Responsible data sharing

Open Science and personal data

Antonio Schettino, Ph.D.
Coordinator Open Science
Erasmus Research Services
schettino@eur.nl

Erasmus University Rotterdam

slides: *https://osf.io/h3u56/*

OUTLINE

- What is Open Science?
- What is Open Data?
- Responsible Open Data
  - informed consent
  - anonymization
  - restricted access

Antonio Schettino

OPEN SCIENCE: DEFINITION

"Open Science represents a new approach to the scientific process based on cooperative work and new ways of diffusing knowledge by using digital technologies and new collaborative tools. The idea captures a systemic change to the way science and research have been carried out for the last fifty years: shifting from the standard practices of publishing research results in scientific publications towards sharing and using all available knowledge at an earlier stage in the research process".

Open innovation, open science, open to the world (European Commission, 2016)

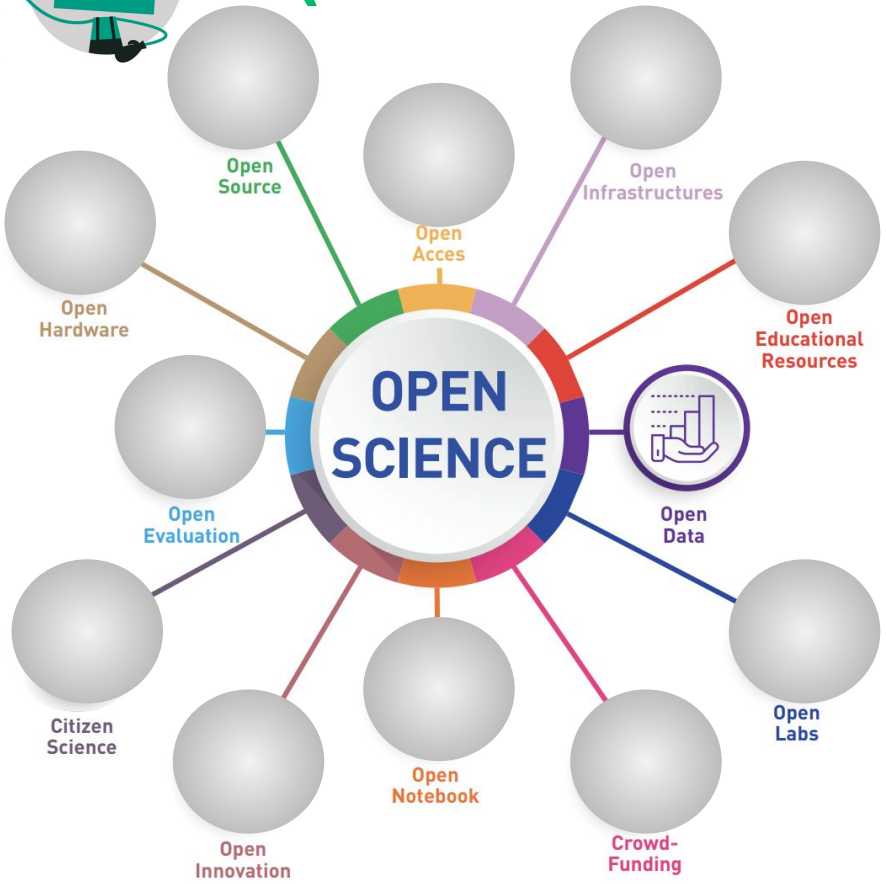
Antonio Schettino

OPEN SCIENCE: TAXONOMY

Towards a UNESCO Recommendation on Open Science (UNESCO, 2020)

Antonio Schettino

(RESPONSIBLE) OPEN DATA



- online publication of data
- available for access, (re-)use, (re-)distribution
- permissive licenses (e.g., *Creative Commons*)
- human- and machine-readable (e.g., *FAIR*)
- example:
  - *COVID-19 in EU/EEA* (*ECDC*)
  - *DANS Easy*
- maximize data value, increase trustworthiness:
  - increase research transparency
  - enable replication (when relevant)
  - explore new research questions
  - reduce participant burden
  - educational opportunities

Towards a UNESCO Recommendation on Open Science (UNESCO, 2020)

INFORMED CONSENT

- participants' views on sharing their own data
  - mostly in favor of data sharing, to *advance science* and *help others*
  - often *assumed or expected*
  - sometimes even *condition for participation*!

- informed consent
  - *enhanced consent forms* effectively improve participant understanding
    - brief, simple language, good readability (e.g., large font size)
    - who will be able to access the data? e.g., researchers, public, …
    - how will the data be treated? e.g., anonymization, restricted access, ...
  - *EUR templates*

ANONYMIZATION

- anonymize personal data
  - any information relating to an identified or identifiable natural person (e.g., name, identification number, IP address, ...)
  - only collect data *necessary* for *current* research questions
- anonymization scheme (created *before* data collection)
  - use pseudonyms instead of actual names
  - aggregate information (e.g., age range, region, …)
  - redact information
  - (sometimes) useful techniques:
    - randomization: add "noise" to increase uncertainty of observations
    - synthetic data: statistical properties similar to original data
- ask your privacy officer (*Annemieke Wiersema*)

RESTRICTED ACCESS

- prerequisite: good data management
  - *FAIR* principles
  - EUR-approved tools & services
  - ask your Data Steward (*Dr. Anna Volkova*)

- restricted access
  - grant access based on clear rules (**no** “data upon request”!)
    - project collaborators
    - editors & reviewers
    - anyone who wishes to reuse data for research purposes
  - *ideally* someone other than researchers can grant access

TIPS

- **plan, plan, plan!**
  - think of open data *before* starting your project
- **collaboration & cooperation**
  - share workload with colleagues
- **ask a professional (data stewards, library, ERS)**
  - advice, guidance, practical help

Antonio Schettino

THANKS FOR YOUR ATTENTION!

Rotterdam

OSCR

Open Science Community Rotterdam

website: *openscience-rotterdam.com*

Twitter: *@OSCRotterdam*

Erasmus

Antonio Schettino