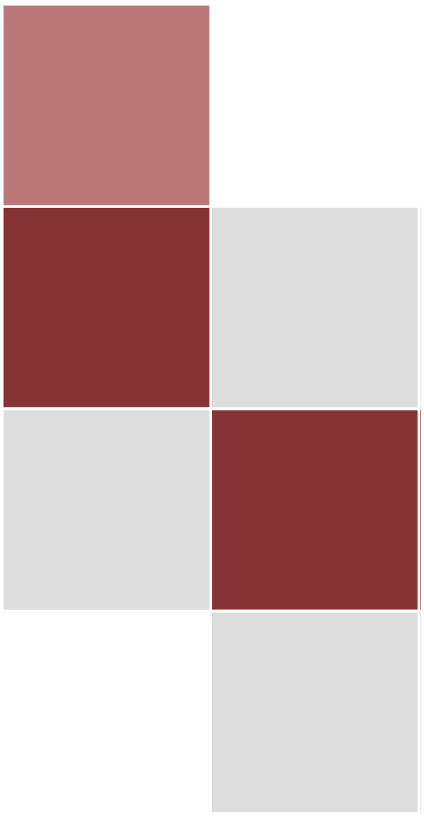
CGIAR

ILRI

INTERNATIONAL LIVESTOCK RESEARCH INSTITUTE



ILRI policy procedure on disclosure of confidential research data

February 2020

ILRI policy procedure on disclosure of confidential research data

Authors:	▪ Jane Poole (Head of Research Methods Group) ▪ Reviewed by IREC members and Legal	Board committee (if applicable)	
		IMC approval date	
Lead manager responsible for operational delivery:	▪ Head of Research Methods Group	Version:	1.0
		Effective date:	O1 February 2020

Revision history

Version No.	Effective date	Approved by and date	Summary of changes	Next scheduled review
1.0	O1 February 2020	ILRI IMC 22 January 2020		

Related documents

ILRI policy(ies)	Research Compliance: ILRI's open-access and data sharing policies Data protection and privacy policy
Global framework	
CGIAR framework/policy	CGIAR open-access and data management policy
ILRI procedures	IREC obtaining prior informed consent in research activities
Other relevant documents	ILRI non-disclosure agreement
Appendix	

Contents

Section 1: Executive summary

1.1 Context and purpose

These policy procedures are intended to help ILRI staff implement our research compliance responsibilities with respect to the disclosure of confidential research data collected within ILRI research activities.

The protection of confidential research data, most commonly personally identifiable information (PII), is an ethical and legal requirement for anyone collecting or using primary research data involving human participants. ILRI's values have at their core the respect and protection of research participants, most of which are vulnerable communities in poor countries. As such, ILRI takes professional and ethical practice very seriously when it comes to the protection and sharing of participant research data.

These procedures provide guidance to ILRI staff on how to protect these data and the conditions under which data may be shared.

1.2 Scope

These policy procedures provides guidelines to ILRI staff on sharing confidential data collected within ILRI research activities (this excludes data from ILRI development activities, which is instead covered by ILRI's Data protection and privacy policy). These research activities must be approved through ILRI's research compliance process and its relevant committee(s): Institutional Research Ethics Committee (IREC), Institutional Animal Care and Use Committee (IACUC) and/or the Institutional Biosafety Committee (IBC), and hence these procedures accompany ILRI's research compliance policy.

The sharing of non-confidential data follows the CGIAR open-access and data management policy and ILRI's own open-access and data sharing policies. A brief mention is made at the end of these procedures of our obligations to open-access as a reminder to staff.

It is the responsibility of both ILRI as an institute, our staff and research partners to ensure that no confidential research data, as defined above, is accessible or shared outside of ILRI without appropriate conditions as described in

Section 2: The procedures.

Refer to What are confidential research data? In Section 3.1 for the definition of confidential research data.

1.3 Primary and other audiences

The primary audience for this document is all ILRI staff involved in research activities, including research technicians and officers, scientists and program management staff. It is also recommended that partners with whom we have contractual agreements are aware of these procedures and may follow them, in the absence of equivalent procedures in their own institutes, where data are co-owned.

ILRI staff who collect primary data are 'custodians' of the data and in many cases this responsibility will sit with the Principal Investigator for a project. ILRI, and partners where appropriate (as per the Contractual Research Agreement, CRA), own the data. The guidance in these procedures is for all custodians and the responsibility for applying the guidance sits with our custodians.

1.4 Key changes since last version

N/A

1.5 Summary of key country/regional-specific differences in application

There are no country or regional-specific differences in application of these procedures.

Section 2: The procedures

"Rules" for sharing confidential research data

ILRI has a process of signing a non-disclosure agreement (NDA) for release of confidential data when used for research purposes only. **Note that confidential data cannot be shared outside of a project and its partners** (who have agreements which detail the sharing of confidential data) **for non-research purposes.**

The following conditions must be met before confidential research data may be released:

- Confidential data can only be shared upon the recipient's institution completing the non-disclosure agreement; confidential data may not be shared with an individual who has no affiliation to an institution (or a national IRB approval[1]) due to the inherent risks.
- Confidential data shall only be shared for research purposes or where there is a legal requirement to notify authorities, such as with communicable disease outbreaks;
- Confidential data should only be shared if its use by others was declared on the initial consent form (if data from humans);
- Confidential data can only be shared if ethics approval for use of the data has been obtained from a relevant Institutional Review Board (IRB) and submitted to ILRI by the recipient. Note that it would be very rare for an individual to have obtained this approval if not affiliated to an institute, except possibly instances of national IRB approval;
- Only confidential data that is required and explicitly requested will be shared (e.g. for spatial analysis don't need the name but do need the GPS coordinates);
- If, for any reason, ILRI requests the recipient to make non-confidential data publicly available (open-access), the recipient must ensure they separate the confidential information from other data relating to an individual, i.e. these publicly available datasets should not show the confidential information: the names of individuals or geographically specific location information unless they are coded (e.g. household ID, "district" shown but village coded etc.);
- Publications authored by the recipient and based on data relating to individuals should ensure that an individual cannot be identified from the publication (e.g. a map could show areas for a survey but not individual points showing households, tables show summaries of individuals at an aggregated level etc.); and
- Confidential data may not be shared, even for research purposes, if prohibited by donor or partner level agreements.

The conditions above are all, except the first and the last, included within the ILRI ILRI non-disclosure agreement form.

Extreme caution must be given to sharing personal data where the data use by the recipient includes: 'as a sampling frame' and where participants may be revisited. For this the NDA should be clear that any follow-up visit must request new consent from participants and also respect previous interactions. The data custodian and legal team may refuse the NDA application if not satisfied that this issue has been resolved.

[1] It would be exceptional if an applicant has national IRB approval but is not affiliated to an institution.

Refer to Completing the non-disclosure agreement (NDA) for ILRI and How to keep confidential research data confidential in Section 3.2 for guidance on how to implement these "rules".

Obtaining informed consent for data sharing

During the consent process with potential research participants, scientists can ensure they are protecting confidentiality of participants while still providing the potential for additional valuable analyses by sharing the data with the broader research community. It is crucial that research participants understand what confidential data is being collected and whether or not the confidential information may be shared outside of the project.

Consent forms should be explicit in **what** information will be kept confidential and **how**. They should also be clear on what information will be shared or published. Refer to Example informed consent statements for protection of confidential data in Section 3.2 describing the protection of confidentiality while allowing both non-confidential, and in some cases confidential, data sharing.

It is recommended to establish a term or phrase that refers to the confidential information (e.g. personal information, information that will identify you etc.). Use another term or phrase that refers to the non-confidential information (e.g. anonymised answers, coded information etc.). Use these terms consistently in the consent form to avoid confusion and be particularly careful to ensure appropriate translation when needed. Ensure that relevant conditions detailed in Obtaining informed consent for data sharing above are also included in the consent form, e.g. including the option for confidential data to be shared for use by others.

Open-access of non-confidential research data

All non-confidential research data collected by ILRI staff and project partners **should be made** open-access according to the CGIAR open-access and data management policy and ILRI's own open-access and data sharing policies. At ILRI we have an open-access data portal (www.ilri.org/portal) and most of our project partners also have their open data repositories.

Therefore, ILRI scientists should never need to share directly non-confidential data from our projects outside of ILRI and our project partners. Any requests should direct the requestor to the relevant data portal or repository!

If you are responsible for research data and it should be made open-access, please contact the ILRI Research Methods Group (or project partner contacts if ILRI is not the lead centre) for any guidance on making non-confidential data open-access.

Section 3: Appendix

3.1 Definitions

What are confidential research data?

Confidential research data means: location (e.g. GPS coordinates, village name), name (e.g. of people or businesses) and ethically sensitive[2] data (e.g. religion, health status). These data are mostly (except for the ethically sensitive information) personally identifiable information - those which allow a link to be made between an individual and the individual's data (e.g. responses to a survey) thereby providing trace back to an individual's identity and/or location. Note that higher level geographic information, for example District names, may be considered non-confidential data and increases the value of the data for further investigation and analyses. The collection of these data is regulated by the ILRI IREC which approves all research activities involving humans.

Confidential research data may also include data from on-station trials (e.g. results of vaccine trials) which may be stipulated as private through clauses in contract agreements or patent applications. The collection of these data is regulated by the ILRI IACUC which approves all research activities involving animals.

Personally identifiable information (PII)

Personally identifiable information (PII) is any data that could potentially identify a specific individual. Any information that can be used to distinguish one person from another and can be used for de-anonymizing anonymous data can be considered PII.

3.2 Tools and forms

How to keep confidential research data confidential

Confidential information needs special care once it is in the hands of a researcher. While certain research projects require by design to collect some confidential data, many projects do not need to collect certain confidential information, so always question "what data do I really need?" For example, do you really need the full name of the participant? Do you really need to know their religion? I**f you don't need it, then it is better you don't collect it.**

To enable confidential data to be locked or removed from datasets we first have to ensure that we can still link data together for analysis (e.g. household information with animal performance, or simply data from different tables in a questionnaire). The best way to do this is to create a Unique Identifier (ID) at the start of any study – this ID may be for an individual, a household, an animal or can be a combination of more than one of these (e.g. Household ID with additional number or letter for the individual in the household). The unique ID could be numeric or alphanumeric and relate to the design of the study (e.g. ABCD – A = District, B = Village, CD = Household number) or be random (e.g. 0001, 0002, 0003,…0546).

For small administrative area (e.g. villages) where households are few (e.g. < 100-200 households) it is advisable to anonymize the names by using village codes to prevent identification of households from their measured characteristics. These should be coded at the start of the study and in all documentation (e.g. survey training manuals) to enable open-access of the code but not the name.

[2] Detailed in the ILRI research compliance form (valid 2017)

Geographic position systems (GPS) coordinates can be altered, for example by taking all household points as the centre of a village, by adjusting the points by a random longitude and latitude or rounding to 1 decimal place (approx. 10 x 10km area), to maintain the value of the data for further spatial analysis while maintaining confidentiality. These "altered" data may also be considered non-confidential, as it will not allow identification of a specific individual.

With electronic datasets (e.g. databases, excel spreadsheets) the easiest way to ensure you do not share confidential data is to create a clean version of the data with the unique ID and any coded variables but without the confidential information. For the full dataset it is not possible to separate the unique ID from the confidential data; therefore to ensure confidentiality these data need to be protected, for example by password protected files and computers. For datasets on a database server in SQL it is possible to lock confidential fields so the database can be viewed with or without the confidential data and the non-confidential data only shared. For Excel and other spreadsheets it is not possible to do this and therefore the full dataset needs to be well protected (e.g. with a password). Staff can ask for support from ICT or the Research Methods Group to ensure their full datasets are appropriately protected.

It is recommended to keep the consent forms as paper and separate from the electronic dataset and/or the paper data collection sheets. It is likely that the consent form also includes the Unique ID – in this case the paper consent form must be kept well protected, in a locked cabinet. If data are collected on paper then it is recommended to transfer to computer quickly and then destroy the paper (or keep in locked cabinet) after data cleaning as, in the long long-term, it is easier to protect the electronic copy of the data and takes up less space!

Biological samples are often collected in ILRI research from animals and/or humans and these are usually maintained for long periods of time. The same standards apply to ensuring that the confidential data (e.g. names linked to the sample) are kept protected and separate from the biological sample. At ILRI, biological samples can be registered in the biorepository with identification codes. The identification code together with other confidential and non-confidential data should then be kept in a separate database with confidential fields protected.

Example informed consent statements for protection of confidential data

Here are some example statements of informed consent that described the protection of confidentiality while allowing both non-confidential and in some cases confidential data sharing:

- Study staff will protect your personal, directly identifying, information closely so no one will be able to connect your responses to this survey to your personal information. National / state laws may require us to show information, not including your personal information, to university or government officials (or sponsors), who are responsible for monitoring the safety of this study. Your personal information (e.g. names, addresses) will be kept safeguarded by password protecting electronic databases and locking cabinets containing paper copies of data. Your name or other personal information that will identify you will not appear in any publication from this study. (Non-confidential only sharing)
- The information in this study will be used only for research purposes, which may include this project and others, and published in ways that will not reveal who you are. (Non-confidential and confidential sharing)

- The data you provide will be used for research purposes only by the project team and other researchers. No data that can identify you directly, like your name, will be shared publicly. (Non-confidential and confidential sharing)
- Your identifying information will be replaced with codes. Only the research team will have access to information that identifies you to carry out this research study. Your identifying information will not be shared with others outside this research study. (Non-confidential only sharing)
- All personally identifying information collected about you will be destroyed once it is no longer needed for the study. Any personal information that could identify you will be removed or changed before files are shared with other researchers or results are made public. (Non-confidential only sharing)
- During the project, information from this study will be kept in locked files that only the research staff can open. Any personal information that could identify you will be removed or changed before files are shared with other researchers or results are made public. (Non-confidential only sharing)

Completing the non-disclosure agreement (NDA)[3] for ILRI

If you receive an application from non-ILRI or non-project partners (this includes ILRI students not linked to the project) for confidential data then follow the below process to ensure that these data are not released without the conditions described above being satisfied. **ILRI staff and students CANNOT share confidential data without following this procedure.**

1) In most cases, ILRI program managers (PM, or delegate) coordinate this process.
2) Draft NDA forms are completed by the applicant and sent to ILRI PM to process through Steps 3–5.
3) The ILRI data custodians, usually the Principal Investigator of the project which collected the data, must first approve release of the data.
4) The program leader (or equivalent) under which the data fall provides approval.
5) The form is submitted to the Legal and IP team to confirm all conditions have been met (e.g. consent to use the data).
6) The form is sent back to the applicant for signature and returned to PM.
7) The Legal and IP team submit the signed form to the DDG for signature and then send the fully-signed form back to the PM and RMG if the data release needs to come from the ILRI Data Portal (http://data.ilri.org/portal).

Useful references:

Cambridge University – Research Integrity
ICPSR Data Management & Curation
U.K. Data Service
World Health Organisation

[3] Live link to https://cgspace.cgiar.org/handle/10568/49610