



Course for Doctoral Students

RESEARCH DATA MANAGEMENT AND OPEN DATA

23rd July 2015, Social Science Data Archives,
Faculty of Social Sciences, University of Ljubljana

ECPR Summer School 2015

A large, stylized orange dandelion seed head is positioned in the bottom left corner. Several individual seeds are scattered across the bottom of the slide, some appearing to blow away from the left.

INTRODUCTION TO OPEN SCIENCE: OPEN ACCESS AND OPEN DATA

*Janez Štebe,
Social Science Data Archives*



Open Science Principles

Both theories (research publications) and factual evidence (research data) needs to be openly accessible

- in order to be able to check the validity of published findings (including preventing fraud)
- Support multiple (competitive or cumulative) findings (using different theoretical starting points, different analytical approaches) based on the same data.
- Develop unique new line of inquiry otherwise not possible without access to existing data (some of it from the past) by combining available sources of data



The result is more extensive and more reliable knowledge for the same public investment in research.

- Accessible to fellow scientists (perhaps some of them less lucky to obtain a lot of funds for doing expensive research)
- Accessible to students (learning to use methods on real world data)
- Accessible even to the lay public: Public trust into the scientific findings could be enhanced (if able to cross the barrier of understanding the data without specialists knowledge).



Research data as a public good



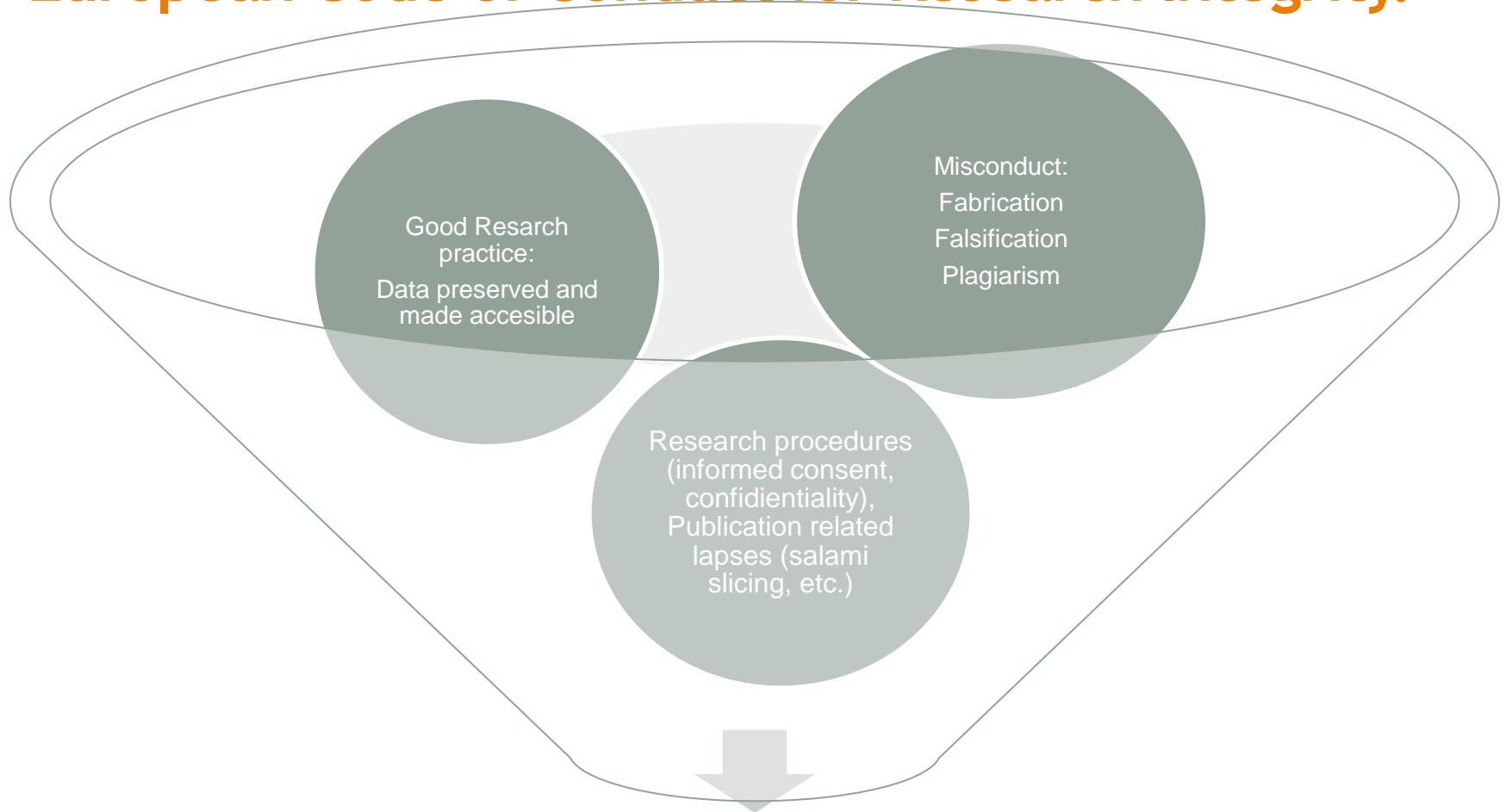
„Publicly funded research data are a **public good**, produced in the public interest, which should be made **openly** available with as few restrictions as possible in a timely and responsible manner.“ (RCUK Common Principles on Data Policy, <http://www.rcuk.ac.uk/research/datapolicy/>)



Who support this?

- ALLEA (All European Academies): Joint Declaration: *"Open Science for the 21st Century"*
- LERU: In December 2013, LERU (League of European Research Universities) produced the [LERU Roadmap for Research Data](#).
- G8: G8 Science Ministers Statement: 3. Open Scientific Research Data
- LIBER: In 2012, the LIBER working group on e-Science formulated [‘Ten recommendations for libraries to get started with research data management’](#). This document has since become one of the most downloaded documents from the LIBER website.
- EU COMMISSION: [Guidelines on Open Access to Scientific Publications and Research Data](#)
- National research funding organisations, research institutions, journals...

European Science Foundation, ALLEA (2011): The European Code of Conduct for Research Integrity.



Research Integrity

What do we mean by data?

- EU Commission H2020 Open Access to research data pilot requirement target object: Raw data and data related to publication
- Raw synonyms: *crude, rough, unfinished, untreated, bare, unprepared, unprocessed ... ?!* (<http://www.collinsdictionary.com/dictionary/english-thesaurus/raw>)

Causes common misunderstanding of the claim...

none of them should apply to proper research data in open access...

- data anonymisation - does this mean that it does not fall under the requirement;
- cleaning (of obvious errors in data),
- processing for easy use (documented, understandable and readable in convenient format)

What do we mean by open?

- EU commission: free, internet (not possible for all forms of data, or practical, or ethical)
- Opt-out reasons: privacy, IPR, security, etc. : not to be read as excuse for relieving of the duty to provide openness:
 - Principle being, to make the data maximally open and usable given the constraints (which needs to be considered proportionate to public benefit)
- Royal Society (Balanced, Qualified/Intelligent Openness) ... not only access, but the context of data retained to enable 'intelligible' reuse... Metadata needed!
„Scientists should communicate the data they collect and the models they create, to allow free and open access, and in ways that are intelligible, assessable and usable for other specialists in the same or linked fields wherever they are in the world. Where data justify it, scientists should make them available in an appropriate data repository.“ (Royal Society (2012): Science as an open enterprise.)

Data specific notion of openness:
Open by default

Licencing

- EU commission recommends using Creative Commons, e.g.



- Does not determine free access over the internet, if other conditions apply. → See further on access arrangements according to the characteristics of data.
- CC0 does not mean, that everything is allowed to do with data:
 - community norms cover some principles, that does not need to be specified in the licence attached; still you are obliged to make an informed judgement about usage of data and, if not unduly unpractical, cite the data in appropriate manner, recommendation being to include a full reference and persistent identifier of access to the version of data used in the references section of published paper.

Research Infrastructure in Social Sciences

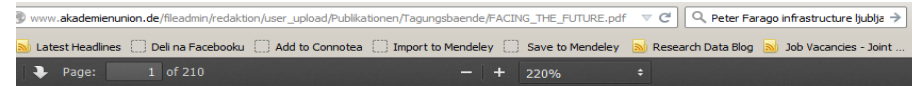
Already conceived in E. K. Scheuch: From a Data Archive to an Infrastructure for the Social Sciences. *International Social Science Journal*, 123, 93- 111.

Contemporary definition:

- ,durable institutions, technical tools and platforms, and/ or services that are put into place for supporting and enhancing research as “public good” resources for the social science community.’ (P. Farago: *Understanding How Research Infrastructures Shape the Social Sciences: Impact, challenges, and outlook*, 2013)

Characteristics:

- User oriented services: data, tools, training, methodological expertise
- Durable and stable on a long-term: cost of infrastructure needs justification in terms of users benefits, including benefits to the society as a whole
- When acquiring new data set: one of the basic criteria for selection is reuse potential that warrant additional processing needed.
- Adaptability to the changing needs of the scientific community



Facing the Future:

European Research Infrastructure for the Humanities and Social Sciences

One example

- “CESSDA provides large scale, integrated and sustainable data services to the social sciences.”

cessda
Consortium of European Social Science Data Archives

Home / National Data Services

Home
About us
CESSDA Services
CESSDA Training
National Data Services
CESSDA Members
CESSDA Observers
Other Data Services

National Data Services

Why should I do make data openly available preferably in a specialist data centre?

- Because funder or journal requires this! (more and more often)
- You can (usually) claim project money to cover additional costs of preparing data and documentation
- It may not necessarily cost you much more: careful planning, avoiding mistakes in study design, data gathering conduct and archiving of the data usually brings high quality data both for own purpose and for sharing and reuse
- Could expect side effects of increased scientific reputation, citations of both data and related original research publications etc.

Who is the enemy of open science?

- Point for a discussion!



Questions?



LEARNING HOW TO ARCHIVE DATA