INDEPTH DATA ACCESS AND SHARING POLICY

The data access and sharing policy sets out the specific procedures (See Annex 1) and prescribes access levels related to various categories of data covered under the policy.

A. **Centre-Specific Data**: When such data are contributed to Network activities they will be subject to the terms of this policy. However, sovereignty over use of the data for centre-specific analyses remains with Member Centres.

B. **Network Core Data**

   a. *Core minimum micro-data sets* will be submitted by all member centres to the iSHARE technical team, who will conduct quality review and disclosure risk assessment on the data in consultation with member centres. Following approval by INDEPTH the data will be immediately released for network use and within 12 months will be made available for public use through licensed access on the INDEPTH Data Repository.

   b. *Aggregated data* will be prepared annually by the designated technical team and released after review by a designated subcommittee of the INDEPTH Scientific Advisory Committee (SAC) and subsequent approval by the Executive Director. This data will be shared through open access on the INDEPTH Data Repository as soon as the approval process is completed.

C. **Network Project Data**

   a. Secondary data will be prepared and quality assured by the working group responsible for the specific Network activity with the involvement of the iSHARE technical team. The working group will report to the INDEPTH Secretariat when the data are available (documented, cleaned, in standard format, and locked for analysis). Within 12 months of data availability or as soon as the primary output from the working group has been submitted for publication (if earlier than 12 months after data availability), the Executive Director will be informed of the readiness of the data for sharing. Following approval by the Executive Director the data will be made available for public use through licensed access on the INDEPTH Data Repository.

   b. All Network activities involving primary data collection should ensure that they include a data sharing plan that is consistent with this policy and with the stated commitment of all centres involved. This plan will be reviewed as part of the proposal approval process. Projects will be required to share data. In general data should be shared for public use within 12 months of the completion of data collection, cleaning, and validation through licensed access on the INDEPTH Data Repository. Where specific circumstances require more restrictive access, this should be motivated to the Executive Director for consideration.

D. **Third Party Data.** These data are subject to the terms of this policy unless otherwise determined through prior agreement.
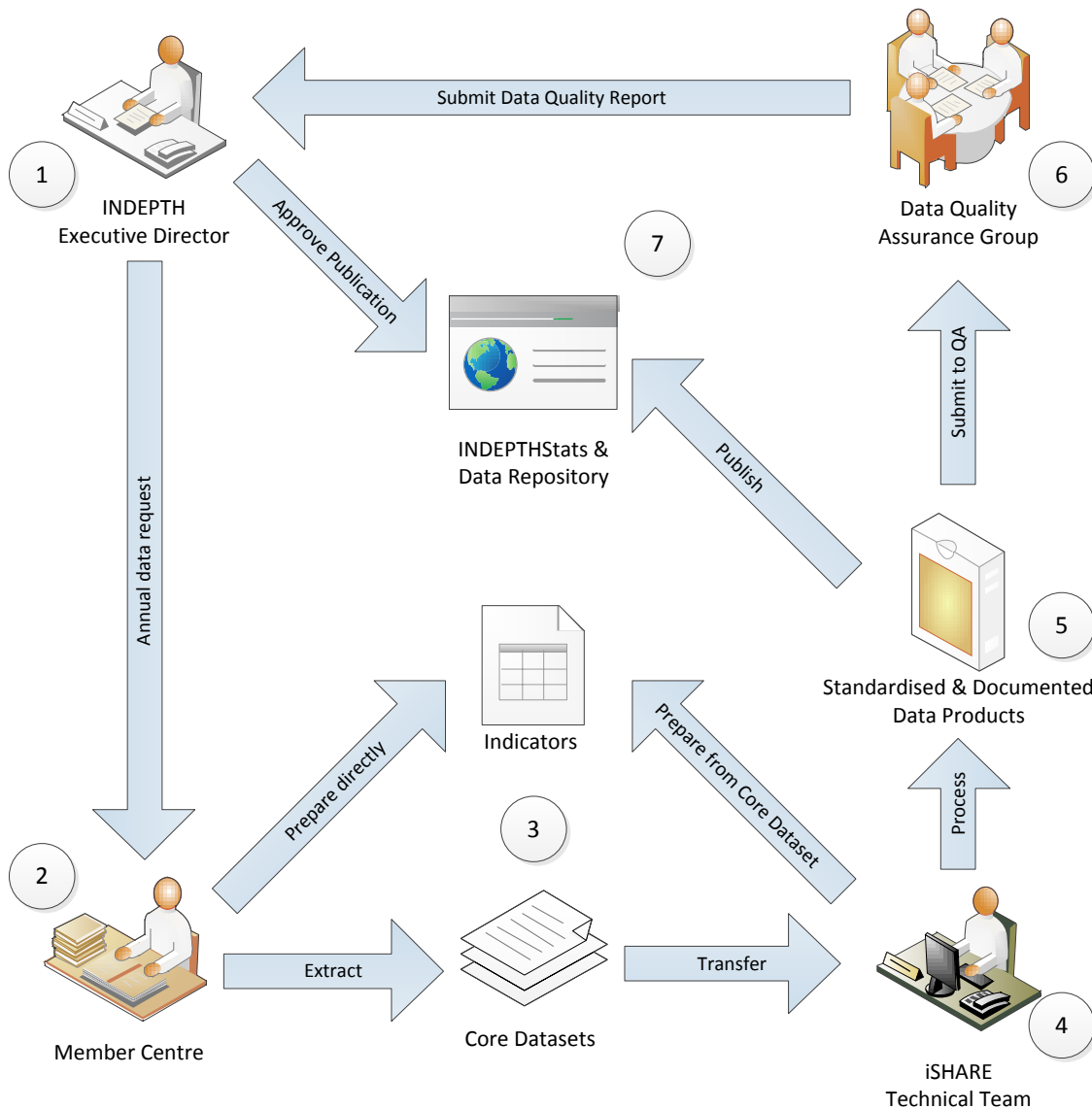
Annex 1 : Process Diagrams

Figure 1 : Network Core Data Process

1. INDEPTH Executive Director (ED) issues a call to Network Member Centres to submit data in terms of agreed Network membership criteria.
2. INDEPTH Member Centre responds to call for data.
3. Member Centre extracts and transfers core minimum dataset to iSHARE technical team. For iSHARE members this process is facilitated through the Centre-in-a-Box (CiB) system. The Centre may prepare indicator (aggregated) data directly or preferably indicator data are prepared on their behalf from their core minimum dataset.
4. iSHARE Technical team processes, standardises and documents data received from Centres, and
5. Produce standardised and documented data products (data sets that are anonymised, uniquely identified and attributed to member centre of origin and the INDEPTH Network).
6. Prior to publication all data products are submitted to the Data Quality Assurance Group for assessment. This group submits a report to the ED with recommendations.
7. Based on the recommendations from the Data Quality Assurance Group, the ED instructs iSHARE Technical Group to publish the data product/s on the INDEPTHStats or Data Repository.
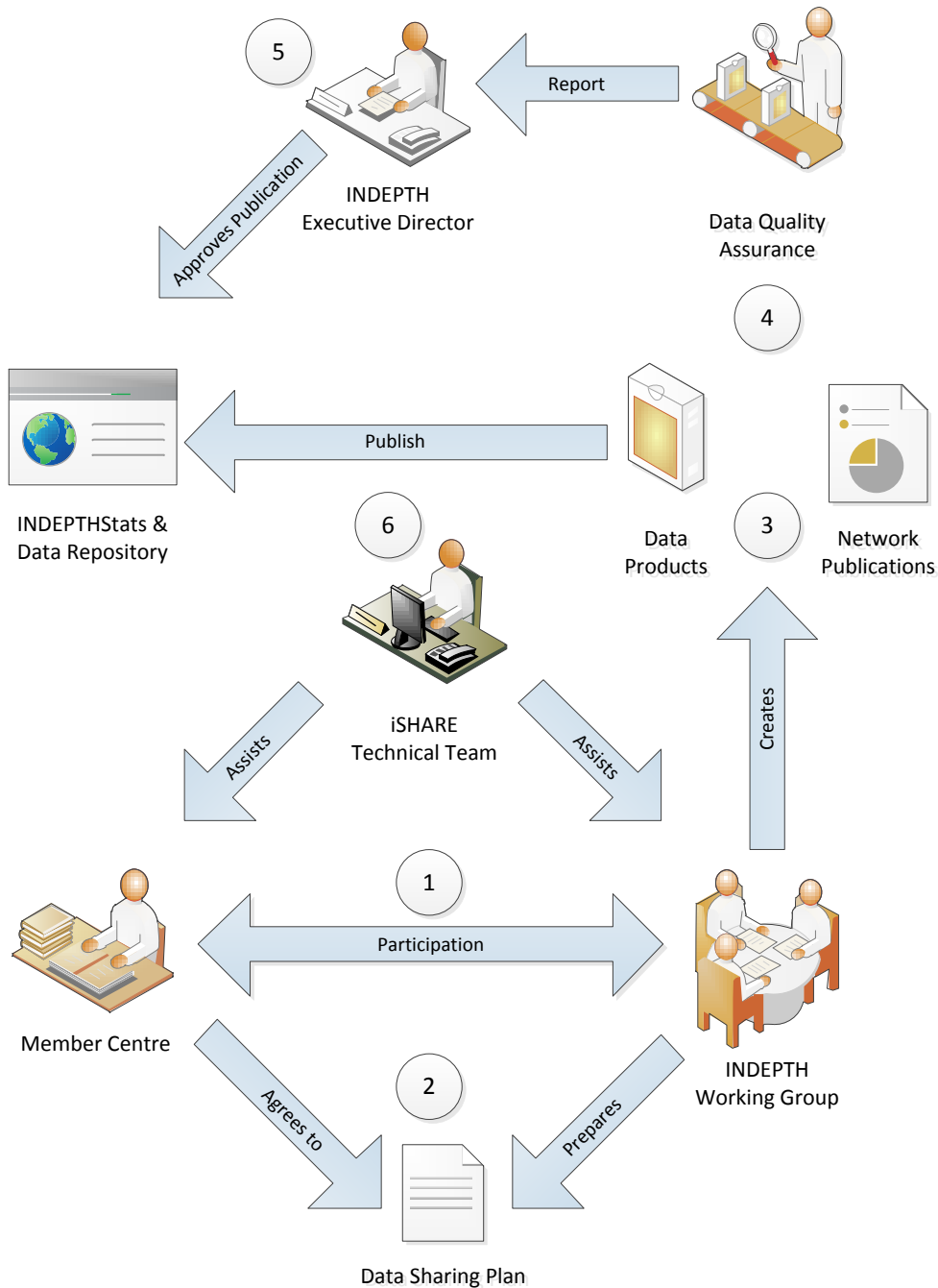
Figure 2 : Network Project Data Process

1. A network activity is initiated by an INDEPTH Working Group (WG) (with ED approval).
2. The WG prepares a data sharing plan that all members agree to.
3. The WG produces network publications and with the assistance of the iSHARE Technical Team prepares the data products underlying their network publications.
4. Data quality assurance of the data products are undertaken by the WG or commissioned by the ED.
5. The ED approves publication of the WG data products
6. The iSHARE Technical Team publishes the WG data products on the INDEPTH Data Repository.

Background

Our underlying principle in developing a Network wide data access and sharing policy is that data are a public good and sharing data is both *ethical* and *beneficial*. Our ultimate goal is to promote the wide use of our data by the scientific community and eventually make more meaningful contribution towards promoting societal wellbeing. However, we remain mindful of the challenges and complexities of data sharing and will continue to make efforts to create a well-functioning research and innovation infrastructure, supported by institutions that are committed to protecting the interests of key stakeholders that are integral to a sustainable data sharing environment.

INDEPTH Network[1] is a southern-based research network comprising currently of 35 independent member centres running 43 health and demographic surveillance systems (HDSSs) which conduct longitudinal health and demographic evaluation of populations in 20 low- and middle-income countries. The centres that constitute the INDEPTH Network cover over 3 million individuals across Africa, Asia and Oceania. This represents one of the largest sources of longitudinal data and scientific evidence on the health and living conditions of the global south in the world today.

The data collected by these centres are unique in several respects and serve a number of functions:

- They provide longitudinal data on the changing dynamics of population and health that more accurately reflects the prevailing conditions in resource poor and geographically defined areas for which such information is hard to obtain;
- They can serve as platforms for conducting policy relevant research and for monitoring and tracking new health threats, such as emerging and re-emerging infectious diseases and drug resistance health conditions so as to alert relevant stakeholders to prepare for appropriate interventions;
- They assist in developing standard data collection methods and in testing and evaluating cost-effective health interventions.

To more fully realise the benefits of sharing such rich datasets, while minimizing the concerns and challenges associated with data sharing, it is imperative that the process be guided by a set of principles and a comprehensive policy. INDEPTH's data access and sharing policy document seeks to ensure that the Network conforms to international best practices of data sharing and also upholds the Network's own position expressed at its July 2011 Nairobi meeting[2]. The Policy reflects the Network's and its individual member's commitment to responsibly, efficiently and widely share public health research data within and beyond the Network in a sustainable manner. The policy is also expected to strengthen the Network's capacity to effectively manage, curate, analyse and publish research outputs and to foster balanced collaborations and partnerships with the scientific community.

This policy was drafted by the INDEPTH Data Access and Sharing Committee (iDASC) operating under the auspices of the INDEPTH Board and will be maintained by the INDEPTH Board. The INDEPTH data access and sharing policy builds on existing Network and centre specific data access and sharing policy documents and identifies various categories of data and access levels associated with each. It also stipulates the terms, conditions, scope and time frame for accessing and sharing the different data categories equitably, ethically and efficiently. The scope of this Network policy is restricted to the

---

[1] http://www.indepth-network.org

[2] Sustainable data sharing in public health research: an INDEPTH-COHRED Position, http://www.indepth-network.org/index.php?option=com_content&task=view&id=1262&Itemid=595

sharing of those data (falling under different data types mentioned in this policy document) that are submitted by member centres to the Network.

Defining terms

The Network
The INDEPTH Network made up of member Centres operating HDSS[3].

Executive Director
The Executive Director of the INDEPTH Network.

Member Centre
A member centre is an institution that operates at least one HDSS and fulfils all the requirements for full membership of the Network.

Network activity
An INDEPTH initiated and/or sanctioned research activity involving three or more Member Centres with a designated Principal Investigator (PI).

Third party sponsored activities
Third party sponsored activities are research activities initiated externally to the Network and involving three or more Network member centres and facilitated by and/or conducted in partnership with the Network.

Contributed data
Data made available by a *Member Centre* to the Network in terms of fulfilling its network membership requirements or in terms of a specific data use agreement between the member centre and the Network.

Micro-data
An observation data collected on an individual object - statistical unit[4]. For the purposes of this policy, micro-data are defined as files of records pertaining to individual respondent units, such as homesteads, households and individuals.

Data Categories
This policy refers to the following categories of data.

A. **Centre-Specific Data** are all data generated by and specific to a particular *member centre*.
B. **Network Core Data** are *data contributed* by a *member centre* as a condition of Network membership. These data can be shared in any one of the following forms:
    a. *Micro-data* at a level of a surveillance unit (individual, household, etc). For the current core minimum micro-data set specifications see Annex 2.
    b. Aggregated (indicator) data, which is:
        i.   Derived from the core minimum micro-data, or
        ii. Provided in aggregated form directly by a Member Centre
C. **Network Project Data** are data resulting from *Network activities* of the following kinds:
    a. *Network activities* that are exclusively based on secondary analysis of data from participating member centres.

---

[3]HDSS.Health and Demographic Surveillance System.
[4] OECD Glossary of Statistical Terms. http://stats.oecd.org/glossary/detail.asp?ID=1656

b. *Network activities* that involve primary data collection.
D. **Third Party Data** are data resulting from *third party sponsored activities*.

Levels of Access

For the purposes of this policy, the following access levels are applicable:

A. **Open Access**: Except for attribution of origin, no conditions and prior registration are applicable to the use of the data.
B. **Licensed Access**: A registration by a prospective data user is required on the INDEPTH data repository with the following minimum information:
    a. Name
    b. Email address
    c. Institutional affiliation
    d. Country
    e. Category of user (e.g. student, researcher, etc)
    f. A statement of purpose for which the data will be used, and
    g. An agreement to the conditions of use detailed in Annex 3.
C. **Restricted Licensed Access:** In addition to the requirements listed under B the prospective data user will be required to provide a formal endorsement letter from the institution s/he is affiliated with.
D. **Closed Access**: This applies to highly sensitive or individually identifiable data. Such data are normally available to prospective users only through controlled-on-site access and/or in collaboration with the member centres involved.

Attribution

All data sets accessed under this policy should be cited in the following manner: "INDEPTH Network data with contribution from member centres …. (centre names), doi[5] xxxxxxxx"

iSHARE

A Network activity that provides technical and infrastructure support to INDEPTH member centres for accessing and sharing Network Core and Project Data on the INDEPTH Data Repository.

INDEPTH Data Repository

A web-based catalogue which lists, describes and facilitates access to all datasets covered by this policy. It also enables users to download available data sets provided that they agree to applicable conditions contained in the policy.

Disclosure Risk

Shared datasets must be anonymised and should not contain data elements that will allow the direct identification of the entity (individual, household or homestead) represented by a data record, such as names, identification numbers or geo-coordinates. It is not possible to completely eliminate the risk of indirect identity disclosure where a combination of data elements may be used in combination with (linked to) other information known to an intruder about the entity in question to identify the entity. Access to datasets should be assessed based on the disclosure risk posed by the dataset. Only datasets that pose a low risk of identity disclosure should be made publicly available. The Executive Director will develop guidelines for the assessment of identity disclosure risk in data to be shared in terms of this

---

[5]Digital Object Indentifier (DOI®). All datasets shared by INDEPTH through iSHARE will be assigned a doi obtained from DataCite.

policy. Such guidelines should include recommendations for specific access restrictions to datasets depending on the measured disclosure risk.

Annex 2: Core minimum micro-data set specifications

Introduction

Any demographic surveillance site requires a standard format for core individual surveillance data in order to calculate individual exposure and basic demographic rates. In the case of INDEPTH Network this format will allow member centres to contribute their surveillance data in a standard way for Network analytical purposes and will facilitate pooling of Network data for multi-site analysis. This format should also allow the calculation of data quality metrics in a standard fashion.

Based on the experience of the INDEPTH Migration, Urbanisation and Health Working group a format based on an event history analysis approach is used as the basis of this specification. This is a flexible format that lends itself to the calculation of basic demographic rates, but is also suitable for event history analysis. This core dataset can be easily expanded through the addition of event attributes or other status events (e.g. educational attainment, employment status or anthropometry measurements). The surveillance data associated with a particular individual over the course of his/her exposure to demographic surveillance are represented by a series of event records. The first event for any individual will either be enumeration, birth or in-migration followed by a variable number of event records for each observed event associated with that individual, such as out-migration or death.

Events

Table 2 defines the different possible events in this specification. Two categories of events are defined, namely, events that changes the residency status of the individual and incidental events (or observations) that do not. Each event has a set of standard attributes (Table 1) that are common to all events, followed by a variable number of attributes specific to the event.

Attribute	Variable Name	Description
Record Number	RecNr	A sequential number uniquely identifying each record in the data file
Centre Identifier	CentreId	An identifier issued by INDEPTH to each member centre of the format CCCSS, where CCC is a sequential centre identifier and SS is a sequential identifier of the site within the centre in the case of multiple site centres.
Individual Identifier	IndividualId	A number uniquely identifying all the records belonging to a specific individual in the data file. For data anonymisation purposes, this number should not be the same as the identifier used by a contributing centre to identify the individual, but the contributing centre should retain a mapping from this identifier to their identifier.
Country Identifier	CountryId	ISO 3166-1 numeric code of the country in which the surveillance site is situated
Location Identifier	LocationId	Unique identifier associated with a residential unit within the site and is the location where the individual was or became resident when the event occurred. For data anonymisation purposes, this identifier should not be the same as the identifier used internally by the contributing centre, but the contributing centre should retain a mapping of this identifier to their internal location identifier.
Date of birth	DoB	The date of birth of the individual
Event	EventCode	A code identifying the type of event that has occurred. (see Table 2)
Event date	EventDate	The date on which the event occurred
Observation date	ObservationDate	Date on which the event was observed (recorded), also known as surveillance visit date
Event count	EventCount[6]	The total number of events associated with this individual in this dataset
Event number	EventNr[1]	A number increasing from 1 to EventCount for each event record in order of event occurrence

Event	Code	Definition	Attributes	Attribute Description
Birth	BTH	The birth of an individual to a resident female.	MotherId	The IndividualId of the mother
			DeliveryEventId	The RecNr of the delivery event associated with this birth
Enumeration	ENU	Starting event for all individuals present at the baseline census of the surveillance area. It is the date on which the individual was first observed to be present in the surveillance area during the baseline census.		
In-migration	IMG	The event of migrating into the surveillance area	Origin	Classification scheme to be developed by MADIMAH
Out-migration	OMG	The event of migrating out of the surveillance area	Destination	Classification scheme to be developed by MADIMAH
Location exit	EXT	The event of leaving a residential location within the surveillance area to take up residence in another residential location within the surveillance area	Destination	The LocationId of the location within the surveillance area to which the individual relocated

---

[6] For data production use only

Location entry	ENT	The event of taking up residence in a residential location within the surveillance area following a location exit event. Note that location exit and entry are actually two parts of the same action of changing residential location and as such happens on the same event date	Origin	The LocationId of the residential location from which the individual moved.
Death	DTH	The death of the individual under surveillance. The date of death is the event date.	Cause1 Cause2 Cause3 Likelihood1 Likelihood2 Likelihood3	Up to three causes of death coded using the WHO list of verbal autopsy death causes. Likelihood values associated with each possible cause of death.
Delivery	DLV	The event of a pregnancy end after 28 weeks of gestation, which may or may not result in the birth of one or more individuals (represented in this dataset by a BTH event linked to this delivery event)	LBCnt SBCnt Parity	Live birth count Stillbirth count The number of live births to this women prior to this delivery
Observation end	OBE	An event inserted when a dataset is right censored at an arbitrary date and this individual remained under surveillance beyond this date. The right censor date is the date of this event		
Last Observation	OBL	An event indicating the last point in time on which this individual was observed to be present and under surveillance. Event date equals observation date in this instance. Normally there should be no individuals with this event as their last event if the right censoring date is prior to the start of the last complete census round.		
Observation	OBS	Used to record characteristics of individuals under surveillance valid at the time of the observation. Could be used to record aspects such as educational attainment, employment status or anthropometry measures. Specific examples of this event is not part of the minimum core individual dataset, but is specified to allow for site or working group needs.		

Annex 3: Conditions of use agreements

A. Data Use Agreement

1. Data and other material provided by INDEPTH will not be redistributed or sold to other individuals, institutions or organisations without INDEPTH's written agreement.
2. Data originating from a single contributing member centre of the INDEPTH Network may not be analysed or reported on in isolation without the express permission of the member centre concerned.
3. No attempt will be made to re-identify respondents, and there will be no use of the identity of any person or establishment discovered inadvertently. Any such discovery will be reported immediately to INDEPTH.
4. No attempt will be made to produce links between datasets provided by INDEPTH or between INDEPTH data and other datasets that could identify individuals.
5. Any books, articles, conference papers, theses, dissertations, reports or other publications employing data obtained from INDEPTH will cite the source, in line with the citation requirement provided with the dataset.
6. An electronic copy of all publications based on the requested data will be sent to INDEPTH.
7. The original collector of the data, INDEPTH, and the relevant funding agencies bear no responsibility for the data's use or interpretation or inferences based upon it.

The following applies to restricted licensed access only:

8. The researcher's organisation must be identified, as must the principal and other researchers involved in using the data. The principal researcher must sign the license on behalf of the organization. If the principal researcher is not authorized to sign on behalf of the receiving organization, a suitable representative must be identified.
9. The intended use of the data, including a list of expected outputs and the organisation's data dissemination policy must be provided.

B. Example Micro Data Set Submission Agreement between Submitting Centre and INDEPTH

Agreement between [providing centre] and INDEPTH regarding the deposit and use of micro data

A. This agreement relates to the following micro datasets:

Nr	Data Set	Access level
1.		
2.		
3.		
4.		

B. Terms of the agreement:

As the owner of the copyright in the materials listed in section A, or as duly authorized by the owner of the copyright in the materials, the representative of [providing centre] grants the INDEPTH permission for the datasets listed in section A to be used by INDEPTH, subject to the following conditions:

1. Micro data (including subsets of the datasets) and copyrighted materials provided by the [providing centre] will not be redistributed or sold to other individuals, institutions or organisations without the [providing centre]'s written agreement. Non-copyrighted materials which do not contain micro data (such as survey questionnaires, manuals, codebooks, or data dictionaries) may be distributed without further authorization. The ownership of all materials provided by the [providing centre] remains with the [providing centre].
2. Data will be used for statistical and scientific research purposes only. They will be employed solely for reporting aggregated information, including modelling, and not for investigating specific individuals or organisations.
3. No attempt will be made to re-identify respondents, and there will be no use of the identity of any person or establishment discovered inadvertently. Any such discovery will be reported immediately to the [providing centre].
4. No attempt will be made to produce links between datasets provided by the [providing centre] or between [providing centre] data and other datasets that could identify individuals or organisations.
5. Any books, articles, conference papers, theses, dissertations, reports or other publications employing data obtained from the [providing centre] will cite the source, in line with the citation requirement provided with the dataset.
6. An electronic copy of all publications based on the requested data will be sent to the [providing centre].
7. The [providing centre] and the relevant funding agencies bear no responsibility the data's use or for interpretation or inferences based upon it.
8. Data will be stored in a secure environment, with adequate access restrictions. The [providing centre] may at any time request information on the storage and dissemination facilities in place.
9. INDEPTH will provide an annual report on uses and users of the listed micro datasets to the [providing centre], with information on the number of researchers having accessed each dataset, and on the output of this research.
10. This access is granted for the duration of the [providing centre] membership of the INEPTH Network.

C. Confidentiality:

The [providing centre] confirms that there are no ethical or legal obligations that prevent the use and sharing of the micro data sets listed in A, at the indicated access level.

D. Communications:

INDEPTH will appoint a contact person who will act as focal person for this agreement. Should the focal person be replaced, the INDEPTH will communicate the name and coordinates of the new contact person to the [providing centre]. Communications for administrative and procedural purposes may be made by email, fax or letter as follows:

Communications made by [providing centre] to INDEPTH will be directed to:

Name of contact person:
Title of contact person:
Address of the recipient centre:
Email:
Tel:
Fax:

Communications made by INDEPTH to [providing centre] will be directed to:

Name of contact person:
Title of contact person:
Address of the recipient centre:
Email:
Tel:
Fax:

D. Signatories

The following signatories have read and agree with the Agreement as presented above:

Representative of the [providing centre]

Name _____

Signature _____ Date _____

Representative of INDEPTH

Name _____

Signature _____ Date _____