
Preparing data for sharing and reuse

Veerle Van den Eynden
UK Data Service
University of Essex

How to Get the Maximum from Research Data
– Prerequisites and Outcomes
University of Tartu Library, 29-30 May 2018

UK Data Service



Overview

- Documentation and metadata
- Organising and formatting
- How and where to publish research data



Warm-up

- Which kind of research data do you create / use?
- Who has published research data?
- If so, why?

- What do you see as pros and cons of sharing data in research with people?



What makes data good for sharing and reuse ?

Other researchers can understand and reuse the data:

- high quality
- accurate
- well organised
- easily accessible
- well documented
- long-term validity

or FAIR data:

Findable, Accessible, Interoperable, Reusable



What is needed

- Quality assurance
- Organisation of files, folders, versions of files, file names
- Context, annotation and documentation of data
- Sustainable (open) file formats
- Deposited in a sustainable data repository
- Accessible to users
- Permissions for future use, e.g. from participants
- Free of identifying information (data protection)



DOCUMENTATION AND METADATA



Documentation (metadata) helps to make sure that data can be found, understood and interpreted by users and facilitate long-lasting usability.



Why documentation / metadata ?

Data doesn't mean anything without metadata and documentation

- A survey dataset is just a block of meaningless numbers when we do not know what the variables and codes mean
- An interview is a block of context-less text when we don't know why it was carried out and what the interviewer aimed to find out
- An archived data collection is useless when it can not be found by users searching for it

If using your data for the first time, what would a new user need to know to make sense of it?



Documentation for a dataset

Data documentation includes:

- Survey questionnaire
- Variable labels in a data table
- Interview schedule
- Consent form + information sheet
- Codebook
- Readme file
- List of interviewees and their demographic characteristics
- Published articles that provides background information



Role of metadata & documentation

- Describe dataset: content, context, provenance
 - Title, abstract, keywords, coverage, collection method,....
- Find a data collection (discovery)
 - Keywords, subjects, controlled vocabularies for search engines to index, standardisation across repositories & disciplines (e.g. registry)
- Assess use potential of data
 - Collection methods, sample size, data content,....
- Know how to access data
 - Access conditions, licence
- Citation of dataset and data use
 - Author/creator, publisher, dates, DOI, cf bibliographic reference
- Understand / interpret / use data
 - Variable labels and codes, file-level data content,...
- Exchange information between systems
 - OAI-PMH



What should be captured?

Any **useful documentation** such as:

- final report, published reports, user guide, working paper, publications, lab books

Information on **dataset structure**

- Inventory of data files
- relationships between those files
- records, cases...

Variable-level documentation

- labels, codes, classifications
- missing values
- derivations and aggregations



What should be captured?

Contextual information about project and data

- background, project history, aims, objectives, hypotheses
- publications based on data collection

Data collection **methodology and processes**

- data collection process and sampling
- instruments used - questionnaires, showcards, interview schedules
- temporal/geographic coverage
- data validation - cleaning, error-checking
- compilation of derived variables
- weighting: factors and variables, weighting process
- secondary data sources used

Data confidentiality, access and use conditions

- anonymisation carried out
- consent conditions/procedures
- access or use conditions of data



Data-level documentation

- Certain types of data file may contain important information which should be preserved:
 - variable/value labels; document metadata; table relationships and queries in relational databases; GIS data layers/tables
- Some examples:
 - SPSS: variable attributes documented in Variable View (label, code, data type, missing values)
 - MS Access: relationships between tables
 - ArcGIS: shapefiles (layers) and tables in geodatabase; metadata created in ArcCatalog
 - MS Excel: document properties, worksheet labels (where multiple)



Data-level documentation: variable names

- All structured, tabular data should have cases/records and variables adequately documented with names, labels and descriptions
- Variable names might include:
 - question number system related to questions in a survey/questionnaire
e.g. Q1a, Q1b, Q2, Q3a
 - numerical order system
e.g. V1, V2, V3
 - meaningful abbreviations or combinations of abbreviations referring to meaning of the variable
e.g. oz%=percentage ozone, GOR=Government Office Region, moocc=mother occupation, faocc=father occupation
 - for interoperability across platforms - variable names should be max 8 characters and without spaces



Data-level documentation: variable labels

- Similar principles for variable labels:
 - be brief, max. 80 characters
 - include unit of measurement where applicable
 - reference the question number of a survey or questionnaire
 - e.g. variable 'q11hexw' with label 'Q11: hours spent taking physical exercise in a typical week' - the label gives the unit of measurement and a reference to the question number (Q11b)*
- Codes of, and reasons for, missing data
 - avoid blanks, system-missing or '0' values
 - e.g. '99=not recorded', '98=not provided (no answer)', '97=not applicable', '96=not known', '95=error'*
- Coding or classification schemes used, with a bibliographic ref
 - e.g. Standard Occupational Classification 2000 - a list of codes to classify respondents' jobs; ISO 3166 alpha-2 country codes - an international standard of 2-letter country codes*



Data-level documentation: transcripts

- Qualitative data/text documents:
 - interview transcript speech demarcation (speaker tags)
 - document header with brief details of interview date, place, interviewer name, interviewee details, context



Example documentation quantitative study

Single user guide or many documents presented separately, e.g. Understanding Society study

DOCUMENTATION

Title	File Name	Size (KB)
Cognitive Ability Measures	6614_cognitive_ability_measures_v1-1.pdf	348
Revisions November 2013	6614_ukhls_2013_revisions.pdf	375
Wave 1 Adult Main Questionnaire	6614_understanding_society_wave1_questionnaire.v04.pdf	2802
Wave 2 Adult Main Questionnaire	6614_understanding_society_wave2_questionnaire_v04.pdf	3726
Waves 1-3 User Manual	6614_usermanual_wave1to3_v1-1.pdf	883
Wave 3 Youth Self-Completion Questionnaire (GB)	6614_w3_youthquestionnaire_gbritain_annotated.pdf	1469
Wave 1 Consent Package	6614_wave1_consent_package.pdf	645
Wave 1 Adult Self-Completion Questionnaire	6614_wave1_main_adult_sc_questionnaire.pdf	429
Wave 1 Youth Self-Completion Questionnaire	6614_wave1_main_youth_sc_questionnaire.pdf	750
Wave 1 Project Instructions for Interviewers	6614_wave1_project_instructions_interviewers.pdf	2426
Wave 1 Showcards	6614_wave1_showcards.pdf	100



Embedded data-level metadata in SPSS file

hse09ai.sav [DataSet2] - PASW Statistics Data Editor

File Edit View Data Transform Analyze Direct Marketing Graphs Utilities Add-ons Window Help

	Name	Type	Width	Decimals	Label	Values	Missing
175	quala10	Numeric	2	0	Which of the qualifications on this card do you have? 10	{-9, No ans...	-99 - -1
176	activb	Numeric	2	0	Activity status for last week	{-9, No ans...	-99 - -1
177	empstat	Numeric	2	0	Manager/Foreman	{-9, No ans...	-99 - -1
178	everjob	Numeric	2	0	Ever had paid employment or self-employed	{-9, No ans...	-99 - -1
179	ftptime	Numeric	2	0	Full-time or part-time	{-9, No ans...	-99 - -1
180	howlong	Numeric	2	0	How long have you been looking	{-9, No ans...	-99 - -1
181	wkstrt2	Numeric	2	0	Able to start work within 2 weeks (Government training scheme)	{-9, No ans...	-99 - -1
182	wklook4	Numeric	2	0	Looking paid work/govt scheme last 4 weeks	{-9, No ans...	-99 - -1
183	nemplee	Numeric	2	0	Number employed at place of work	{-9, No ans...	-99 - -1
184	nssec	Numeric	5	1	NS-SEC - long version (harmonised)	{-9.0, No a...	-99.0 - -1.0
185	othpaid	Numeric	2	0	Ever had other employment (waiting to start work)	{-9, No ans...	-99 - -1
186	payage	Numeric	3	0	Age when last had a paid job	{-9, No ans...	-99 - -1
187	paylast	Numeric	4	0	Year left last paid job	{-9, No ans...	-99 - -1
188	paymon	Numeric	2	0	Month last left paid job	{-9, No ans...	-99 - -1
189	sclass	Numeric	2	0	Social Class	{-9, No ans...	-99 - -1
190	seg	Numeric	2	0	Socio-Economic Group	{-9, No ans...	-99 - -1
191	snemlee	Numeric	2	0	Self employed, how many employees	{-9, No ans...	-99 - -1
192	age	Numeric	3	0	Age last birthday	{-9, No ans...	-99 - -1

Data View Variable View



Variable descriptions

	A	B	C
1	id	Id number, unique within country and season	
2	age	Age of the respondent, in months for children in years for	adults
3	sex	Sex of the respondent	
4	childadult	Child or adult	
5	country	Country of data collection	
6	season	Season of data collection	
7	proteintot~g	Total protein intake per day (gram)	
8	fattotal_g	Total fat intake per day (gram)	
9	chototal_g	total carbohydrates consumed per day (gram)	
10	ca_mg	Total calcium intake per day (gram)	
11	fe_mg	Total iron intake per day (gram)	
12	zn_mg	Total zinc intake per day (gram)	
13	vitaraemcg	Total vitamin A intake per day (microgram)	
14	foodweight	Total food weight per day (gram)	
15	folate_mcg	Total folate intake per day (microgram)	
16	energy_kcal	Total energy intake per day (kcal)	
17	vitamina_RE	Total vitamin A intake per day (RE)	
18	vitaminc_mg	Total vitamin C intake per day (mgram)	
19	speciescount	Number of different species consumed per day	
20	ddscout10	Diet diversity score for women and children	
21	ddscout21	Diet diversity score for women and children 21 food groups	
22	ddscout27	Diet diversity score for women and children 27 food groups	
23	vitamimanar	Nutrient adequacy ratio for vitamin A	
24	vitamimcnar	Nutrient adequacy ratio for vitamin C	
25	simpson	Simpsons measure of diversity	
26	entropy	Entropy	
27	dissimilar~y	Dissimilarity index z-score	
28	zsimpson	Simpsons measure of diversity z-score	
29	zentropy	Entropy z-score	
30	zdds	Diet diversity z-score	
31	zdissimila~y	Dissimilarity index z-score	
32	zMAR	Mean nutrient adequacy ratio z-score	











Variable description - DDI

Dataset: Young People's Social Attitudes, 2003

Variable **ypyngsex**: Is sex under 16 years old wrong? YP237

LITERAL QUESTION

What if it was a boy and a girl who were both still under 16? Please choose an answer from the card to show what your general opinion of this would be.

Values	Categories	N		
1	Always wrong	229		34.5%
2	Mostly wrong	192		29.0%
3	Sometimes wrong	134		20.2%
4	Rarely wrong	46		6.9%
5	Not wrong at all	35		5.3%
6	(Depends)	16		2.4%
8	Don't know	9		1.4%
9	Not answered	2		0.3%

SUMMARY STATISTICS

Valid cases 663
Missing cases 0
This variable is numeric

INTERVIEWER INSTRUCTIONS

CARD Y16 AGAIN.

UNIVERSE

ASK ALL

UK Data Service



Example documentation qualitative study

User guide could contain a variety of documents that provide context: interview schedule, transcription notes, even photos

Q21

NOTES ON THE INTERVIEW SCHEDULE

1. The household

1(c) Respondents are not often able to recite the names of the children in the family from eldest to youngest and the spaces between them. It is useful in these cases to ask where the respondent came in the family and then ask who was older than him and the spaces between the children who were older than him. Then ask about the younger ones. Respondents are sometimes vague about the respective ages of their siblings, e.g. "He came at pretty regular intervals". Try to find out what these intervals were, and if there were any exceptions to the average interval. Respondents sometimes find it easier to write down or tell you the ages and names of their siblings, alive and dead, at the present time.

1(d) When respondents do not know the age of their father when they were born, ask if they know how old their father was when he died (assuming he is dead) and what year that was. Or respondents may know the age their father was when he married and the date. Approximate dates will do.

1(e) See notes on 1(d).

2. Domestic Routine

2(a) Select the house in which respondent spent the longest time he can remember before leaving home.

2(c) Servants in this period who did not live in were usually charwomen or women who came in "to do the rough", i.e. to do the rough housework. There were also washwomen who came in to do the washing and young girls who came in to look after children. Where the respondent as a child came into a lot of contact with the servant, particularly if she looked after the respondent, find out what the relationship was between them, the sort of things she did for the respondent, etc.

2(d) Older children sometimes looked after the younger children, took them out for walks, saw them to school, etc.

3. Meals

3(c) Men and women whose working day started early would often take something with them for breakfast. When asking about meals find out when the respondent took food and what he called those meals and stick to the terminology he uses. Lunch is the midday meal to some, particularly in class 1 and 2, to an agricultural labourer it is a snack eaten at about 11 a.m.. Dinner is the midday meal to the majority of respondents. To some, again in class 1 and 2 it is a meal at about 3 or 5 p.m.. Tea to most respondents is a meal mainly of bread and tea with occasionally something cooked, and is the last meal of the day. To some, in class 1 and 2 mainly, it is a cup of tea and bread and butter and cake at about 4 p.m.. It is usually distinguished as afternoon tea in that case. Supper may be a cup of cocoa and some bread and cheese taken just before bed at 9 pm when tea has been the last meal at about 5 p.m.. Or it may be a meal of two courses either hot or cold eaten at about 7 p.m.

3(k) Sometimes a person might take his plate and sit by the corner of the fire during a meal. Or a person in a hurry might snatch some food standing up.



Transcript template and datalist

Study Number 6124
 Being a Doctor: a Sociological Analysis, 2005-2006
 Nettleton, S

Interview ID	Gender	Description	Date of Interview	No of Pages	Text File Name
x001	Male	Interview with a Hospital Doctor in a Multi-Ethnic Northern City	07/09/2005	36	6124int001
x002	Male	Interview with a Hospital Doctor in a Multi-Ethnic Northern City	09/09/2005	41	6124int002
x003	Male	Interview with a Hospital Doctor in a Multi-Ethnic Northern City	09/09/2005	39	6124int003
x004	Female	Interview with a Hospital Doctor in a Multi-Ethnic Northern City	13/09/2005	36	6124int004
x005	Male	Interview with a Hospital Doctor in a Multi-Ethnic Northern City	13/09/2005	34	6124int005
x006	Male	Interview with a Hospital Doctor in a Multi-Ethnic Northern City	14/09/2005	50	6124int006
x007	Male	Interview with a Hospital Doctor in a Multi-Ethnic Northern City	21/09/2005	31	6124int007
x008	Male	Interview with a Hospital Doctor in a Multi-Ethnic Northern City	21/09/2005	35	6124int008
x009	Male	Interview with a Hospital Doctor in a Multi-Ethnic Northern City	22/09/2005	33	6124int009
x010	Female	Interview with a Hospital Doctor in a Multi-Ethnic Northern City	22/09/2005	23	6124int010
x011	Male	Interview with a Hospital Doctor in a Multi-Ethnic Northern City	22/09/2005	36	6124int011
x012	Female	Interview with a Hospital Doctor in a Multi-Ethnic Northern City	27/09/2005	41	6124int012
x013	Female	Interview with a Hospital Doctor in a Multi-Ethnic Northern City	27/09/2005	21	6124int013
x014	Female	Interview with a Hospital Doctor in a Multi-Ethnic Northern City	30/09/2005	20	6124int014
x015	Male	Interview with a Hospital Doctor in a Multi-Ethnic Northern City	05/10/2005	19	6124int015
x016	Female	Interview with a Hospital Doctor in a Multi-Ethnic Northern City	05/10/2005	27	6124int016
x017	Male	Interview with a Hospital Doctor in a Multi-Ethnic Northern City	07/10/2005	27	6124int017
x018	Male	Interview with a Hospital Doctor in a Multi-Ethnic Northern City	17/10/2005	11	6124int018
x019	Male	Interview with a Hospital Doctor in a Multi-Ethnic Northern City	19/10/2005	33	6124int019
x020	Female	Interview with a Hospital Doctor in a Multi-Ethnic Northern City	07/11/2005	21	6124int020
z001	Male	Interview with Hospital Doctor in Northern Tourist City	19/07/2005	50	6124int021
z002	Male	Interview with Hospital Doctor in Northern Tourist City	10/08/2005	45	6124int022
z003	Male	Interview with Hospital Doctor in Northern Tourist City	17/08/2005	56	6124int023
z004	Male	Interview with Hospital Doctor in Northern Tourist City	07/11/2005	21	6124int024
z005	Female	Interview with Hospital Doctor in Northern Tourist City	14/11/2005	32	6124int025
z006	Male	Interview with Hospital Doctor in Northern Tourist City	15/11/2005	23	6124int026
z007	Male	Interview with Hospital Doctor in Northern Tourist City	16/11/2005	23	6124int027
z008	Female	Interview with Hospital Doctor in Northern Tourist City	17/11/2005	18	6124int028
z009	Male	Interview with Hospital Doctor in Northern Tourist City	18/11/2005	20	6124int029



Metadata

- Highly **structured documentation**
- Standard data collection metadata includes:
 - Components of a bibliographic reference
 - Core information that a search engine indexes to make the data findable
- International standards/schemes
 - Data Documentation Initiative (DDI) (social sciences, incl variable level)
 - ISO19115 (geospatial)
 - Dublin Core (basic)
 - DataCite Metadata Schema (data citation, with doi)



Data paper: summum of metadata

www.nature.com/scientificdata

SCIENTIFIC DATA

OPEN Data Descriptor: Spatial and temporal dynamics of multidimensional well-being, livelihoods and ecosystem services in coastal Bangladesh

Received: 22 March 2016
Accepted: 30 September 2016
Published: 08 November 2016

Helen Adams^{1,2}, W. Neil Adger², Sate Ahmad³, Ali Ahmed³, Dilruba Begam³, Attila N. Lázár⁴, Zoe Matthews⁵, Mohammed Mofizur Rahman⁶ & Peter Kim Streatfield³

Populations in resource dependent economies gain well-being from the natural environment, in highly spatially and temporally variable patterns. To collect information on this, we designed and implemented a 1586-household quantitative survey in the southwest coastal zone of Bangladesh. Data were collected on material, subjective and health dimensions of well-being in the context of natural resource use, particularly agriculture, aquaculture, mangroves and fisheries. The questionnaire included questions on factors that mediate poverty outcomes: mobility and remittances; loans and micro-credit; environmental perceptions; shocks; and women's empowerment. The data are stratified by social-ecological systems to take into account spatial dynamics and the survey was repeated with the same respondents three times within a year to incorporate seasonal dynamics. The dataset includes blood pressure measurements and height and weight of men, women and children. In addition, the household listing includes basic data on livelihoods and income for approximately 10,000 households. The dataset facilitates interdisciplinary research on spatial and temporal dynamics of well-being in the context of natural resource dependence in low income countries.

Design Type	time series design
Measurement Type(s)	Household Environment • anthropogenic habitat • social environment condition
Technology Type(s)	Cluster Random Sampling • defining social-ecological system • survey method
Factor Type(s)	socio-ecological system
Sample Characteristic(s)	Homo sapiens • Sathkhira District • fish farm • Khusha District • mangrove swamp • freshwater fish product • Bagerhat District • Pirgajar District • farm • Barguna District • sea coast • Barisal District • river bank • Bhola District • Patuakhali District • Jhalokati District

¹Geography, King's College London, Strand Campus, London WC2R 2LS, UK. ²Geography, College of Life and Environmental Sciences, University of Exeter, Rennes Drive, Exeter EX4 4RJ, UK. ³Institute for Climate Change and Health, International Centre for Diarrhoeal Disease Research Bangladesh, GPO Box 128, Dhaka 1000, Bangladesh. ⁴Engineering and the Environment, University of Southampton, University Road, Southampton, Southampton SO17 1BJ, UK. ⁵Social Statistics & Demography, University of Southampton, 58 Salisbury Rd, Southampton SO17 1BJ, UK. Correspondence and requests for materials should be addressed to H.A. (email: helen.j.adams@kcl.ac.uk).

SCIENTIFIC DATA | 3:160094 | DOI: 10.1038/sdata.2016.94

<http://www.readcube.com/articles/10.1038/sdata.2016.94>

UK Data Service



Mechanisms and dynamics of wellbeing-ecosystem service links in the southwest coastal zone of Bangladesh

Adams, Helen and Adger, Neil (2016). *Mechanisms and dynamics of wellbeing-ecosystem service links in the southwest coastal zone of Bangladesh*. [Data Collection]. Colchester, Essex: UK Data Archive. [10.5255/UKDA-SN-852356](https://beta.ukdataservice.ac.uk/datacatalog/studies/study?id=10.5255/UKDA-SN-852356)

Delta regions are probably the most vulnerable type of coastal environment and their ecosystem services face multiple stresses in the coming decades. These stresses include, amongst others, local drivers due to land subsidence, population growth and urbanisation within the deltas, regional drivers due to changes in catchment management (e.g. upstream land use and dam construction), and global climate change impacts such as sea-level rise. The ecosystem services of river deltas support high population densities, estimated at over 500 million people globally, with particular concentrations in Southern and Eastern Asia and Africa. A large proportion of these

Working Paper # 2
ESPA Deltas
www.espadeltas.net

Wellbeing-ecosystem service links: Mechanisms and dynamics southwest coastal zone of Bangladesh

Helen Adams¹, W. Neil Adger², Hamidul Hossain³, Rezaur Rahman⁴, Mas Salehin⁵

¹Geography, College of Life and Environmental Sciences, University of Exeter, UK
²Institute of Land and Water Studies, China, Bangladesh
³Geography, University of Chittagong, Bangladesh
⁴International Centre for Integrated Mountain Development, Bangladesh
⁵Faculty of Social and Human Sciences, University of Southampton, UK
⁶Faculty of Engineering and Environment, University of Southampton, UK

June 2013

^{*}Corresponding author: h.adams@exeter.ac.uk

SCIENTIFIC DATA

Data Descriptor: Spatial and temporal dynamics of multidimensional well-being, livelihoods and ecosystem services in coastal Bangladesh

OPEN

Received: 22 March 2016
Accepted: 09 September 2016
Published: 08 November 2016

Helen Adams^{1,2}, W. Neil Adger², Sate Ahmad³, Ali Ahmad³, Dilruba Begum⁴, Atila M. Lima⁵, Zoe Matthews⁶, Mohammed Mofizur Rahman⁶, Peter Kim Streetfield⁶

Populations in resource dependent economies gain wellbeing in many forms from the natural environment. The dimensions of well-being range from material aspects associated with income and consumption, through to health outcomes and perceived life satisfaction. Both well-being, and ecosystems that provide many aspects of material and subjective well-being, have high temporal and spatial variation. We designed and implemented a questionnaire survey to collect data on material, subjective and health dimensions of wellbeing in the context of natural resource use, particularly agriculture, aquaculture, mangroves and fisheries in delta environments in Bangladesh and administered the survey to 1500 households. The questionnaire included questions on factors that mediate poverty outcomes: mobility and remittances; loans and micro-credit; environmental perceptions; shocks; and women's empowerment. The data are stratified by socio-ecological system to take into account spatial dynamics and the survey was repeated with the same respondents three times within a year to incorporate seasonal dynamics. The dataset includes salinity of drinking water and blood pressure, height and weight of men, women and children. In addition, the household listing includes basic data on livelihoods and income for approximately 10000 households. The dataset facilitates interdisciplinary research on spatial and temporal dynamics of well-being in the context of natural resource dependence in low income countries.

Design Type Data Descriptor
Measurement Type(s) Questionnaire
Technology Type(s) Survey
Factor Type(s) Socio-ecological system
Sample Characteristic(s) Household listing

¹Geography, King's College London
²Environmental Science, Natural and Health, International Centre for Integrated Mountain Development, Bangladesh
³Geography and the Environment, University of Southampton, UK
⁴Geography, University of Chittagong, Bangladesh
⁵Geography, University of Exeter, UK
⁶Geography, University of Exeter, UK
(email: helen.adams@kcl.ac.uk)

SCIENTIFIC DATA | 3.166093 | DOI: 10.1038/sdata.2016.04

Gateway to Research

Assessing health, livelihoods, ecosystem services and poverty alleviation in populous deltas

Lead Research Organisation: University of Exeter
Department Name: Geography

Abstract

Delta regions are probably the most vulnerable type of coastal environment and their ecosystem services face multiple stresses in the coming decades. These stresses include, amongst others, local drivers due to land subsidence, population growth and urbanisation within the deltas, regional drivers due to changes in catchment management (e.g. upstream land use and dam construction), and global climate change impacts such as sea-level rise. The ecosystem services of river deltas support high population densities, estimated at over 500 million people globally, with particular concentrations in Southern and Eastern Asia and Africa. A large proportion of these people experience extremes of poverty and are severely exposed to vulnerability from environmental and ecological stress and degradation. In areas close to or below the poverty boundary, both subsistence and cash elements of the economy tend to rely disproportionately heavily on ecosystem services which underpin livelihoods. Understanding how to sustainably manage the ecosystem services in delta regions and thus improve health and reduce poverty and vulnerability requires consideration of all these stresses and their complex interaction. This proposal aims to develop methods to understand and characterise the key drivers of change in ecosystem services that affect the environment and economic status in the world's populous deltas. This will be done through analysis of the evolving role of ecosystem services, exploring the implications of changes for the livelihoods of delta residents, and developing management and policy options that will be beneficial now and in the future in the face of the large uncertainties of the next few decades beyond the extensive coastal fringe of the Ganges-Brahmaputra-Meghna Delta within Bangladesh has been selected as the pilot study area for this work. This is because Bangladesh is almost entirely located on one of the world's largest and most dynamic deltas. It is characterised by densely populated coastal lowlands with significant poverty, supported to a large extent by natural ecosystems such as the Sundarbans (the largest mangrove forest in the world). It is under severe development pressure due to many growing cities, eg. Khulna and the capital, Dhaka. At present the importance of ecosystems services to poverty and livelihoods is poorly understood. This is due to the complexity of interactions between physical drivers, environmental pressures and the human responses to stress and the resultant impacts on ecosystems. Government policy rarely takes up the ecosystems services perspective and as a result an holistic overview of their value is often overlooked. This project aims to address this gap by providing policy makers with the knowledge and tools to enable them to evaluate the effects of policy decisions on people's livelihoods. This will be done by

Funded Value: £264,619
Funded Period: Mar 12 - Sep 16
Funder: NERC
Project Status: Active
Project Category: Research Grant
Project Reference: NE/J00892/1
Principal Investigator: Neil Adger

Spatial and temporal dynamics of multidimensional well-being, livelihoods and ecosystem services in coastal Bangladesh

Adams, Helen and Adger, Neil and Ahmad, Sate and Ahmed, Ali and Begum, Dilruba and Matthews, Zoe and Rahman, Mohammed Mofizur and Streetfield, Kim (2016). *Spatial and temporal dynamics of multidimensional well-being, livelihoods and ecosystem services in coastal Bangladesh*. [Data Collection]. Colchester, Essex: UK Data Archive. [10.5255/UKDA-SN-852179](https://beta.ukdataservice.ac.uk/datacatalog/studies/study?id=10.5255/UKDA-SN-852179)

Populations in resource dependent economies gain wellbeing in many forms from the natural environment. The dimensions of well-being range from material aspects associated with income and consumption, through to health outcomes and perceived life satisfaction. Both well-being, and ecosystems that provide many aspects of material and subjective well-being, have high temporal and spatial variation. We designed and implemented a questionnaire survey to collect data on material, subjective and health dimensions of wellbeing in the context of natural resource use, particularly agriculture, aquaculture, mangroves and fisheries in delta environments in Bangladesh and administered the survey to 1500 households. The questionnaire included questions on factors that mediate poverty outcomes: mobility and remittances; loans and micro-credit; environmental perceptions; shocks; and women's empowerment. The data are stratified by socio-ecological system to take into account spatial dynamics and the survey was repeated with the same respondents three times within a year to incorporate seasonal dynamics. The dataset includes salinity of drinking water and blood pressure, height and weight of men, women and children. In addition, the household listing includes basic data on livelihoods and income for approximately 10000 households. Interdisciplinary analysis of the data demonstrates the spatial and temporal dynamics of wellbeing in the context of natural resource dependence in low income countries.

Creators:	Creator Name	Email	Affiliation	ORCID (as URL)
	Adams, Helen	helen.j.adams@kcl.ac.uk	King's College London	http://orcid.org/0000-0003-1732-9833
	Adger, Neil	n.adger@exeter.ac.uk	University of Exeter	http://orcid.org/0000-0003-4212-2854

- RELATED RESOURCES**
- Data collections**
 - Mechanisms and dynamics of wellbeing-ecosystem service links in the southwest coastal zone of Bangladesh
 - Publications**
 - Wellbeing-ecosystem service links: Mechanisms and dynamics in the southwest coastal zone of Bangladesh
 - Spatial and temporal dynamics of multidimensional well-being, livelihoods and ecosystem services in coastal Bangladesh
 - Website**
 - Assessing health, livelihoods, ecosystem services and poverty alleviation in populous deltas

via DataCite to researcher's ORCID profile



Application: mapping datasets



STUDY | FACULTIES | RESEARCH | ABOUT US

NEWS & EVENTS

LSHTM Data Compass

Home | About | Browse options | Year | Faculty | Research Centre | LSHTM Author | Projects | Location

Statistics | Login

Welcome to LSHTM Data Compass

[More search options...](#)

LSHTM Data Compass addresses the School's mission, to improve health and health equity in the UK and worldwide and maximise the benefit and impact of its research by ensuring the underlying data can be safeguarded, shared and cited.



How to deposit

Please take a look to our guides to depositing data for help and further information.

[Submit your data to Data Compass](#)

Latest additions

Choko, A, Webb, E and Corbett, L (2015) [Uptake, accuracy, safety and linkage into care over two years of promoting annual self-testing for HIV in Blantyre, Malawi: a community-based prospective study.](#) [Data Collection]

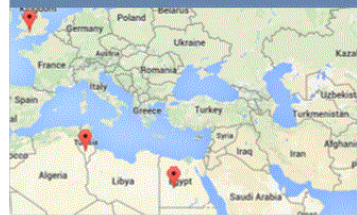
Lal, S (2015) [Effectiveness of provider and community interventions to improve treatment of uncomplicated malaria in Nigeria: a cluster randomised controlled trial.](#) [Data Collection]

Michaels-Igbokwe, C (2014) [Young people's preferences for family planning service providers in rural Malawi: a stated preference data set.](#) [Data Collection]

Fielding, K, Lewis, J and Jiang, S (2013) [Effectiveness of electronic reminders to improve medication adherence in tuberculosis patients: a cluster-randomised trial.](#) [Data Collection]

Lal, S (2015) [Measuring patient adherence to malaria treatment: A comparison of results from self-report and a customised electronic monitoring device.](#) [Data Collection]

Collections map



Tweets

[Follow](#)

UK Data Service



LSHTM Data Compass

[Home](#) [About](#) [Browse options](#) [Year](#) [Faculty](#) [Research Centre](#) [LSHTM Author](#) [Projects](#) [Location](#)

[Statistics](#) | [Login](#)

Geographical locations of data collections

[More search options...](#)

Please note that the pins shown are centred in a geographical bounding box and therefore represent an approximation of the location of the study.



UK Data Service




Application: harvest metadata

Research Data Discovery Service (Alpha)

Datasets Organisations About Reports FAQ

Home > Datasets

Filter by location [Clear](#)



Map data © OpenStreetMap contributors
Tiles by Stamen Design (CC BY 3.0)

Organisations

- UK Data Service (129)
- NERC (13)
- Archaeology Data Se... (2)
- University of Hull (1)

Groups

There are no Groups that match this search

Taas

View as JSON

denmark

Order by Relevance

[Advanced Search](#)

145 datasets found for "denmark"

Urban Population of 18th Century Denmark

Abstract copyright UK Data Service and data collection copyright owner. The study examines the conditions in a major Danish provincial town, Odense, in the eighteenth century...

Data from the Photographic Camera on-board the Piper PA31 Navajo Chieftain G-...

The Airborne Research & Survey Facility (ARSF, formerly Airborne Remote Sensing Facility) is managed by NERC Scientific Services and Programme Management. It provides the UK...

Globalisation, Technology and Wage Inequality, 1870-1970

Abstract copyright UK Data Service and data collection copyright owner. The aim of the originating project was to increase understanding of the past and possible future impact...

<http://ckan.data.alpha.jisc.ac.uk/dataset>

UK Data Service



Tools to create metadata

- [Create a codebook](#) about your research to accompany the dataset (DDI Alliance, 2017a)
- York University (n.d.) [Library Metadata Template](#), Dublin Core
- Georgia Tech Library (n.d.) [Metadata Template](#);
- Use the [Dublin Core Metadata Generator](#) (dublincoregenerator, n.d.)
- Cornell University (n.d.) [guide to writing “readme” style metadata](#) (with downloadable template)
- [ISO 19115-2 Metadata Editor](#) (GRIIDC (2015)) web application



Application: explore variables

Exploring relationships and provenance of multiple longitudinal studies

The screenshot shows the CLOSER Discovery website interface. On the left is a dark sidebar with navigation links: Home, Search, Explore, List, Help, and Feedback. The main content area features the CLOSER logo and tagline 'Promoting excellence in longitudinal research' at the top left, and logos for E-S-R-C (Economic & Social Research Council) and MRC (Medical Research Council) at the top right. A light blue banner below the header contains a welcome message and a 'Tour this Page' button. Below the banner is a search section with filters for 'Item type', 'Studies', and 'Life Stages', all currently set to 'All'. There is a search query input field and a blue 'Search' button. To the right of the search section is a 'Welcome' panel with introductory text and a list of studies: Avon Longitudinal Study of Parents and Children, 1970 British Cohort Study, Hertfordshire Cohort Study, Millennium Cohort Study, 1958 National Child Development Study, MRC National Survey of Health and Development, Southampton Women's Survey, and Understanding Society. At the bottom of the main content area are five data cards: 8 Studies, 81 Sweeps, 55,226 Variables, 192 Data Files, and 26,286 Questions.

<http://discovery.closer.ac.uk/>

UK Data Service



Question | **Related Variables**

Name qi_114_a

Label I14 a

Question Text How much alcohol do you drink?

Multiple Choice Response Options

Codes cs_qi14_a

- 1 never drink alcohol
- 2 very occasionally (less than once a week)
- 3 occasionally (at least once a week)
- 4 drink 1-2 glasses* nearly every day
- 5 drink 3-9 glasses* every day
- 6 drink at least 10 glasses * a day

Caring for a Toddler

349 questions before...

I13 How many cigarettes per day do you currently smoke?

I14 a How much alcohol do you drink?

* by glass we mean a pub measure (1oz) of spirits, half a pint (1/4 litre) of lager or cider, a wine glass of wine, etc)

67 questions after...

[View the complete instrument](#) | [Caring for a Toddler](#)

<http://discovery.closer.ac.uk/item/uk.alspac/7a8dd904-43dc-4be3-9d46-cbb2e19a5900>



Summary

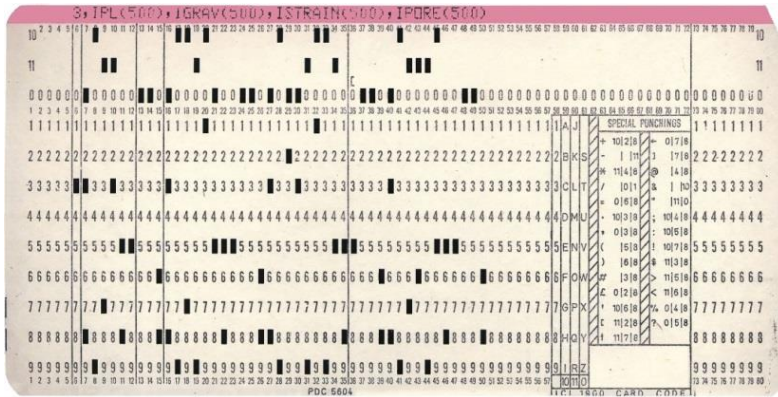
- The role of metadata for:
 - Research documentation
 - Data searching and retrieving
 - Sharing and reuse of data
 - Citation of datasets and data reuse



FORMATTING AND ORGANISING



Can you understand/use these data?



File formats

A file format is a specific way of structuring digital data so that a machine, and therefore a person, can understand and manipulate it.

Choice of software format for digital data dependent on:

- planned data analyses
- software used - availability/cost dependent
- hardware used - data type dependent e.g. audio capture
- discipline-specific standards and customs



File formats

Digital data is software dependent, so endangered by obsolescence of software/ hardware

Best formats for long-term preservation:

- should be readable by as many types of 'system' as possible
- without compromising the purpose of the data
- standard, interchangeable, open
- *e.g. tab-delimited, comma-delimited (CSV), ASCII, RTF, PDF/A, OpenDocument format, SPSS portable, XML*

Recommended file formats

Audio data	Free Lossless Audio Codec (FLAC) (.flac)	<p>MPEG-1 Audio Layer 3 (.mp3) if original created in this format</p> <p>Audio Interchange File Format (.aif)</p> <p>Waveform Audio Format (.wav)</p>
-------------------	--	---

Image data	TIFF 6.0 uncompressed (.tif)	<p>JPEG (.jpeg, .jpg, .jp2) if original created in this format</p> <p>GIF (.gif)</p> <p>TIFF other versions (.tif, .tiff)</p> <p>RAW image format (.raw)</p> <p>Photoshop files (.psd)</p> <p>BMP (.bmp)</p> <p>PNG (.png)</p> <p>Adobe Portable Document Format (PDF/A, PDF) (.pdf)</p>
-------------------	------------------------------	--

Textual data	<p>Rich Text Format (.rtf)</p> <p>plain text, ASCII (.txt)</p> <p>eXtensible Mark-up Language (.xml) text according to an appropriate Document Type Definition (DTD) or schema</p>	<p>Hypertext Mark-up Language (.html)</p> <p>widely-used formats: MS Word (.doc/.docx)</p> <p>some software-specific formats: NUD*IST, NVivo and ATLAS.ti</p>
---------------------	--	---

Video data	<p>MPEG-4 (.mp4)</p> <p>OGG video (.ogv, .ogg)</p> <p>motion JPEG 2000 (.mj2)</p>	AVCHD video (.avchd)
-------------------	---	----------------------



Quality control of data

- Integral part throughout the research project
 - During data collection
 - Data entry
 - Data checking
- Data collection and entry
 - Calibration of instruments
 - Taking multiple measurements
 - Using standardised methods and protocols
- Data checking
 - Checking inputted correctly
 - Checking data completeness
 - Adding variable and value labels

Depending on the type of data, you may be able to automate aspects of this process using:

- statistical software to check e.g. frequencies on quantitative data
- consistency checking with data manipulation tools like OpenRefine



Minimise errors in survey data entry





- Check the completeness of records: correct number of records, number of variables or length of the records
- Reduce burden at manual data entry
- Minimise the number of steps
- Conduct data entry twice (or %), e.g. manual entry or scanning from paper questionnaires
- Perform in-depth checks for selected records, e.g. 5–10% of all records
- Perform logical and consistency checks:
 - Check the value range (e.g. a respondent over the age of 100 is unlikely);
 - Check the lowest and highest values and extremes;
 - Check the relations between associated variables (e.g. educational attainment should correspond with a minimum age, the total number of hours spent doing various activities should not exceed 100% of the available time);
- Automate checks to prevent meaningless values from being entered, e.g.
 - set the range of valid values
 - apply filters to manage the data entry



Quality control for transcription

- Ensure high quality recording
- Transcriber instructions or guidelines with required transcription style, layout and editing
- Confidentiality agreement with transcriber
- Check through proofreading

Can you understand these data?

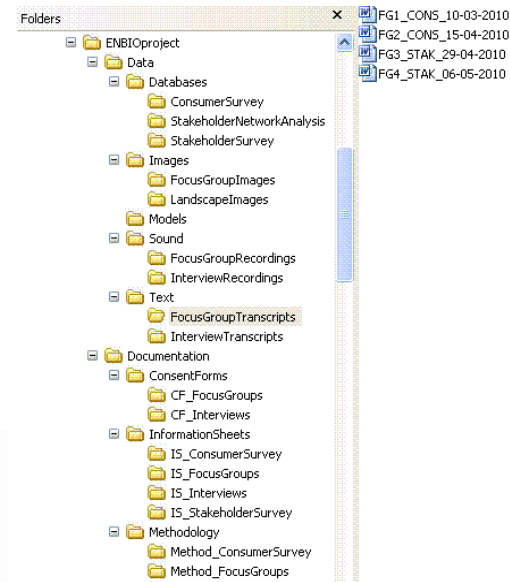
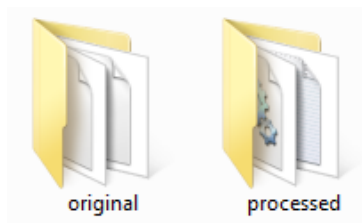
SrvMthd Draft.doc	
SrvMthd Final.doc	
SrvMthd Final 2.doc	
SrvMthd Final Version.doc	

Organising data

- Plan in advance how best to organise data
- Use a logical structure and ensure collaborators understand

Examples

- Hierarchical structure of files, grouped in folders, e.g. audio, transcripts and annotated transcripts
- Survey data: spreadsheet, SPSS, relational database
- Interview transcripts: individual well-named files



File naming

- file name = principal identifier of file
- use logical naming i.e. easy to identify and retrieve the file
- naming provides organisation, context & consistency
- name elements: version number, date, content description, creator name

Best practice

- name independent of location
- relevant to content
- no special characters, dots or spaces
- for separation use underscores _
- versioning via filename: ascending, decimal version numbers
- avoid very long file names



Example file names

- FG1_CONS_2010-02-12.rtf : interview transcript of the first focus group with consumers, that took place on 12 February 2010
- Int024_AP_2008-06-05.doc : interview with participant 024, interviewed by Anne Parsons on 5 June 2008
- BDHSurveyProcedures_00-04.pdf : version 4 of the survey procedures for the British Dental Health Survey

Version control

Keep track of different copies or versions of data files

- Useful for files kept in multiple locations
- Or which have multiple users
- A way to safeguard against accidental changes

Versioning can be done through:

- date recorded in file name, e.g. HealthTest-2008-04-06
- version number in file name, e.g. HealthTest-00-02 or HealthTest_v2
- file history or version control table where versions, dates, authors and details of changes to the file are recorded
- version control software

File name	Changes to file
Interviewschedule_1.0	Original document
Interviewschedule_1.1	Minor revisions made
Interviewschedule_1.2	Further minor revisions
Interviewschedule_2.0	Substantive changes



Example: version control table

Title:		Vision screening tests in Essex nurseries	
File Name:		VisionScreenResults_00_05	
Description:		Description of the data files	
Created By:		Chris Wilkinson	
Maintained By:		Sally Watsley	
Created:		04/07/2007	
Last Modified:		25/11/2007	
Based on:		VisionScreenDatabaseDesign_02_00	
Version	Responsible	Notes	Last amended
00_05	Sally Watsley	Version 00_03 and 00_04 compared by SW	25/11/2007
00_04	Vani Yussu	Entries checked by VY, independent from previous	17/10/2007
00_03	Steve Knight	Entries checked by SK	29/07/2007
00_02	Karin Mills	Test results 81-120 entered	05/07/2007
00_01	Karin Mills	Test results 1-80 entered	04/07/2007

PUBLISH DATA

UK Data Service



Data publishing

- Properly documented with metadata
- Reviewed for quality
- Searchable and discoverable in catalogues (or databases)
- Citable in articles



Which repository

Recommended by OpenAIRE:

- A **(trusted) domain repository** already established for your research domain
- If a domain repository isn't available, use an **institutional research data repository**
- If none of the above is available, use a general purpose repository like [Zenodo](#), [Figshare](#) or [Harvard Dataverse](#)
- Find your own at [re3data.org](#): a registry of over 1500 research data repositories

- Choose between self-archiving (quick and easy) or expert help



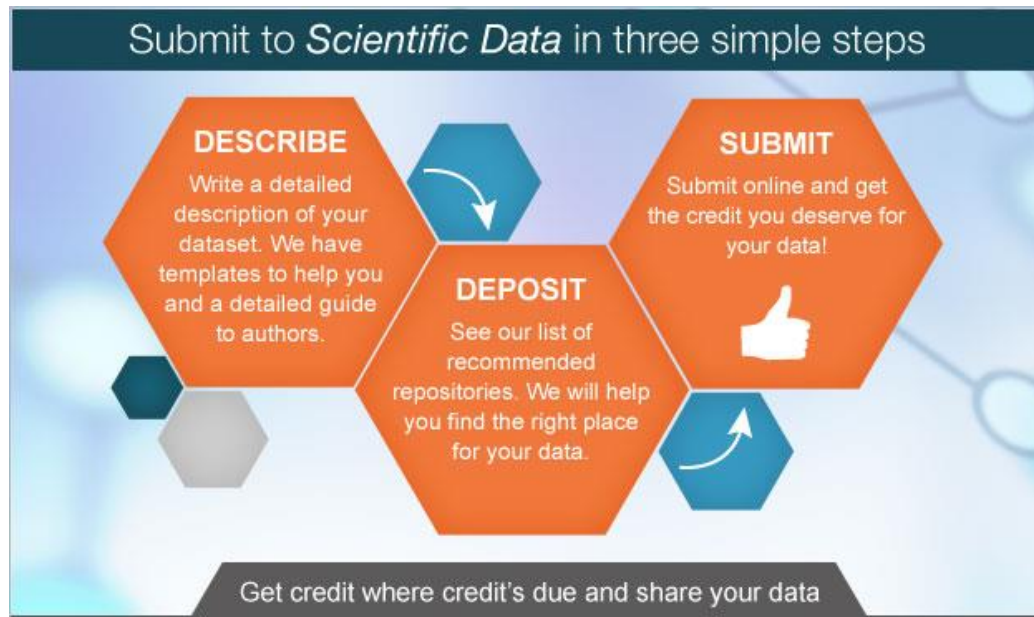
CESSDA data repositories

- Accessible and protected when needed
- Comprehensibility
- Findability and visibility
- Accessibility and reusability
- Longevity
- Quality



Data journals

- Publish a detailed journal style article describing the data and how it was collected
- Recommends or provides a place of deposit
- e.g. Nature Scientific Data (<http://www.nature.com/scientificdata/>)



Citing data

- Citation a fundamental part of research and academia in general
- Just as articles are cited, data which has contributed to research should be cited
- Example:

University of Essex. Institute for Social and Economic Research and NatCen Social Research, *Understanding Society: Waves 1-3, 2009-2012* [computer file]. *5th Edition*. Colchester, Essex: UK Data Archive [distributor], November 2013. SN: 6614, <http://dx.doi.org/10.5255/UKDA-SN-6614-5>



Connecting research and researchers

- Research in the digital realm is becoming increasingly linked up
- Leverage this to increase your profile
- Get an **ORCID** (Open Researcher and Contributor ID) and identify yourself as a unique researcher
- ORCID provides a persistent digital identifier that distinguishes you from every other researcher i.e. *that* Dr. John Smith
- Looks something like: <http://orcid.org/xxxx-xxxx-xxxx-xxxx>
- Simple and free to register at: <http://orcid.org/>

ORCID

Connecting Research
and Researchers

UK Data Service



Our data management guidance

- Online best practice guidance: ukdataservice.ac.uk/manage-data.aspx
- [Managing and Sharing Research Data – a Guide to Good Practice](#): (Sage Publications Ltd)
- Helpdesk for queries: ukdataservice.ac.uk/help/get-in-touch.aspx
- Training: www.data-archive.ac.uk/create-manage/advice-training/events

The screenshot displays the UK Data Service website interface. At the top, there is a navigation menu with links for 'About us', 'Get data', 'Use data', 'Manage data' (which is highlighted), 'Deposit data', and 'News and Events'. Below the navigation, the page title is 'Home > Manage data'. The main heading is 'Prepare and manage data'. A quote reads: "Good data habits from the moment you start planning your research". To the right of the quote is an illustration of three interlocking green gears. Below the quote, there is a 'SHARE' button with a left-pointing arrow. The text below the quote states: "Data created from research are valuable resources that can be used and re-used for future scientific and educational purposes." It then explains that good data management practices are essential for high-quality, well-organized, documented, preserved, and accessible research. It also notes that data management should be planned early in the research cycle and that roles and responsibilities should be explicitly established. On the right side of the page, there is a search bar with the text 'DISCOVER UK DATA SERVICE' and a 'GO' button. Below the search bar, there are two tabs: 'Data' and 'Website'. A 'LOGIN / REGISTER' button is located above the search bar. At the bottom right, there is a 'QUICK ACCESS TO' section with a link for 'FAQ about managing data' and a 'RELATED LINKS' section with a link for 'UK Data Archive'.



UK Data Service



Questions

Veerle Van den Eynden

veerle@essex.ac.uk

