NCEPH Data Management

Ivan Hanigan

National Centre for Epidemiology and Population Health (NCEPH)
ANU

2012-09-04
1. Overview

2. Data storage and management
Introduction

The NCEPH data management and analysis environment is:

- State-of-art cyber infrastructure for organisation and synthesis of many disparate, large and unwieldy datasets across biological, social and environmental domains.

- The system needs to perform operations in the domains of data acquisition, data transformation, visualisation, mathematical operations, statistical analysis, and publication of outputs.

- Integrating data from disparate locations and sources with eclectic structures and formats that has been stored as well as captured in real time.
The purpose of the NCEPH Data Management and Analysis Environment Policy is to describe the procedures for managing

- The data collection
- Exposing the data
- Managing access for analysis
- Other roles and duties of the Data Manager.
Guiding principles

The fundamental principles underlying the policies are
- Sustainability,
- Security and
- Flexibility.
Contact

For more information please contact the Data Management Officer on

- ext 57767 or
- email ivan.hanigan@anu.edu.au
1 Overview

2 Data storage and management
The Structure

NCEPH’s data management and analysis environment

Data Providers

Data creation

VM cluster (College, Supercomputer, Research Cloud)

- Database servers (postgres, oracle)
- Analytics servers (stata, R)
- Catalogue (ddi-index)

NCEPH IT

Data Manager

Restricted

Unrestricted

NCEPH User

Students

Papers

Conferences

Figure: HaniganEcologicalDataManagement20120831.jpg
The Function

- acquire the raw data
- external sources
- collected by researcher
- file server
- database server
- simulation
- calculate new data
- analyse using stats package
- results
- metadata database
- communicate the results
- search engine web server
- technical documentation
- journal publication
- cleaning
- repurposed data
- destroy
- archive at end of project

Figure: transformationsOverview.pdf

Ivan Hanigan (NCEPH)
NCEPH Data Management
2012-09-04 9 / 9