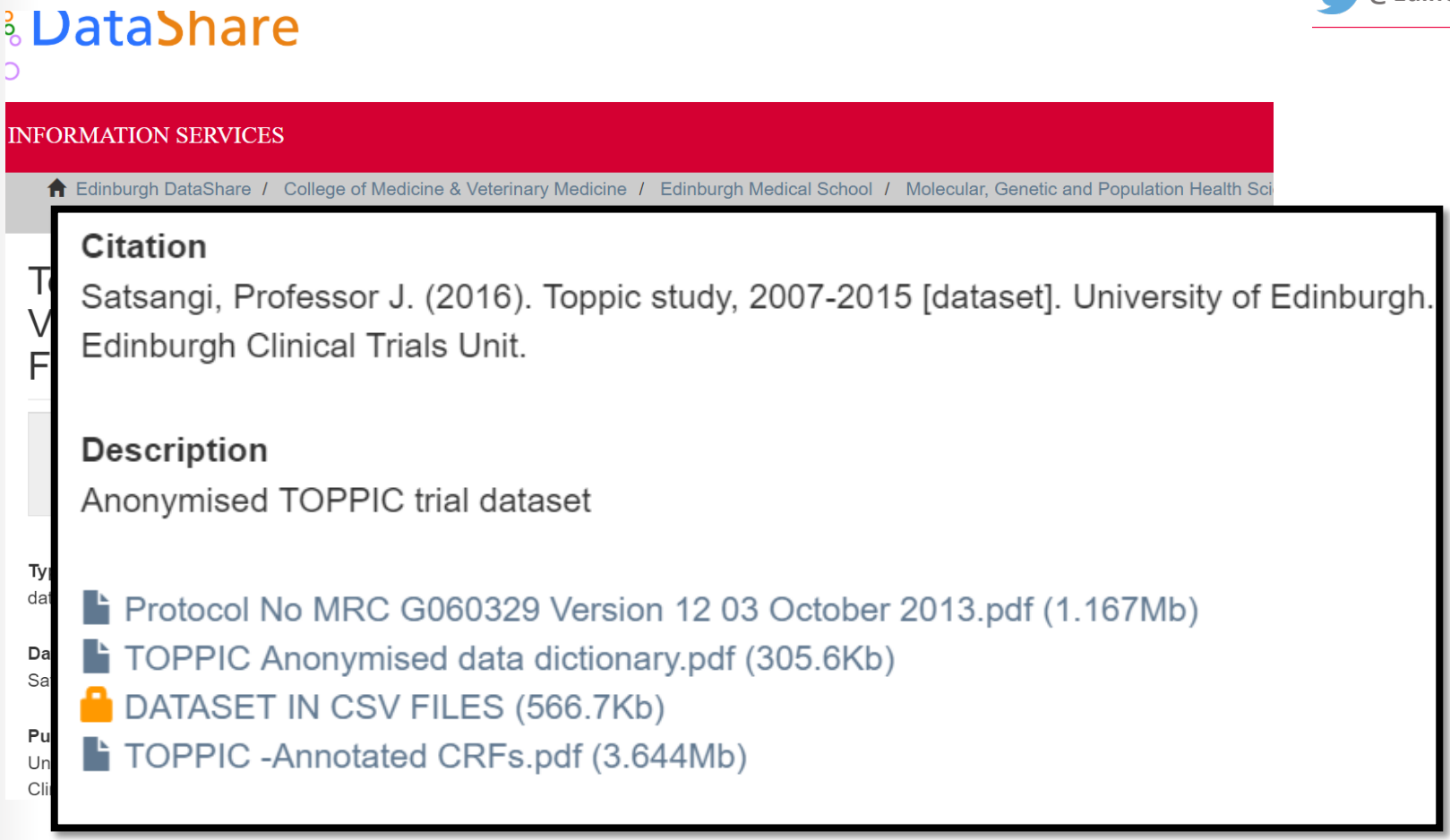
# Summary on anonymization

Usher Institute
Population Health Sciences & Informatics
www.ed.ac.uk/usher
@EdinUniUsher

- Generic rules can be created for direct identifiers
- Decisions on indirect identifiers on a trial-by-trial basis
- Balance anonymisation, data utility and practicality

# Using Datashare

Usher Institute
Population Health Sciences & Informatics
www.ed.ac.uk/usher
@EdinUniUsher



# Application process

Usher Institute
Population Health Sciences & Informatics
www.ed.ac.uk/usher
@EdinUniUsher

**Apply**
- Standard form (credentials, research question)

**Review**
- Committee review (ECTU + Chief Investigator)

**Contract**
- Data Access Agreement

**Release**

# Questions?

Usher Institute
Population Health Sciences & Informatics
www.ed.ac.uk/usher
@EdinUniUsher

- Thanks to:

ECTU
Edinburgh Clinical Trials Unit

TOPPIC
Trial of Prevention of Post-operative Crohn's Disease

Asthma UK Centre for Applied Research

asthma UK