Data Practices, Data Management and FAIR Principles: An Educational Scenario by the EnTIRE project

Background

Professor Brown, an epidemiologist working in a public university, won a prestigious grant three years ago to develop an interdisciplinary research consortium that investigates the impact of environmental, genetic and clinical factors on the prevalence of obesity in highly urbanized areas. Professor Brown's multidimensional approach involves collecting data from public databases (Geographic Information System (GIS), Google Street View), surveys and interviews. Furthermore, the approach includes retrospective data collection from hospital records and a study on genetic samples stored in the hospital's biobank. In addition to Professor Brown, the consortium involves 20 scholars, including senior researchers, postdoctoral researchers and doctoral students. Having reached the half-way point of the project, the group has demonstrated significant progress. However, they are going through a very intensive and challenging period of the project. Three datasets using GIS and Google Street View have already been finalized (D1, 2 and 3). During the early phase of the project, a brief commentary was published by members of the consortium on conceptual and methodological issues (M1). They have just submitted a promising manuscript (M2) based on a retrospective study analyzing patient's clinical data. Currently, they are preparing two further manuscripts (M3 and M4)

Issue 1

The GIS and the Google Street View study generated large datasets (D1, 2 and 3) about urban areas that influence the daily level of physical exercise of those living in those areas. A member of the consortium, Mr. Green, plans to send the datasets to a public repository for scholars who wish to use them. He argues that the grant agreement requires the consortium to follow FAIR principles that describe how research outputs should be organized so they can be more easily accessed, understood, exchanged and reused. The FAIR principles

require that research data should be findable, accessible, interoperable and re-usable. Professor Brown criticizes Mr. Green for his plan. Professor Brown argues that the university controls and owns the data and it should stay there. Brown refers to GDPR rules and legal advice regarding the university's property rights. Other senior scholars in the consortium argue that the data should not be transferred prior to the publication of manuscripts M3 and M4 as they worry that the early usage of their dataset by competing research groups will endanger their publication plans.

1a. Questions for Researchers

1. Should we make our data public? Do we have such an obligation?
2. Does the institution where the research took place own the copyright to the data?
1. In case the consortium is considering transferring research data to public depositories, at what stage of the project should this be done?
2. In case we deposit data in public platforms for reuse, can we still claim credit for generating the data? How does data citation work?

1b. Questions for Research Administrators

3. As a research administrator working for Professor Brown's department, and also for the project consortium, you are responsible for a variety tasks related to project budget, communication, managing contracts and issues related to regulatory compliance. The consortium will collect and process a significant amount data, some of which is sensitive and personal. How would you ensure fair data management in such a large project that collects and processes a wide diversity of data?
4. How could you ensure transparency in accessing or making use of project research data and materials?

Issue 2

Dr. White, a postdoctoral researcher, was a member of the consortium for the first three years of the project. Subsequently, she received a tenured teaching position at another university. She is planning to continue some of her research interests that were developed through her involvement with Professor Brown's project. She is planning to download considerable portions of the dataset - that she was responsible for obtaining - so she can continue her research and use them at her new institution. Professor Brown insists that as long as he is the principal investigator, all the data generated by the project belongs to him and his university. Dr. White is disappointed. She believes that Professor Brown's response is unfair and unacceptable. As she still has access to that part of the dataset, she decides to download it.

2. Questions for Researchers

1. Who owns the research data? Dr. White? Professor Brown? The university? The organization that provided the grant?
2. How could the conflict between Dr. White and Professor Brown over access to the datasets have been avoided?
3. Under what ethical conditions is it appropriate to download the data for personal research use?

Issue 3

M2 was submitted to one of the leading journals in the field. This study was mostly based on retrospective data collection from hospital records. As an appendix, the manuscript includes a table containing raw data, a set of clinical data (symptoms,

laboratory data, histological results, outcome) and some further personal details about individual patients (age, sex, residential area, occupation). The editor of the journal was concerned about the use of patient information and the lack of research ethics committee (REC) approval for its use, raising a query with the corresponding author. The corresponding author responded, stating that the data was collected from routine samples and hospital records taken a decade ago. As a result, patient consent was never obtained. The patients were no longer contactable. In addition, there was no REC approval for the use of this data. However, the corresponding author mentioned that the consortium had discussed the use of existing patient data with the institutional REC, which advised that the proposed study was exempt from review. The editor rejects the manuscript, requesting that the authors should contact their REC.

3. Questions for Research Ethics Committees and Research Integrity Offices

1. Are patients identifiable from the information provided in the manuscript?
2. If one removes the age and sex of all patients, would that be sufficient to render the data anonymous?
3. Should we approve the research without the documented consent of patients?

Suggested Resources

ECCRI: The European Code of Conduct for Research Integrity

ICMJ: http://www.icmje.org/recommendations/browse/roles-and-responsibilities/protection-of-research-participants.html

ICMJE: Copyright

FAIR: The FAIR Guiding Principles for scientific data management and stewardship

EDPS opinion on GDPR and data protection

Preparing raw clinical data for publication: guidance for journal editors, authors, and peer reviewers

Anonymization techniques, Opinion 05/2014 on Anonymization Techniques.

Anonymization, practical guidance: Personal Data Protection Commission of Singapore: Guide To Basic Data Anonymization Techniques.

Joint Declaration of Data Citation Principles

Related Scenarios

This scenario has been inspired by the following case studies:

"Data Acquisition and Management" Responsible Conduct of Research, Course at Columbia University, http://ccnmtl.columbia.edu/projects/rcr/rcr_data/q_a/index.html

COPE, Case number 16-05, Data Anonymity https://publicationethics.org/case/data-anonymity

COPE, Case number 06-25, Consideration of publishing raw data, https://publicationethics.org/case/consideration-publishing-raw-data