



## **Implementation of the Data Seal of Approval**

The Data Seal of Approval board hereby confirms that the Trusted Digital repository DataFirst Data Portal complies with the guidelines version 2014-2017 set by the Data Seal of Approval Board.

The afore-mentioned repository has therefore acquired the Data Seal of Approval of 2013 on October 13, 2014.

The Trusted Digital repository is allowed to place an image of the Data Seal of Approval logo corresponding to the guidelines version date on their website. This image must link to this file which is hosted on the Data Seal of Approval website.

Yours sincerely,

The Data Seal of Approval Board

## Assessment Information

Guidelines Version:	2014-2017   July 19, 2013
Guidelines Information Booklet:	<a href="#">DSA-booklet_2014-2017.pdf</a>
All Guidelines Documentation:	<a href="#">Documentation</a>
Repository:	DataFirst Data Portal
Seal Acquiry Date:	Oct. 13, 2014
For the latest version of the awarded DSA for this repository please visit our website:	<a href="http://assessment.datasealofapproval.org/seals/">http://assessment.datasealofapproval.org/seals/</a>
Previously Acquired Seals:	None
This repository is owned by:	<b>DataFirst, University of Cape Town</b> School of Economics Building Middle Campus University of Cape Town Rondebosch 7700 Cape Town Western Cape South Africa  T +2721 6505708 E <a href="mailto:info@data1st.org">info@data1st.org</a> W <a href="http://www.datafirst.uct.ac.za/">http://www.datafirst.uct.ac.za/</a>

# Assessment

## 0. Repository Context

### Applicant Entry

#### *Self-assessment statement:*

DataFirst is a Data Service based at the University of Cape Town, South Africa. Our data service accepts survey and administrative microdata (data at unit record level) from African governments and research projects and gives researchers access to this data. Our curation procedure is as follows:

1. DataFirst accepts the data (Submission Information Package) from data producers.

2. Data assurance: This includes

2.1 Data quality checks - errors are tagged and corrected in consultation with data depositors and data quality notes recorded in the metadata. Corrections/updates to data are reflected by versioning of datasets. We version at data file level.

2.2 Disclosure control - direct identifiers are removed and indirect identifiers recoded - see <http://www.datafirst.uct.ac.za/images/docs/datafirst-disclosure-control-flowchart.pdf>

3. Data description: DDI compliant metadata is created for each dataset, and can be viewed on our data portal <http://www.datafirst.uct.ac.za/dataportal/index.php/catalog/central>

4. Data archiving: An archival copy of each dataset (Archival Information Package) is preserved on a secure server at the University of Cape Town. All iterations of each dataset (versioned) are maintained on the server.

5. Data dissemination:

The final dataset for sharing (Dissemination Information Package) is made available either as public use or licensed use data or as secure data:

5.1 We share public use and licensed use data online via our data portal  
<http://www.datafirst.uct.ac.za/dataportal/index.php/catalog/central>

We assist researchers to use the data via our online helpdesk  
<http://www.datafirst.uct.ac.za/services/online-helpdesk>

We offer formal training courses in microdata analysis.

DataFirst also trains African data managers in microdata curation and we work with African microdata producers to improve the quality of their data products through the OECD funded Accelerated Data Program  
<http://adp.ihsn.org/>

We provide official data from Statistics South Africa <http://beta2.statssa.gov.za/> and have a Service Level Agreement to make the microdata from the National Income Dynamics Survey <http://www.nids.uct.ac.za/> available for research purposes

5.2 We share confidential or sensitive data with accredited researchers through our Secure Data Service  
<http://www.datafirst.uct.ac.za/services/secure-data-services>

View our data curation model at  
<http://www.datafirst.uct.ac.za/images/docs/8-20131024-datafirst-data-curation-model-v4.pdf>

## Reviewer Entry

*Accept or send back to applicant for modification:*

Accept

*Comments:*

**Data Seal of Approval Board**

W [www.datasealofapproval.org](http://www.datasealofapproval.org)

E [info@datasealofapproval.org](mailto:info@datasealofapproval.org)

**1. The data producer deposits the data in a data repository with sufficient information for others to assess the quality of the data, and compliance with disciplinary and ethical norms.**

*Minimum Required Statement of Compliance:*

3. In progress: We are in the implementation phase.

**Applicant Entry**

*Statement of Compliance:*

4. Implemented: This guideline has been fully implemented for the needs of our repository.

*Self-assessment statement:*

Ethical Norms:

DataFirst has an agreement with the South African government's National Statistics Office, Statistics SA, to share official data for research purposes.

Each dataset is provided to DataFirst with a metadata report detailing data collection methods. Data collectors within the National Statistical System are also legally obliged to carry out data collection according to South African legal requirements (SA Statistics Act 1999). Statistics SA also requires users to cite them when repurposing the data. We pass this instruction on in the metadata we provide with their data on our data portal.

We work with other data depositors to ensure the data deposited with us for researchers adheres to legal and ethical standards. Datasets deposited by other research organisations are checked to ensure the research has been approved by ethics committees of the relevant institutions. The archive manager has undertaken disclosure control training at Michigan University and is responsible for ensuring each dataset undergoes disclosure control procedures before it is made available online. Direct identifiers are removed and indirect identifiers recoded to ensure there is no disclosure of personal information in the data we distribute. Our disclosure control process is detailed on our website here <http://www.datafirst.uct.ac.za/images/docs/datafirst-disclosure-control-flowchart.pdf> We have a number of datasets which we do not make available because the quality of the data is dubious or the data may be disclosive. High-quality, sensitive data is shared through our Secure Data Service to protect data confidentiality.

Disciplinary Norms:

We undertake data quality checks on all datasets we redistribute. The norms of data quality we work within are related to data accuracy, comparability, relevance, timeliness and interpretability. We provide feedback to data

**Data Seal of Approval Board**

W [www.datasealofapproval.org](http://www.datasealofapproval.org)

E [info@datasealofapproval.org](mailto:info@datasealofapproval.org)

producers on these data quality dimensions, to improve the data. Our work also directly supports data accessibility, another quality attribute.

Communicating metadata to data users:

DataFirst creates DDI-compliant metadata for all data deposited with us. We use methodology reports and other documentation provided with the data to create the metadata, but where data quality information such as provenance and data ownership information is missing we contact the data depositor for confirmation and further details, which we include in the final metadata available online for data users.

Our online helpdesk allows data users to report back to us on missing documents. We pass these queries on to data depositors and work with them to find missing documents and make these available to data users online. We have had variable success with older datasets as data depositing organisations for these may no longer exist. However, researchers on our Data Quality Project and our project examining historical data have visited research institutions and government departments and interviewed retired government employees in attempts to trace valuable information on South African data.

We pass all this information on to data users in the metadata records we create for the data and which is available for each dataset from our data portal <http://www.datafirst.uct.ac.za/dataportal/index.php/catalog/central>

## **Reviewer Entry**

*Accept or send back to applicant for modification:*

Accept

*Comments:*

## 2. The data producer provides the data in formats recommended by the data repository.

*Minimum Required Statement of Compliance:*

3. In progress: We are in the implementation phase.

### Applicant Entry

*Statement of Compliance:*

4. Implemented: This guideline has been fully implemented for the needs of our repository.

*Self-assessment statement:*

Data depositors are provided with deposit information and contact details at <http://www.datafirst.uct.ac.za/services/deposit-data> DataFirst works with depositors to ensure that all data files are included in the deposited dataset.

A list of preferred data file formats is included on this page (Preferred file formats are ASCII, SAS, SPSS or Stata). We try to obtain files in ASCII if these are available, to support future format migration. We also encourage data depositors to provide as much provenance and usage documentation as possible.

However, we do take other file formats. This is because data sharing is still a novel idea in African institutions and we do not want to make depositing data difficult as this will be a disincentive to share data. We also have the skills and software at DataFirst to convert files to preferred formats. We are also in the process of obtaining administrative data from government departments, and converting this to formats suitable for data analysis. In this case the departments deposit data in file formats dependent on the administrative systems they use, and we do the converting. This is aiding our staff to develop knowledge and skills around data conversion, which is a skill in short supply in South Africa, and will be beneficial for our work.

We have been asked to house data for research projects but our policy is that we do not archive data without concomitant permission to pass this on to the research community. This is because our mission is not only data preservation but data sharing to make African microdata widely available to support better research and policy-analysis.

## **Reviewer Entry**

*Accept or send back to applicant for modification:*

Accept

*Comments:*



### **3. The data producer provides the data together with the metadata requested by the data repository.**

*Minimum Required Statement of Compliance:*

4. Implemented: This guideline has been fully implemented for the needs of our repository.

#### **Applicant Entry**

*Statement of Compliance:*

4. Implemented: This guideline has been fully implemented for the needs of our repository.

*Self-assessment statement:*

We work with the data depositor to ensure they provide all documentation for the data, and try and locate key documents where these are missing from older datasets, e.g. codebooks. We have a project to locate and share early South African data and we digitise the documentation for this data and make it available online with the data, e.g. the Manpower Surveys from the 1950s <http://www.datafirst.uct.ac.za/dataportal/index.php/catalog/315>

We create a metadata record for each dataset, using the free, DDI compliant NESSTAR metadata editing software. The metadata is available with the data and documentation for each dataset we disseminate.

We also provide metadata for other African survey datasets, and link to the websites from where the data may be downloaded, e.g. the Afrobarometer survey series [http://www.afrobarometer.org/index.php?option=com\\_content&view=article&id=132&Itemid=80](http://www.afrobarometer.org/index.php?option=com_content&view=article&id=132&Itemid=80)

Metadata and documents for all datasets is available online with the data from our data portal <http://www.datafirst.uct.ac.za/dataportal/index.php/catalog/central>

#### **Reviewer Entry**

*Accept or send back to applicant for modification:*

Accept

*Comments:*

**Data Seal of Approval Board**

W [www.datasealofapproval.org](http://www.datasealofapproval.org)

E [info@datasealofapproval.org](mailto:info@datasealofapproval.org)

#### **4. The data repository has an explicit mission in the area of digital archiving and promulgates it.**

*Minimum Required Statement of Compliance:*

4. Implemented: This guideline has been fully implemented for the needs of our repository.

#### **Applicant Entry**

*Statement of Compliance:*

4. Implemented: This guideline has been fully implemented for the needs of our repository.

*Self-assessment statement:*

DataFirst's Mission Statement appears on our website homepage

DataFirst is a data service dedicated to making South African and other African survey and administrative microdata available to researchers and policy analysts. We promote high quality research by providing the essential research infrastructure for discovering and accessing data and by developing skills among prospective users, particularly in South Africa. We undertake research on the quality and usability of national data and encourage data usage and data sharing

<http://www.datafirst.uct.ac.za/>

The continuity of our work and the preservation of our digital resources is confirmed by being part of the University of Cape Town which demonstrates their support for DataFirst by providing funding for equipment used by the unit. DataFirst has also recently signed an agreement with the University to curate university student data for reuse. The unit has also been involved in writing research data management policy documents for the University. Thus University buy-in insures the future of the digital assets held by DataFirst.

DataFirst's Mission Statement is implemented through the data services we offer and the ongoing data curation and data quality assessments we undertake.

Promotional activities include presentations to Departments and Faculties within the University of Cape Town and to other Universities and Higher Education institutions in South Africa, as well as to government ministries.

#### **Reviewer Entry**

**Data Seal of Approval Board**

W [www.datasealofapproval.org](http://www.datasealofapproval.org)

E [info@datasealofapproval.org](mailto:info@datasealofapproval.org)

*Accept or send back to applicant for modification:*

Accept

*Comments:*

**5. The data repository uses due diligence to ensure compliance with legal regulations and contracts including, when applicable, regulations governing the protection of human subjects.**

*Minimum Required Statement of Compliance:*

4. Implemented: This guideline has been fully implemented for the needs of our repository.

**Applicant Entry**

*Statement of Compliance:*

4. Implemented: This guideline has been fully implemented for the needs of our repository.

*Self-assessment statement:*

DataFirst is a service at the University of Cape Town and therefore needs to comply with both South African data legislation and UCT policies.

Public Use and Licensed Use data incorporate online license conditions to which users must agree in order to download the data (you will see this if you attempt to download any of the microdata). In essence the data requestor agrees to comply with the following conditions:

1. Access is limited to the Lead Researcher and their research team
2. The data will only be used for the purpose stated on the data request form.
3. Data matching intentions must be stated and details provided of the datasets to be matched
4. The Lead Researcher undertakes that no attempt will be made to identify any individual person, family, business, enterprise or organization.
5. The data source will be cited in publications based on the data and a copy emailed to DataFirst

To comply with legal and policy requirements, all data deposited with DataFirst undergoes disclosure control. DataFirst's Manager has been working with data for more than 20 years and has taken a training course in disclosure control at the University of Michigan's Summer School. Other staff are trained in Statistical analysis. Suitable anonymised data is shared as Public Use or Licensed Use Data.

Data that still has some disclosure risk is shared with researcher through DataFirst's Secure Data Service (SDS). The SDS provides potentially disclosive data to researchers in a Secure Research Data Centre <http://www.datafirst.uct.ac.za/documentation/13-sds-brochure/file>

The SDS has several legal documents which need to be signed by the accredited researcher, their organisation, the data depositor and DataFirst. These have been approved by UCT's Contracts and IP Office and are available on the SDS homepage <http://www.datafirst.uct.ac.za/services/secure-data-services>

## **Reviewer Entry**

*Accept or send back to applicant for modification:*

Accept

*Comments:*

## **6. The data repository applies documented processes and procedures for managing data storage.**

*Minimum Required Statement of Compliance:*

4. Implemented: This guideline has been fully implemented for the needs of our repository.

### **Applicant Entry**

*Statement of Compliance:*

4. Implemented: This guideline has been fully implemented for the needs of our repository.

*Self-assessment statement:*

Preservation policy

It is the policy of DataFirst's Data Service to preserve South African socio-economic microdata for the long term. Unit record data on South Africans has value into the future for policy analysis. It thus needs to be preserved in perpetuity, migrated to new formats in a timely manner, and made available on an ongoing basis.

The original data files are preserved unchanged and user-ready versions maintained and shared in 3 different formats (SAS, SPSS, Stata). Documents are stored in original formats and as pdf files. Preservation copies of datasets (data, documents and any accompanying programme files) are dated and versioned and stored on a University Server maintained and backed-up by UCT's ICT Services. Servers are replaced every 3 to 4 years or if they become outdated before this period, with funding provided by UCT. Secure data is kept on a secure server at ICT Services which is not connected to any UCT intranet or the internet. This is backed up manually on a regular basis by senior ICT staff.

#### **File Naming and Versioning**

Versioning and dating of files is used to ensure consistency across archival copies, and to confirm the dissemination copy of the data is the most up-to-date version available. Names of original files received from data producers are retained. The dissemination copy is renamed according to the DDI standard and has version and date information added. Data quality improvements to the file will result in a new file version and date.

NOTE: As DataFirst versions at file level as well as at dataset level the version numbers of the data files of the dataset will not always match. Notes on this should be included in the metadata to explain the difference. The advantage of this is that researchers will not need to download/recheck data files that have not changed.

**Data Seal of Approval Board**

W [www.datasealofapproval.org](http://www.datasealofapproval.org)

E [info@datasealofapproval.org](mailto:info@datasealofapproval.org)

This policy can be accessed on our site at <http://www.datafirst.uct.ac.za/services/data-curation-process>

## **Reviewer Entry**

*Accept or send back to applicant for modification:*

Accept

*Comments:*

## **7. The data repository has a plan for long-term preservation of its digital assets.**

*Minimum Required Statement of Compliance:*

3. In progress: We are in the implementation phase.

### **Applicant Entry**

*Statement of Compliance:*

3. In progress: We are in the implementation phase.

*Self-assessment statement:*

The University of Cape Town recently agreed to take a certain amount of financial responsibility for DataFirst, as a university-wide resource. We are funded by grant money so it is through the university taking greater responsibility for supporting our activities that we aim to provide for sustainability. Data resources are finally being seen as part of the University's intellectual assets, as indicated by our Research Office undertaking policymaking for research data management. Thus it is unlikely that an expert organisation such as DataFirst will not continue to be valued and supported financially by our parent body.

We use non-proprietary software but not all our data are available in ASCII format so data migration is something we currently do to ensure continued access. This applies also to documentation, e.g. we migrated document files from WordPerfect to MS Word using legacy software to ensure usability of these documents

### **Reviewer Entry**

*Accept or send back to applicant for modification:*

Accept

*Comments:*



## 8. Archiving takes place according to explicit work flows across the data life cycle.

*Minimum Required Statement of Compliance:*

3. In progress: We are in the implementation phase.

### Applicant Entry

*Statement of Compliance:*

4. Implemented: This guideline has been fully implemented for the needs of our repository.

*Self-assessment statement:*

The data curation process at DataFirst has been documented on our website and modelled in our data curation model <http://www.datafirst.uct.ac.za/images/docs/8-20131024-datafirst-data-curation-model-v4.pdf>

Disclosure limitation procedures are provided in an online flowchart <http://www.datafirst.uct.ac.za/images/docs/datafirst-disclosure-control-flowchart.pdf>

Data curation and access procedures for our Secure Data Service (SDS) have been written up in a procedures manual which is available on our website <http://www.datafirst.uct.ac.za/images/docs/sds02-procedures-manual.pdf>

Data access procedures for the SDS have also been modelled on our website <http://www.datafirst.uct.ac.za/images/docs/datafirst-sds-flowchart.pdf>

All employees of the University of Cape Town have to have a job-description which is recorded by our HR department. This job description will include an organigram indicating the position of the staff member in the organisational structure. DataFirst also keeps a record of these for our staff.

### Reviewer Entry

*Accept or send back to applicant for modification:*

Accept

*Comments:*

**Data Seal of Approval Board**

W [www.datasealofapproval.org](http://www.datasealofapproval.org)

E [info@datasealofapproval.org](mailto:info@datasealofapproval.org)

## **9. The data repository assumes responsibility from the data producers for access and availability of the digital objects.**

*Minimum Required Statement of Compliance:*

4. Implemented: This guideline has been fully implemented for the needs of our repository.

### **Applicant Entry**

*Statement of Compliance:*

4. Implemented: This guideline has been fully implemented for the needs of our repository.

*Self-assessment statement:*

Ownership of the data does not pass to DataFirst but we have curatorship approval and sharing agreements. We have formal written agreement with South Africa's official data producer, Statistics South Africa, to curate and share their data. We have a Service Level Agreement with the National Income Dynamics Study team to undertake this for their project. We have memorandums of understanding with other data depositors. We always obtain our agreements in writing as the depositors needs to agree at what level they would like their data shared - Public or Licensed Access or Secure Access. This ensures a paper trail for data sharing.

Depositors of secure data will use the data deposit form provided from the SDS site <http://www.datafirst.uct.ac.za/documentation/14-sds-data-deposit-agreement> and will also need to sign their final approval on the researcher accreditation document <http://www.datafirst.uct.ac.za/documentation/15-sds-researcher-accreditation-form>

### **Reviewer Entry**

*Accept or send back to applicant for modification:*

Accept

*Comments:*

## **10. The data repository enables the users to discover and use the data and refer to them in a persistent way.**

*Minimum Required Statement of Compliance:*

3. In progress: We are in the implementation phase.

### **Applicant Entry**

*Statement of Compliance:*

3. In progress: We are in the implementation phase.

*Self-assessment statement:*

The African survey and administrative microdata we provide for research purposes is made available in three different data analysis software formats: SAS, SPSS and Stata. The research community use these software programmes to undertake high-level analysis of the data

Datasets available via our data portal can be searched at dataset or variable level using keyword searches. Dataset searchers use the metadata we create for searching. Variable searches use variable labels for searching. Variable searchers will yield variables covering the search topic in all datasets, and users can select the variables of interest to see summary statistics from these variables before choosing to download the microdata files.

OAI harvesting of metadata files in xml and dbf (Dublin Core) can be done as the data curation software we use for metadata creation and data dissemination is Open Source software with this capability.

The data repository has a web-based interface which allows users to register on the site and access and download data <http://www.datafirst.uct.ac.za/dataportal/index.php/catalog/central>

Currently the url provided for each dataset is a permanent locator up to a point (we can do redirects when we upgrade our software but this is not ideal). We are currently applying for Direct Object Identifiers for all our data and these will be applied to ensure that the links to the data are permanent.

### **Reviewer Entry**

*Accept or send back to applicant for modification:*

Accept

*Comments:*

**Data Seal of Approval Board**

W [www.datasealofapproval.org](http://www.datasealofapproval.org)

E [info@datasealofapproval.org](mailto:info@datasealofapproval.org)

## 11. The data repository ensures the integrity of the digital objects and the metadata.

*Minimum Required Statement of Compliance:*

3. In progress: We are in the implementation phase.

### Applicant Entry

*Statement of Compliance:*

3. In progress: We are in the implementation phase.

*Self-assessment statement:*

We began a process some years ago to use the built-in command in Stata to generate checksums for our data files. Lack of person-power has meant this is still not completed but we are aiming to complete this by the end of 2014. Archive versions of the data are not accessible to users but we want to include checksums in all data files, including those provided to users.

We are the only data distributors in South Africa using versioning. We version according to the Data Documentation Initiative (DDI) data file naming convention. We version at dataset and file level. Version 1 of a dataset is the one we receive from the data producer. Any data quality changes we make will lead to version updates e.g. version 1.1., 1.2 and so on. If the producer recalls the dataset and re-issues this it becomes version 2. Any changes we make will then be 2.1, 2.2, etc. Our data users alert us to data quality issues once the dataset is disseminated and we then work with the producers to fix errors and issue a new version. The metadata will be updated to include change information for the new version, and the metadata itself will receive a new version (in line with DDI standards).

### Reviewer Entry

*Accept or send back to applicant for modification:*

Accept

*Comments:*

## 12. The data repository ensures the authenticity of the digital objects and the metadata.

### *Minimum Required Statement of Compliance:*

3. In progress: We are in the implementation phase.

### **Applicant Entry**

#### *Statement of Compliance:*

4. Implemented: This guideline has been fully implemented for the needs of our repository.

#### *Self-assessment statement:*

Data changes are identified through versioning. Notes on each version and changes to new versions are provided in the online metadata for each dataset available from our data portal. All datasets on our portal include essential documentation and detailed metadata to assist usage.

All data deposited with us undergoes extensive data quality checks for accuracy and usability. Disclosure risk analysis is undertaken for all data. Stata is used as a data management tool to check different versions of the same file, and we isolate issues. For example recently our National Statistics Office released a dataset online and on CD and these had different variable labels, which we drew to their attention.

Data depositors are generally our national data producer or government departments, or survey projects of Universities or Research Institutes. In South Africa the research community is relatively small and DataFirst has been in existence for 11 years so we know most of the data depositors or at least their organisations. We generally work with them to ensure the data is ready for deposit before it is deposited with us and we do not accept online deposits, although we are investigating secure methods of doing this in the future, to save data depositors time and effort.

### **Reviewer Entry**

#### *Accept or send back to applicant for modification:*

Accept

#### *Comments:*

### **13. The technical infrastructure explicitly supports the tasks and functions described in internationally accepted archival standards like OAIS.**

*Minimum Required Statement of Compliance:*

3. In progress: We are in the implementation phase.

#### **Applicant Entry**

*Statement of Compliance:*

4. Implemented: This guideline has been fully implemented for the needs of our repository.

*Self-assessment statement:*

We use the OAIS reference model as our standard. Our adapted OAIS model is available on our website at

<http://www.datafirst.uct.ac.za/images/docs/8-20131024-datafirst-data-curation-model-v4.pdf>

We use the DDI and Dublin Core standards for data documentation and a modified DataCite standard for data citations.

We are continually upgrading our data curation infrastructure and use Open Source software developed by the International Household Survey Network for metadata creation and data dissemination <http://www.surveynetwork.org/> We are the African test site for the software which is developed with funding from the World Bank. We have provided input into the development of the software since its installation in 2010 and recently provided the idea for a citations module which is now a component of the software.

In 2012 we set up a Secure Data Service at the University of Cape Town to provide researchers with sensitive data not previous available to them. Currently we share income and education data through the SDS, which was set up with expert advice from the UK Data Archive. We hope to evolve this service to an online secure service with funding support in the future.

#### **Reviewer Entry**

*Accept or send back to applicant for modification:*

Accept

*Comments:*

**Data Seal of Approval Board**

W [www.datasealofapproval.org](http://www.datasealofapproval.org)

E [info@datasealofapproval.org](mailto:info@datasealofapproval.org)

## **14. The data consumer complies with access regulations set by the data repository.**

### *Minimum Required Statement of Compliance:*

4. Implemented: This guideline has been fully implemented for the needs of our repository.

## **Applicant Entry**

### *Statement of Compliance:*

4. Implemented: This guideline has been fully implemented for the needs of our repository.

### *Self-assessment statement:*

End user licenses are provided online which are signed when data users request the data. These are based on standard end-user agreements and users cannot access the data unless they agree to the terms of the license. The Public Use files and Licensed Data files require agreements to preserve confidentiality of data subjects. However these files are anonymised so this is unlikely. Licensed Data users also agree not to pass on data to third parties.

Sensitive or potentially disclosive data is shared through our Secure Data Service based at the University of Cape Town. Users sign an End-User agreement with the service <http://www.datafirst.uct.ac.za/documentation/16-sds-end-user-agreement> which describes specific penalties for breaches of the agreement.

## **Reviewer Entry**

### *Accept or send back to applicant for modification:*

Accept

### *Comments:*

**15. The data consumer conforms to and agrees with any codes of conduct that are generally accepted in the relevant sector for the exchange and proper use of knowledge and information.**

*Minimum Required Statement of Compliance:*

4. Implemented: This guideline has been fully implemented for the needs of our repository.

**Applicant Entry**

*Statement of Compliance:*

4. Implemented: This guideline has been fully implemented for the needs of our repository.

*Self-assessment statement:*

The data we provide online is anonymised and the public use data files can be downloaded without an agreement. To obtain licensed data the user signs an online usage agreement concerning what they can and can't do with the data. These online licenses are signed by the user when they request the data. The wording of the online license is;

The representative of the Receiving Organization agrees to comply with the following conditions:

1. Access to the licensed data will be limited to the Lead Researcher and other members of the research team listed in this request.
2. Copies of the restricted data or any data created on the basis of the original data will not be copied or made available to anyone other than those mentioned in this Data Access Agreement, unless formally authorized by the Data Archive.
3. The data will only be processed for the stated statistical and research purpose. They will be used for solely for reporting of aggregated information, and not for investigation of specific individuals or organizations. Data will not in any way be used for any administrative, proprietary or law enforcement purposes.
4. The Lead Researcher must state if it is their intention to match the restricted microdata with any other micro-dataset. If any matching is to take place, details must be provided of the datasets to be matched and of the reasons for the matching. Any datasets created as a result of matching will be considered to be restricted and must comply with the terms of this Data Access Agreement.



5. The Lead Researcher undertakes that no attempt will be made to identify any individual person, family, business, enterprise or organization. If such a unique disclosure is made inadvertently, no use will be made of the identity of any person or establishment discovered and full details will be reported to the Data Archive. The identification will not be revealed to any other person not included in the Data Access Agreement.
6. The Lead Researcher will implement security measures to prevent unauthorized access to licensed microdata acquired from the Data Archive.
7. Any books, articles, conference papers, theses, dissertations, reports, or other publications that employ data obtained from DataFirst will cite the source of data in accordance with the citation requirement provided with the dataset.
8. A digital copy of all publications based on the requested data will be sent to DataFirst.
9. The original collector of the data, the Data Archive, and the relevant funding agencies bear no responsibility for use of the data or for interpretations or inferences based upon such uses.
10. This agreement will come into force on the date that approval is given for access to the licensed dataset and remain in force until the completion date of the project or an earlier date if the project is completed ahead of time.
11. If there are any changes to the project specification, security arrangements, personnel or organization detailed in this application form, it is the responsibility of the Lead Researcher to seek the agreement of DataFirst to these changes. Where there is a change to the employer organization of the Lead Researcher this will involve a new application being made and termination of the original project.
12. Breaches of the agreement will be taken seriously and DataFirst will take action against those responsible for the lapse if willful or accidental. Failure to comply with the directions of this agreement will be deemed to be a major breach of the agreement and may involve recourse to legal proceedings. The Data Archive will maintain and share with partner data archives a register of those individuals and organizations which are responsible for breaching the terms of the Data Access Agreement and will impose sanctions on release of future data to these parties.

DataFirst signs data distribution agreements with data depositors and the University's Contracts office will check these as the University takes on joint responsibility with DataFirst for data deposit agreements. For University project we have service level agreements or MOUs. These are not online but can be provided as separate documents.

DataFirst's Manager recently undertook training in Disclosure Control at the University of Michigan and is responsible for information sessions with researchers applying to use the Secure Data Service. DataFirst works with data depositors to investigate confidentiality issues in their datasets and help with anonymisation. A disclosure control flowchart of procedures is available to interested parties at <http://www.datafirst.uct.ac.za/images/docs/datafirst-disclosure-control-flowchart.pdf>

DataFirst's Manager also teaches a data curation module in the Masters in Library and Information Science at the University of Cape Town, which covers disclosure control issues. She also runs workshops in data curation which provide data managers with an overview of data confidentiality issues.

## **Reviewer Entry**

*Accept or send back to applicant for modification:*

Accept

*Comments:*

## **16. The data consumer respects the applicable licences of the data repository regarding the use of the data.**

*Minimum Required Statement of Compliance:*

4. Implemented: This guideline has been fully implemented for the needs of our repository.

### **Applicant Entry**

*Statement of Compliance:*

4. Implemented: This guideline has been fully implemented for the needs of our repository.

*Self-assessment statement:*

Data users sign an online license which lists usage requirements (see above).

Secure Data Service users sign an End-user license which provides data usage requirements.

License agreements deal largely with breaches of confidentiality. Publication of personal information would result in researchers organisations being liable to prosecution but we have not had any cases of this since the establishment of the service so do not know how the censures would be managed in practice.

The Secure Data Service has set penalties for breaches listed in the End-user agreement <http://www.datafirst.uct.ac.za/documentation/16-sds-end-user-agreement>. The majority of the penalties could easily be enforced by the University and the remaining breaches would relate to the Statistics Act and would therefore be handled by Statistics South Africa according to national legislation.

### **Reviewer Entry**

*Accept or send back to applicant for modification:*

Accept

*Comments:*