Legal & Regulatory Restrictions

Background

This category covers situations where the availability of a dataset or part of its content, when posted at a repository and/or related to a research object (e.g. journal publication), raises a concern in relation to disputes over intellectual property or a breach of a national legal framework or international regulations around research practice and/or data handling.

Licenses

When depositing datasets to a repository, the authors apply a license agreement to the data. The license is a legal arrangement that designates what a user is allowed to do with the data. Repositories usually provide a limited set of licenses that authors are required to use or select from. Many use forms of Creative Commons licenses for data (software should be under a separate software license). A CC0 license places the work in the public domain, and the author waives all copyright and related rights. Other Creative Commons licenses (open licenses) designate which uses are allowed and any expectations for those re-using the data (e.g., attribution to the source). Open licenses cannot be revoked once applied.

Copyright law does not generally apply to raw data or factual information, but some types of research products that might be used in a similar way (e.g., images and audiovisual information) can be copyrighted, as can the compilation of the data into databases or collections in some jurisdictions. Submitters should therefore ensure they have the right, or permission, from any rightsholders, to deposit such copyright-protected material.

In certain countries, the university, higher education institution or research organization where the researcher conducts the work is the legal owner of datasets generated from grants to the university or funds by the organization. In university settings, the researcher is the custodian of

the data, and researchers are authorised to make the research data openly available, provided there are no commercial, legal or ethical restrictions. However, the expectations may vary across countries and institutions and researchers should thus check the framework applicable at their setting.

For access to commercial data, researchers usually enter into an agreement with a company. Best practices are that agreements should be documented, reviewed by legal representatives for the researchers' employer, and indicate how and when the researcher may publish results and make the data available under an open license (see https://science.sciencemag.org/content/357/6353/759). Usually, additional reach-through restrictions on the use of deposited data are not allowed and supported by repositories.

Restricted data

Governments pass legislation regarding data protection of their residents, and these have implications on how researchers can share data collected from research in each country, particularly regarding the protection of the identities of patients or participants in human subjects research[1]. Individual countries and jurisdictions have their individual privacy laws and regulations toward protecting identifiable personal data and metadata.

Any breach in regulatory expectations about privacy and ethical conduct of research involving human subjects (e.g., informed consent[2]) incurs a risk to participants in human subjects research, or to individuals or communities who may be identified via the dataset. These concerns and recommendations for handling cases involving such a risk are outlined in the Risk document.

In addition to national laws, there are also international treaties and frameworks by intergovernmental bodies such as the EU, UN, UNESCO, OECD and the WMA. These are generally

[1] The WHO defines research with human subjects as '*any social science, biomedical, behavioural, or epidemiological activity that entails systematic collection or analysis of data with the intent to generate new knowledge, in which human beings:*

- *are exposed to manipulation, intervention, observation, or other interaction with investigators either directly or through alteration of their environment; or*
- *become individually identifiable through investigator's collection, preparation, or use of biological material or medical or other records.*

[2] The Declaration of Helsinki states: '*Participation by individuals capable of giving informed consent as subjects in medical research must be voluntary. [...] In medical research involving human subjects capable of giving informed consent, each potential subject must be adequately informed of the aims, methods, sources of funding, any possible conflicts of interest, institutional affiliations of the researcher, the anticipated benefits and potential risks of the study and the discomfort it may entail, post-study provisions and any other relevant aspects of the study. The potential subject must be informed of the right to refuse to participate in the study or to withdraw consent to participate at any time without reprisal.*'

not legally binding instruments under international law, but instead draw their authority from the degree to which they have been codified in, or influenced, national or regional legislation and regulations. A relevant example is the Declaration of Taipei by the WMAs which focuses on Health Databases and Biobanks. The Declaration of Taipei aims to address any use of health data and specimens and is not restricted to research, so it applies to commercial, administrative and political use of such data.

In addition to national regulatory frameworks, there are agreements developed by relevant stakeholders or communities (research funders, journals and societies) which outline scientific best practice and etiquette agreements for all relevant parties regarding research data. Examples include the Bermuda Principles for the release of DNA sequence data, the Fort Lauderdale Agreement covering pre-publication data sharing of genetic data, the Toronto Statement that widened this to cover other areas of high throughput biology, the Nagoya Protocol on Access and Benefit-sharing which regulates "Access to Genetic Resources and the Fair and Equitable Sharing of Benefits Arising from their Utilization to the Convention on Biological Diversity", and the San Code of Ethics which outlines the values and community approval process expected of researchers intending to engage with the San indigenous communities.

Other local laws may apply to data sets, concerning for example hate speech or the protection of vulnerable populations.

Examples

Cases may arise around breaches to copyright, licenses, or in the context of legal and regulatory frameworks. Additional specific examples are listed under the Resources section.

Licenses

Examples of concerns may arise in relation to licenses include:

- A user breaches copyright or the uses stipulated by the license that applies to the dataset
- An author deposits a dataset where a third party holds rights over the dataset or has not approved public availability
- An author deposits a dataset with an open license that is not compatible with their institution's ownership/requirements for the dataset
- An author has permission to use commercial data for certain purposes but there are restrictions on further use or open deposition that the author or repository does not know or respect

Restricted data

Concerns in relation to breaches of legal or regulatory frameworks may be related to:

- The open sharing of the dataset breaches the privacy laws or other laws of the country where the data was collected
- The requirements for data sharing by a stakeholder/community framework are inconsistent with the regulatory framework in the country where the authors conducted the research
- The open sharing of the dataset breaches institutional biosafety and biosecurity protocols and any similar national or international recommendations relevant to the research field (e.g., policies and guidelines relating to Dual Use Research of Concern).

How cases may arise

Concerns about a dataset may arise in the context of the dataset itself (record at data repository, data paper) or about a research object the dataset underlies (e.g., journal article, preprint, report). If a legal concern arises in relation to a dataset or to an article (e.g., a figure), it is relevant to establish whether there are associated research objects that may also need to be scrutinized, so that the hosts of the outputs can consider whether there are legal concerns in relation to their record.

It is expected that these cases could arise from a variety of stakeholders:

- Repository managers or data curators -- may identify legal concerns as part of the data submission/deposition process
- Editors or reviewers -- may identify legal concerns during a manuscript's peer review or publication process, or may be contacted by readers after publication.
- Data users or producers -- may be contacted by regulators and in turn raise the issue with the data repository or journals
- Data owners, for example academic institutions
- Readers, including human rights or patient groups
- Regulators/government agencies or law enforcement bodies

In a first instance, it is recommended to raise any concerns directly with the author and the host of the dataset (e.g. data repository, journal), rather than via public commentary for example on social media or blogging sites.

Recommendations

If the dataset hosted in the repository is identified as breaching local or national regulations, the repository is likely to be required to take action on the dataset to address the legal breach.

While intergovernmental legislation and legal instruments may apply to the repository, national laws from one country will not apply to another, so if the data publisher and the data submitter are

in different countries then there may not be a legal requirement to act, provided the repository is in alignment with the regulations that apply to its setting.

Data publishers should have clear policies and public terms of service around any legal or regulatory restrictions that may apply, including the types of data subject to specific regulations and community standards. The publisher's policy/terms of use should also include information on how the repository would handle legal challenges and their obligations should a legal breach be identified. While the repository may in some cases not be legally compelled to take action if a concern arises, the policy/terms of use should outline expectations for the authors.

Data submitted to any repository should be in compliance with relevant institutional biosafety and biosecurity protocols and any national or international recommendations relevant to the research field, e.g., the WHO information DURC for life sciences research. Across all research disciplines, submitters should be made aware of dual-use concerns related to their work and take steps to minimise misuse of their work. Where submitted data is deemed to present a potential dual-use risk, the repository may ask submitters to provide details of how such a risk has been mitigated and how it complies with their institutional and funder's requirements, as well as any national regulations. And where guidelines have been breached, the repository should reserve the right to ensure the corresponding data is redacted, removed or retracted.

In relation to issues associated with geographical boundaries in maps, it is recommended that the data repositories, journals and any other data hosts include a public statement indicating that they remain neutral on any jurisdictional claims expressed or implied in published data, manuscript texts, maps and institutional affiliations. As such, the data host would not pursue requests for changes related to jurisdictional claims.

In all the scenarios below, the data publisher that first received the concern (e.g., data repository, journal) should take reasonable steps to establish whether another party (e.g., related journal) should be notified and where necessary, communicate to the other party that an issue has arisen. It may not always be possible for a data repository to establish whether associated objects exist for the dataset, and thus, the author is also responsible for notifying the hosts of objects associated with the dataset. Once a resolution is reached, the data publisher that first received the concern should notify the person raising the concerns.

What actions should be taken if the dataset has not yet been published? Who needs to be involved in this decision?

- Repositories/Journal publishers:
    - Follow up with the author noting the issue. If relevant, note that these type of data cannot be supported without restricted access (if relevant, can provide suggested alternative repositories)

- If applicable, the data publisher may wish to request documentation from the authors regarding the permissions obtained to collect and share the data
- If the author agrees there is a legal issue, they would withdraw the deposition of the dataset/the manuscript
- If the author disputes the legal issue, and the data publisher has ongoing concerns about the dataset in the context of applicable regulations and its policies/terms of use, the data publisher can at this point take the decision not to publish the dataset
- In case of disputes of data ownership or in case of misconduct concerns in relation to datasets, the repository may inform the author's institution
- If necessary, and if this is part of the data publisher's framework for follow up, the data publisher may seek legal advice (ideally working in their jurisdiction) or consult with an expert or with the body who provided permission for the data collection/publication
- The extent of the follow up when a concern is identified (e.g., in relation to documentation requested from the authors or further legal advice) will vary from one data publisher to another, depending on their scope and frameworks for data handling, e.g., whether they provide restricted access to the dataset, whether they can get legal advice within their organization. If such a framework does not exist at the data repository, the repository can take a decision to decline publication of the dataset based on concerns about a breach in its policy/terms of use.

What actions should be taken for a published dataset? Who needs to be involved in this decision?

- Repositories:
  - Follow up with the author, explaining the issue, and pointing to repository policy/terms of use, giving a timeline for when action will be taken on the published dataset. Ask the author for information about research objects that rely upon the dataset.
  - If applicable, the data repository may wish to request documentation for permissions obtained to collect and share the data
  - Data may need to be removed immediately and replaced with an updated version that does not incur legal risk, if that is an option
  - If replacing the dataset with a new version is possible (e.g. removal of metadata or parts of the data to ensure compliance and replacement with de-identified data, addition of controlled access related to geography and/or move data to servers in a different location), the dataset is updated with a new version that complies with legal/regulatory frameworks. The metadata for the data record may need to be updated if any information needs to be removed or replaced to ensure compliance, or if it relates to changes in API access. The repository may also consider posting a comment/note to alert users about any legal considerations related to dataset use.

- If a revised form of the dataset is not possible (e.g., third party who holds ownership does not provide permission to share, request by national regulatory body, the data repository does not handle dataset versioning), consider removing the dataset completely
    - If replacing the dataset with a new version or posting a tombstone page for a removal, the data repository would need to decide whether or not the notification can include a mention/direct readers to the data which incurred a legal breach.
    - If removed, ensure the persistent identifier (e.g DOI) goes to a tombstone page;according to any workflows at the repository for notifications to indexing services, take reasonable steps to notify places where the dataset may be mirrored or aggregated
- If necessary, and if this is part of its framework for follow up, the data repository may seek legal advice (ideally working in their jurisdiction) or consult with an expert or with the body who provided permission for the data collection/publication
- The extent of the follow up when a concern is identified (e.g. in relation to documentation requested from the authors or further legal advice) will vary from one data repository to another, depending on their scope and frameworks for data handling e.g. whether they provide restricted access to the dataset, whether they can get legal advice within their organization. If such a framework does not exist at the data repository, the repository can take a decision to remove the published dataset without such follow up, based on concerns about a breach in its policy/terms of use
- If applicable, and if the information is known, notify the journal(s) that has related manuscript(s) or the relevant academic institution(s)

- <u>Journal publishers:</u>
    - Consider if the data affects the paper, to what extent, and what action should be taken on the article
    - If the published paper is affected, the journal would contact the author and any other relevant party e.g. data repository, seek legal counsel if necessary
    - According to the outcome of the contacts with the authors and any other relevant party, the data publisher takes action as needed
        - If replacing the dataset with a new version is possible, the journal can post a Correction/Notice of republication designating there was a change after publication, and consider whether the notice can refer/link to the original version of the dataset; the journal may need to republish the article to replace the original dataset that represented a legal breach with the updated acceptable dataset version

- If a revised form of the dataset is not possible, the journal should consider removing the dataset completely. This would require a republication of the article to remove the original dataset file, and may also involve removal or redaction of parts of the article content if necessary to ensure regulatory compliance. The journal would need to determine whether the legal concerns with the dataset impact the status of the publication and warrant any action under its publisher policies and/or COPE retraction guidelines, and if so, what type of notification should be issued (e.g. Correction, Expression of Concern, Retraction, removal of the article).

To whom and when does it need to be reported?

If any legal concern is identified regarding the published data, if possible it is worth notifying entities hosting research objects that display or mirror the dataset through automated or manual means, as the information described in the article or other output may also incur a legal risk - although it should be noted that breaching regulations from a particular setting does not in itself imply concerns about the rigor of the work or whether the research is ethically acceptable.

It is recommended that the party which identified the concerns take reasonable steps to notify all parties (platforms) which host research outputs that are known to be associated with the dataset (e.g. journal or other). It may not always be possible for the hosts of the research objects to establish whether associated objects exist for the dataset, and thus, the author is also responsible for notifying the hosts of objects associated with the dataset.

The repository may have a legal obligation to inform the authorities in its own jurisdiction if a legal breach is identified, but no obligation to notify authorities in other jurisdictions.

How should the public be notified?

If an updated version of the dataset has been posted, this can be identified on the landing page per any existing data repository procedures for versions. If geographic/data access restrictions are added to the data record, this should be documented via a notification on the dataset record.

If the dataset is removed, disclosures around data removal should be prepared in a manner that minimizes potential risks that may ensue by drawing attention to the removed data. Flagging any legal challenges can lead to the public finding older versions or recovering downloaded versions of these data, if such a risk exists, the recommendation is to not notify any further than the action taken. In some cases where there is a risk if users are directed to the original record of the dataset, a tombstone page should be displayed that just denotes there was previously a record which is no longer available without explicitly noting the reasons behind the dataset removal.

How do we handle inaction or silence from stakeholders (e.g, the publisher, the authors, the institution)?

The data publisher should consider whether they fall within the jurisdiction of the body which raised the concerns. It should be noted that the authors or the repository may be at risk of prosecution if they receive a legal challenge by law enforcement bodies and fail to take adequate steps.

If the author does not respond and cannot be reached, the repository/journal may need to report the concerns to the relevant research integrity authority; for example the author's institution, funder, or national research integrity office. The repository/journal may also need to take steps to remove the dataset on the basis of the identified legal breach, independent of the author's response.

If the repository/journal has a legal obligation (e.g. are directly affected by national or international law) they may need to report the incident to the appropriate legal authorities. If the repository or publisher has no legal obligation (e.g. are outside the jurisdiction and not directly affected by national law) they should assess if they have an ethical obligation to make any updates to the dataset. If applicable, the data publisher may want to add a note on the data to point out that release or re-use of this data may have legal implications in some jurisdictions but not others.

If a concern is raised to a journal and this does not respond, the matter should be referred to the publisher. In the lack of response by the journal/publisher, the matter may be raised to the author's institution.

Resources

- Regulation like GDPR (the EU General Data Protection Regulation) and other equivalent national legislation such as the Chinese Personal Information Law have implications on protecting identifiable patient data and metadata beyond national level or regional borders
- Dispute regarding access to the National Health Insurance Database for academic use: https://digitalcommons.pace.edu/pilr/vol28/iss1/2
- Implications of GDPR for the data collected as part of the International Genomics of Alzheimer's Project: https://www.sciencemag.org/news/2019/11/european-data-law-impeding-studies-diabetes-and-alzheimer-s-researchers-warn
- Sharing of Liver cancer genomes sequenced in Hong Kong for the Asian Cancer Research Group by the EBI in the non-controlled access ENA repository, following the legislation of the country where the data was collected. The International Cancer Genome Consortium

refused to include the data because it didn't meet US regulations and their policies on controlled data access.

- Requests for retraction for publications involving data originating from samples removed and collected without authorization at the Alder Hey Children's Hospital: https://www.nature.com/news/2011/110816/full/476263a.html
- Withdrawal by the Chinese authorities of the licenses granted to the collaborative projects Comparative Genetic Study of Psychosis in Han Chinese (UCLA and Shanghai Jiaotong University) and CONVERGE Genetic Foundation of Depression in Chinese Women (Oxford University and Peking University) following an update to their Regulation on Human Genetic Resources. The decision by Chinese authorities conflicts with international consensus on genomic data sharing per the Bermuda Principles and the Fort Lauderdale Agreement. https://www.nature.com/articles/d41586-018-07222-2
- Requirements by some countries on how to display map information, in the context of internationally disputed geographical boundaries: https://wwimesw.thighereducation.com/news/journal-articles-tacitly-support-china-territory-grab
- Data sharing restrictions on SARS-Co-V2 imposed by a national body, which conflict with internationally recognised expectation e.g. under the WHO Global Influenza Surveillance and Response System: https://www.scmp.com/news/china/society/article/3052966/chinese-laboratory-first-shared-coronavirus-genome-world-ordered
- Example of a publisher's statement regarding neutrality toward jurisdictional claims in maps and institutional affiliations: https://www.nature.com/srep/journal-policies/editorial-policies#submission-policies