OPEN SCIENCE AND THE GDPR: CAN'T WE JUST ALL GET ALONG?

JESSICA HRUDEY

FGB RESEARCH DATA OFFICER

research.data.fgb@vu.nl

WHAT IS OPEN DATA? WHAT IS FAIR DATA?

- Open data are openly published so that others can freely re-use/analyze/adapt/share the data
  - Metadata must also be published so that re-users can correctly interpret and utilize the data
- FAIR-principles ≠ open data
  - Guiding principles for data management practices/infrastructure
  - Promotes creation of rich provenance for research data
  - Allows users to assess the usefulness of data for their purposes
  - Facilitate knowledge, discovery and innovation
  - Emphasize need for machine-actionability of data and documentation; help machines find, access and reuse data, with little-to-no user intervention
- Data that are FAIR are:
  - Findable
  - Accessible
  - Interoperable
  - Reusable

WHAT IS OPEN DATA? WHAT IS FAIR DATA?

- NOTE: “Accessible” doesn’t mean EVERYONE has access
  - “As open as possible, as closed as necessary”
  - If there are privacy/IP concerns, its possible to opt-out of open data obligations

    → FAIR principles still apply! (conditions for access should be documented in the metadata)
- Benefits of open and FAIR data
  - Improves transparency and trust in research
  - Improves reliability and validity of research
  - Reduces burden on participants (e.g. patient groups with rare diseases that are frequently asked to participate in research)
  - Promotes collaboration
  - Promotes innovation
  - Promotes big data practices that may lead to more discovery and development
  - Improves research quality as a whole

WHAT IS THE GDPR?

- EU privacy legislation enacted to protect the privacy, rights and freedoms of EU residents
- Many similarities to previous privacy legislation in NL and EU directives
- Main updates/concerns for researchers
  - Enhanced rights of data subjects
  - Higher standards/stricter requirements for informed consent
  - Processing registers
  - Strict requirements for data transfers outside the EU
  - DPIAs
  - Data anonymity
  - Re-use of research data for new purposes

CAN THE GDPR AND OPEN DATA WORK TOGETHER?

- Can participant consent allow for data to be published openly?
  - How much detail can be provided about future re-use of data during the informed consent process?
    - Research purposes are considered ”compatible with the original processing purposes”
    - Researchers are not expected to know nor are they obligated to describe all possible purposes at the outset
    - But researchers and secondary users are obligated to inform participants about new purposes once they are defined→ HOW?
  - Is it possible to obtain valid consent to share data with the whole world?
    - Not possible to define all future purposes, nor explain all privacy risks of all countries
    - Not possible for participants to revoke consent or exercise right of data erasure
    - Once it’s on the Net it’s there forever
    - Are we ok with all of the above?

CAN THE GDPR AND OPEN DATA WORK TOGETHER?

- Ok, can't we just anonymize the data so that the GDPR doesn't apply?
  - Very complex; current EU recommendations for anonymization require:
    - No singling-out
    - No linkability between or within datasets
    - No inferences can be made from the data
- EU requirements seem to require 100% anonymization
  - This is not feasible
  - The more anonymization is applied, the less useful the data is for re-use

CAN THE GDPR AND OPEN DATA WORK TOGETHER?

- EU viewpoint: anything that is distinguishable is an identifier
  - Non-EU experts: if a variable is not reproducible, knowable **and** distinguishable it is not an identifier
- EU viewpoint: if data are more sensitive, or data subjects are more vulnerable, data may be more "desirable", which makes the risk of re-identification HIGHER. The opposite does not apply (????): if data are non-sensitive/data are from non-vulnerable subjects, re-identification risk is not lower, even though the impact of re-identification may be lower
  - Non-EU experts: anonymization involves risk assessment. "Desirability" of the data plays a role in that risk assessment and it is only logical that if the "desirability" is lower the re-identification risk is lower
- Must every dataset containing variables that create distinguishable profiles about participants be treated as non-anonymous/GDPR-applicable forever????

CAN THE GDPR AND OPEN DATA WORK TOGETHER?

- Future technological developments may make data that is currently considered anonymous re-identifiable?
  - E.g. using gait characteristics or EEGs as biometric data
- Do we have to consider all possible future technologies that could allow for re-identification?
- Are there ways we can control and manage the re-use of these data, without having to go through all the GDPR hoops, e.g. terms of use agreements?
  - If we control and manage re-use, are the data no longer open? Will FAIR be enough?

CAN THE GDPR AND OPEN DATA WORK TOGETHER?

- Ok, fine! If it's too much work to anonymize data, why not just publish data (FAIR-ly) in a repository, but with controlled access (the depositor decides who gets to re-use the data)?
  - Who decides who gets access on the long term? Not every department has data managers who can deal with this and many depositors have only temporary positions at institutions
  - Ongoing debate, but essentially restricted access datasets are not open datasets
    - They can still be FAIR! However, infrastructure is needed at institutions and repositories to support creation of FAIR metadata, particularly for documenting conditions for access
  - Informing participants of new uses? How to do this and when do exceptions apply? When does it become information overload for participants?
  - Data sharing agreements (and extra measures for outside EU transfers) are required. Who will deal with this on the long term?
  - Processing agreements need to be in place with the data repository (there isn't one between the VU and OSF)
  - Is restricting access appropriate and necessary for data that are currently considered anonymous, but *maybe* in the future could be re-identifiable? When does it become overkill?

CAN WE GET OPEN DATA AND THE GDPR TO PLAY NICE TOGETHER?

- Plan ahead: obtain consent for data publishing during the informed consent process
  - Inform participants as much as possible, without overwhelming them
    - If you plan to publish data completely openly (no access restrictions), you must emphasize that once data is COMPLETELY open online, there is no deleting it and that the data is accessible by ANYONE in the world! Participants must realize and understand this!
  - FGB, UU and many other parties are working on informed consent guides/templates for open data publishing
  - Updating participants on new uses remains a complex issue
    - Will require support from data repositories/National Co-ordination Point for RDM
    - There are exceptions to informing participants, but the reasoning for not informing participants needs to be documented (e.g. in a Data Management Plan/Data Protection Impact Assessment!)

CAN WE GET OPEN DATA AND THE GDPR TO PLAY NICE TOGETHER?

- Agreements on anonymization methods and standards
  - Education of researchers (what is and isn't anonymous)
  - Trusted Third Parties for carrying out anonymization process
  - National agreements regarding "anonymity" of low-risk data
    - Standardized methods to assess risk
    - Agreements that low-risk is the best we can hope for regarding anonymity
    - Evidence to make anonymization defensible
    - Potentially measures to "control" data re-use; limit re-use to researchers
  - National Co-ordination Point for RDM (LCRDM) has a task group working to get approval of best practices from Dutch Data Protection Authority

CAN WE GET OPEN DATA AND THE GDPR TO PLAY NICE TOGETHER?

- Plans are needed for long-term management of published, restricted access data that cannot be anonymized. Who determines and manages access to and sharing of the data?
  - Departmental data managers?
  - Data access committees (similar to ethics committees)?
  - Responsibility of the department head?
  - FAIR documentation of conditions for access will need to be standard practice

DISCUSSION