INDEPTH Data Repository Fact Sheet

**Wellcome Trust:** “We are proud to be a long-term supporter of INDEPTH, a unique network of demographic surveillance sites that collect, maintain, validate and analyse vital population data that are taken for granted in developed health systems. Making these valuable datasets widely accessible and providing the tools to make sense of them are significant achievements, and we believe such an approach to data sharing has tremendous potential for informing global public health policies.” *Dr Jimmy Whitworth, Head of International Activities.*

**The William and Flora Hewlett Foundation:** “The new data sharing platform is a tremendous achievement for the INDEPTH network. It will permit researchers to realize the full potential of INDEPTH’s unique data resources, answering questions about health and demographic trends that are fundamental for better social and economic policies and programs. By voluntarily making data available to their peers, researchers are making their field stronger and contributing to the best science.” *Ruth Levine, Director, Global Development and Population Program.*

**WHO:** “INDEPTH data and statistics are critical to fill information gaps on health and populations in many countries. This is an important step towards improving evidence-based reviews and decision-making in the context of national health sector strategic plans.” *Ties Boerma, Director, Health Statistics & Information Systems.*

**COHRED:** “COHRED congratulates INDEPTH on achieving this important milestone that has challenged the INDEPTH network to overcome substantive technical and ethical obstacles. Now that the data of the INDEPTH network are shared, we need to tackle another challenge: increasing the analytical capacity of the 'global south' so that it does not simply become a data producer and exporter, but is able to produce and use that data to generate new knowledge, new insights, and new solutions for its own health and development problems. Making the data available was a key first step. Using them for optimal health, equity and development is our next aim. We look forward to our continued collaboration with INDEPTH to address this.” *Carel Ijsselmuiden, Executive Director and Debbie Marais, Programme Director.*

**UNESCO:** “The launch of INDEPTH Data Repository and INDEPTHStats is a remarkable achievement designed to further bridge the gap between scientific research, data analyses and policymaking in Africa and beyond. Both instruments provide policymakers and relevant stakeholders working on population issues, migration, health, mortality and related fields, a wide range of statistical data and analyses that will assist in better informing and shaping policy decisions. INDEPTH Network should be commended for its leadership in this area.” *Dr. Abdul Rahman Lamin, Social Scientist (UNESCO Office, Accra, Ghana)*

**Rockefeller Foundation:** “Since its inception over a decade ago the INDEPTH Network has continually proved to be an important provider of technical resources, key household level
Mwihaki K. Muraguri, Associate Director; Nairobi, Kenya.

Sida: “Sida congratulates INDEPTH Network on this remarkable and historic achievement. Making INDEPTH and its HDSS partners data publically accessible will strengthen governments efforts to provide evidence-based and equitable health care to its populations. We are proud to support INDEPTH and their efforts to contribute in bridging the gap between research, policy and practice.” Anders Granlund, Head of Research Swedish International Development Agency, Stockholm, Sweden.

What is the INDEPTH Data Repository?

- The INDEPTH Data Repository is an online archive of various fully documented, high-quality datasets from INDEPTH member HDSS centres.
- Its goal is to enable INDEPTH member HDSSs and associated researchers to contribute and share HDSS datasets with the scientific community in support of the Network's mission.
- Every dataset is documented, using an internationally accepted metadata standard developed by the Data Documentation Initiative (DDI), enabling data users to quickly identify and obtain the data they require.
- Through the use of digital object identifiers (DOI) the documentation promotes the citing of data sets by data users and facilitates the recognition of the efforts by the INDEPTH Network to make this valuable resource of population and health data for low- and middle-income countries (LMICs) available.

Why INDEPTH Data Repository?

- There is an expectation from a growing body of scientists, research funders, and indeed the general public, that the data underlying published research or generated by research projects should be freely accessible.
- Sharing research data can add value to the data at little cost by, for example, using the data to answer questions not anticipated by the original research, or facilitating meta-analyses by pooling similar data from many sources. Individuals who participate in and taxpayers who directly or indirectly support research are justified in expecting that the maximum benefit will be obtained from the research.
- It is important to ensure that the means and capacity to share and actively participate in the analysis of those data are in the hands of those who generate the data and not only in those who want to analyse it.
- The INDEPTH Data Repository will be the first data repository that specialises in longitudinal individual level data from LMICs and presents an opportunity to exert a powerful and sustained influence on the availability of well documented and high...
quality longitudinal individual level exposure and cause specific mortality data from LMICs where such data has traditionally been difficult to obtain.

- The lack of research data management skills and capacity is generally acknowledged to be a major factor in why such data is not more generally available. As far back as 2006 the INDEPTH Network has taken important steps to address this through the establishment of the iSHARE project with core support funding from Sida/Research Cooperation Unit, Hewlett Foundation, Gates Foundation, Rockefeller Foundation and Wellcome Trust.

What is the importance of having effective data management systems in health research?

In short, effective data management systems in health research results in better science. The data life cycle identifies the phases in health research where effective data management is important:

a. *Discovery and planning new research.* Through the analysis of existing data and publications scientist decide on which new research questions to ask, how to prioritise those questions and how to investigate those questions in ways that are most likely to provide reliable answers. Quick and comprehensive access to research data is clearly of critical importance here.

b. *Data collection.* To answer a chosen research question, a particular research methodology (laboratory experiment, clinical trial, survey) has to be implemented and the data collected. Effective research data management determines whether the data is accurate, comprehensive and collected timeously. Proper research data management is also essential in protecting research participants, for example, by maintaining confidentiality and identifying any adverse effects of the intervention being investigated.

c. *Data preparation and analysis.* The operational research data collected in the process of conducting the research are rarely in a form that can be statistically analysed directly by the study investigators and statisticians. The data needs to be extracted, edited and formatted into analytical datasets suitable for analysis to produce research results.

d. *Publication and sharing.* The previous step resulted in research outputs that can now be published. Consistent with the Panton Principles¹ more and more scientific journals require the underlying data to research findings to made available, in part so that other researchers can reproduce the published research findings and so verify them. Effective research data management is critical to ensure such data sets are properly documented, citable, and anonymised to protect the confidentiality of research participants.

e. *Long term management.* It is important for research data to be properly archived and protected against technological changes, so that future researchers still have access to historical data. Such historical data has proved invaluable in climate and

¹ [http://pantonprinciples.org/]
demographic research. Again effective research data management is crucial in this regard.

What are the benefits of data sharing/data management in the field?

The benefits of sharing research data essentially boils down to the following main premises:

a. Sharing of data reinforces the principles of open scientific enquiry as expressed in the Panton Principles quoted above.

b. Sharing data allows more value to be extracted from the data beyond what the original investigators may be able to do. This is particularly true for basic data such as genomic data, but is also relevant for public health data such as the longitudinal population-based data that INDEPTH will share. Re-analysis of the data may lead to new insights and new research.

c. The process of data sharing and the support provided by funders to do so, enhance research data management capacity and practice. This contributes to the improvement of measurement and data collection methods and advances analytical techniques.

d. Shared data represents a valuable resource to train new scientists.

Available Data on the INDEPTH Data Repository

- The INDEPTH Data Repository is a long term project of the Network and the datasets available in the repository will continue to expand in concert with the Network’s effort to build its research data management capacity. At the launch of the Repository on 1st July 2013, the repository will contain the detail datasets underlying the indicators on INDEPTHStats for eight Network Centres and the data from an INDEPTH collaboration on the study of the epidemiology of epilepsy in demographic sites (SEEDS).

- The detail datasets will contain data in event-history format for approximately 800,000 individuals representing more than 3.7 million person years of observation. The dataset format corresponds to the standard micro-dataset format recently published by INDEPTH² and contains a data record for each observed individual demographic event.

Data that will be available at the launch will come from Africa Centre, Agincourt and Dikgale HDSSs (South Africa), Dodowa, Kintampo and Navrongo HDSSs (Ghana), Kilifi, Kisumu and Nairobi HDSSs (Kenya), Magu and Rufiji HDSSs (Tanzania), Ouagadougou HDSS (Burkina Faso), Farafenni HDSS (Gambia), Niakhar HDSS (Senegal), Manhica HDSS (Mozambique), Vadu HDSS (India), Matlab and Chakaria HDSSs (Bangladesh) and Wosera HDSS (Papua New Guinea) Chililab HDSS (Vietnam).

---

New data will be added annually on 1st July each year.

In the near future, we will be adding data from the following HDSSs: Mbita HDSS (Kenya), Iganga/Mayuge and Rakai HDSSs (Uganda), Ifakara HDSS (Tanzania), Karonga HDSS (Malawi), Nouna, Kaya, Sapone, Nanoro HDSSs (Burkina Faso), West Kiang HDSS (Gambia), Mlomp and Bandafassi HDSSs (Senegal), Taabo HDSS (Ivory Coast), Nahuche (Nigeria), Gilgel Gibe, Butajira and Kilite Awlaeelo HDSSs (Ethiopia), Kanchanaburi HDSS (Thailand), Purworejo HDSS (Indonesia), PIH HDSS (Papua New Guinea), Bandarban HDSS (Bangladesh), Bandim HDSS (Guinea Bissau), Filabavi and Dodalab HDSSs (Vietnam) and Ballabgarh HDSS (India).

INDEPTHT worked closely with COHRED to develop a set of ethical guidelines that will ensure that the data sharing repository and other activities are implemented in line with the highest ethical standards that apply to this new field of scientific endeavour.

For more information please visit the INDEPTH Network website: http://www.indepth-network.org or the INDEPTH Data Repository website: www.indepth-ishare.org