AFRICA FOR RESULTS INITIATIVE

# GUIDELINES FOR ADOPTING REPRODUCIBLE PRINCIPLES IN EVALUATING DEVELOPMENT INTERVENTIONS IN AFRICA

*From The African Community of Practice On Managing for Development Results at The African Capacity Building Foundation*

Guideline

N° 11

## SYNOPSIS


Development interventions such as the recently adopted Sustainable Development Goals (SDGs), at least to the year 2030 and its predecessor the Millennium Development Goals (MDGs) play a crucial role in motivating world leaders to commit to global development. However, the success in achieving landmark initiatives like the SDGs, to a large extent will depend on country level ability to design, implement and evaluate the performance and achievement of initiatives aligned to the development intervention. The objectives of these guidelines are to a) define the reproducibility principle, b) discuss the various pieces of software that can facilitate the application of reproducibility principles in evaluation contexts, and c) propose the next steps for applying the reproducible principles in the evaluation of development interventions.

This guide is based on a synthesis of existing literature and software packages to demonstrate how the reproducible principles can be used to facilitate effective assessment of the performance of development interventions across Africa. Using information from secondary sources, the paper provides a step-by-step approach for applying reproducible principles, delineates the key capacity requirements that can foster easy adoption and implementation of reproducible principles. The paper demonstrates that the reproducible principles can facilitate consistency in data generation, analysis and coherence in assessment of development performance and achievements of results over a period of time. It also underscores the potential of reproducible principles in improving the design, conduct and reliability of quantitative program evaluation and highlights some of the applicable softwares.

In conclusion, the document alludes to the fact that, although reproducible principles have been around for a while, the concept has not taken root among evaluators. Findings from evaluation studies often lead to recommendations with direct policy implications. Therefore, pushing the reproducibility principles will change how practitioners approach program evaluations and forces them to systematically manage data collection processes, develop rules for archiving data before transforming it, sharing code used for data analysis, and developing and lodging pre-analysis plans before collecting and analyzing data. The main recommendations proffered are that there is a need to build capacities of members of the African Community of Practice on Managing for Development Results (AfCoP/MfDR) about the application of the reproducible principles. Furthermore, the paper suggests developing user friendly software that will facilitate the management of complex workflows, and educate AfCoP members and practitioners on how to explore, adopt and use such approaches and methodologies.

## Introduction

Over the last two decades, there has been an increase in the demand for and investment in capacities aimed at strengthening the conduct and use of evaluations to guide policies and programs across African countries. More and more stakeholders are investing in initiatives aimed at enhancing evaluation capacities of development actors especially the public sector. The effort is partly motivated by the increased appreciation of the value of quality, relevant and timely evidence-based information by program managers and public sector decision makers to improve policy decisions and programmatic choices as well as to enhance the efficiency and effectiveness of development interventions and their impact on the citizenry (McCall 2009). The drive for accountability in development interventions is currently motivated by the call from both the internal and external development stakeholders for good governance including transparency and accountability of governments and public sector office holders as well as duty bearers of the impact of their stewardship and use of financial resources (Chan et al. 2010).

Learning from the data management challenges encountered during the implementation of the Millennium Development Goals (MDGs) (Sanga 2011), this African Community of Practice (AfCoP) knowledge product seeks to provoke evaluators and members of the AfCoP to apply the reproducibility principles when undertaking country level evaluations. The Sustainable Development Goals (SDGs), which were approved in 2015, carry a lot of aspirations in terms of stimulating a more encompassing development strategies that will improve the well-being of all mankind especially the poor and the marginalized (Dereck Osborn, Cutter Amy, and Ullah Farooq 2015). However, the ability to transform the SDGs into real improvements in people's lives depends on the countries' ability to design, implement and assess the progress and achievement of the expected results. This is an element that brings to life the reproducible principles framework as countries are expected to domesticate and monitor the progress and the impact of the SDGs over the next fifteen years.

Therefore, reproducible principles can help improve traditional monitoring and evaluation workflows. The guide outlines details on how to use reproducible principles when conducting development evaluations, presents the software used when undertaking reproducible work, proposes a road-map for enhancing the use of reproducible principles in development evaluations and offers recommendations on the way forward.

## Methodology

The material presented in this guide is based on an extensive review of secondary literature on the role of reproducible principles and how they can be used to enhance the conduct of development evaluations. Among the literature reviewed, Gandrud (2013) provides a comprehensive workflow for undertaking reproducible work, and the software needed to manage the files and organizing analysis code. Kipf et al. (2016) discusses the need to develop platforms that are user friendly for non-ICT[1] professionals in order to integrate data collection and analysis for impact evaluations. The work of Peng (2011) shows how improvements in computer science make it possible for other disciplines to develop new methods for conducting research work and reporting research results and it provides appropriate technologies for implementing reproducible principles in many fields.

## What is the reproducible principle?

The view that "Science advances faster when people waste less time pursuing false leads…. No research paper can ever be considered to be the final word, but there are too many that do not stand up to further study" is widely shared among the scientific community. The reproducibility principle is the idea it is not adequate to publish research findings alone without sharing the raw data and the code used to

[1] Information and Communication Technologies

process the all raw data (Peng 2011). One of the first major reproducible projects was conducted in the field of Psychology. It sought to replicate the findings from more than 100 studies. However, the process of conducting reproducible work was documented as early as 1992 (Kipf et al. 2016). The project faced challenges in accessing data from older studies (Collaboration 2012) as most of the studies for which data was requested, data could not be availed. A similar project in cancer biology found that on average, it took more than two months for authors to avail the data used in their publications (Collaboration 2012).

Reproducibility therefore, is about allowing other people to independently follow through the analysis steps and procedures used in deriving research findings and assess whether a similar result can be obtained (Monya 2016). Replicating studies requires that the initial study carefully documents all the protocols, statistical tests, and analysis plans. The documentation of data, code, clear documentation, and sharing the data is often considered a key benchmark in undertaking reproducible work (Peng 2011). Sharing data generated from research or evaluations is often the most challenging but once clear data access policies are defined, some organizations and researchers end up availing their data online (Vines et al. 2013). Sharing data is always a challenge as discussed earlier as raw data needs to be processed further and anonymized before it is shared. The final version of any dataset will then be accompanied by the code used to clean the data. While this process could be cumbersome, if done properly, it will increase the credibility of the findings and allow other researchers to verify the pathways used to arrive at conclusions (Wadern 2016).

## Software for conducting reproducible work

The advances in Open Source Software has allowed researchers and scientists to communicate across multiple software platforms and this fostered the spirit of collaboration and reproducible science (Hampton et al. 2015). Some of the key tools for collaborative open and reproducible science are listed in Table 1. Software for conducting reproducible research can be classified in four ways namely: literate programming, version control, complex workflows, and automation of the research process.

Literate programming is a way of automatically generating statistical reports by combining the data, code, and text. Literate documents comprise of steps for preparing the data, computing the results and readable text that describes, explains and presents the analysis (Martinez and Clark 2014). Version control software on the other hand, allows users to track changes to their documents using such software as Github and Gitbucket. The complex workflow software is used to keep track of types and versions of software used in producing research reports. The fourth set of tools allow for automation of the entire research process including the formulation of hypothesis and changes from the initial plan.

Table 1: List of Reproducible research software

| **Name of Software** | **Description** |
|---|---|
| Git and Gitbucket Dropbox Google Drive | This is a file versioning system that can be used to keep track of the changes to the survey tools and analysis plans. |
| Knitr (Xie 2016) | This is a general-purpose tool for dynamic report generation in R allowing users to generate MS Word, HTML and PDF files. |
| Packrat | This is used to manage software dependencies and to ensure that local files will be archived/saved when the software is upgraded. |
| The Reproducible | It combines several packages that ensure that codes are run successfully especially when version of your software is continuously updated. |

| | |
|---|---|
| R Toolkit (RRT)[2] | |

*Source: Compiled by author*

Compared to the various desktop pieces of software described earlier, there are new platforms that offer an integrated environment for managing reproducible workflows. The Open Science Framework is one such platform currently hosting the psychology and cancer replication studies. It offers an easy to use interface that allows researchers to register the research plans and document the entire research process. The visTrails also offers a unified interface for data analysis and visualization interface that seeks to address the problem of integrating varied tools as well as automatically documenting the methods and parameters employed (Koop, Freire, and Silva 2013). Pre-registration of analysis plans is also important in eliminating the biases in results interpretation. Several research sites such as "Experiments in Governance and Politics" also offer platforms for researchers and evaluators to register evaluation plans before conducting their studies. While there are several tools of undertaking reproducible research, these guidelines focuse on software that is used within the R-Statistics programming environment.

## Proposed steps for applying reproducible principles in evaluation of development programs

This section delineates a step-by step approach and guidance on how the principles associated with reproducible research can be implemented in undertaking evaluation studies or used by other target audience. It delineates the key principles and the procedures involved at each stage of the process. As shown in Table 2, the reproducible principles can be incorporated into the ten step traditional workflow for conducting development evaluations as follows:

Table 2: Step for applying reproducible principles in evaluation of development programs

| Step | Procedures |
|---|---|
| STEP 1: Clarify what is to be evaluated | This assesses the evaluability of a program or project<br>- ***No changes required*** |
| STEP2: Engage stakeholders | This stage is intended to stimulate stakeholder buy-in and ownership. It is critically important to ensure that stakeholders are committed to and are likely to use the findings and recommendations from the evaluation<br>- ***No changes required*** |
| STEP 3: Assess the scope and resource requirements | An assessment of the scope of the evaluation and resources required will help determine if the proposed effort and allocated resources are adequate<br>This will also require allocating resources to ensuring that the evaluation is conducted in a reproducible manner. Measures should be put in place to ensure timely collection and analysis of data that is intended to be repetitive overtime (electronic tools where feasible) and achieved/hosted in a secure database. |
| STEP 4: Determine evaluation questions | This process involves identifying and determining the questions that will facilitate the achievement of the goal and specific objectives of the evaluation.<br>***No major changes required as these will be consolidated into an evaluation plan that will be locked or registered on a file versioning system.*** |

[2] https://cran.rproject.org/web/packages/checkpoint/vignettes/checkpoint.html

| | |
|---|---|
| STEP 5: Determine appropriate methods of measurement procedures | This involves defining appropriate methodologies and approaches that will be feasible for securing the required evidence in line with the evaluation objectives. ***No changes required*** |
| STEP 6: Develop an evaluation plan | The evaluation plan details how initiatives will be monitored and evaluated and often provides transparency to stakeholders and funding agencies ***This needs to be lodged into a file versioning system in order to ensure that any changes are tracked.*** |
| STEP 7: Collect data | This process involves data collection, selection and training of enumerators, pilot tests, modifications to tools, analysis of pilot data to further refine tools ***There is need to use electronic data collection tools in order to keep a permanent record of the original data set and a file versioning system that accompany the changes to the tools*** |
| STEP 8: Process data and analyze results | In traditional data collection, this involves monitoring data collection, entering data after collection, cleaning and validating data before analysis. Data is also organized into a format for visualization and archiving. ***There is need to adopt electronic data collection platforms particularly for quantitative surveys. Some softwares allow mobile devices to write the data directly to the database and will not permit edits to the initial submissions. Data needs to be exported into statistical software for analysis. It will be important to also document the steps and procedures used to clean the original data into the final dataset used to derive evaluation findings. In some cases, there is a tendency to delete outliers without documenting them properly. An analysis script will be an important deliverable at this stage in addition to integral survey records on the database.*** |
| STEP 9: Interpret and disseminate results | Step 9 involves sharing the findings with stakeholders and tease out the recommendations. ***If Step 9 is done properly, it will be easier to demonstrate to other stakeholders how data has been used to derive the results. If changes are suggested, it will be easier to reconstruct the data analysis with updated changes. In addition to the evaluation report, the datasets and the analysis code can also be shared - this does not need to be done concurrently.*** |
| STEP 10: Apply evaluation findings | This involves identifying the actionable outcomes and developing action plans. ***This will also involve refining the evaluation methodology and data analysis procedures. This is also a key resource from which other evaluators can learn from and improve on the procedures used to collect and analyze data***. |

*Source: Adapted from: At a glance: The ten steps for conducting an evaluation. Toronto, ON: Queen's Printer for Ontario; 2015.*

The following three key reproducible principles are critical for conducting reproducible development evaluation or research work.

1. ***Sharing the analysis code***

Sharing the analysis code that accompanies a given dataset is critically important for the conduct of reproducible development evaluation or research work. Mandating that the data and code are shared will force individuals and organizations to carefully

document all the procedures used in arriving at the final decisions. This will facilitate further processing of the data when compiling regional and universal indicators. The goal of sharing the code is not only to ensure that it can be re-run and reproduce the same results, but also to ensure that users are able to understand and apply it appropriately (Peng 2011).

*2.* ***Reproducible***

Reproducible refers to the ability to replicate the findings using the available methods, data and code. In computational statistics, the focus is on literate programming, which is a way of connecting the text and analysis code to reproduce a paper or manuscript. This could be a steep curve for the evaluation community. However, having access to the data and code used for the assessment will be a reasonable starting point (Gandrud 2013).

***3. Online Archiving***

The third principle is the establishment of a repository that will facilitate the publishing and dissemination of reproducible evaluations (Peng 2011). This is important considering that some of the data for evaluation will be collected by stakeholders who have not traditionally collected regular monitoring and evaluation data such as private companies.

These three reproducibility principles do not need to be met at once but any effort toward meeting any one will improve the implementation of reproducible evaluations. It will also increase transparency and accountability and minimize the chances of misreporting.

## Conclusions and policy implications

This paper concludes that, reproducibility does not address results accuracy but allows for greater learning among practitioners and leads to improved policy decision making. While reproducible principles offer a unique potential for transforming how evaluations of development programs are conducted and managed, there are limitations that currently exist. Key among them is the ownership of data and the availability of user friendly software that makes it possible to apply reproducible principles in the evaluation of development programs. In order to promote the use of reproducible principles, there is a need for additional partnerships between software developers and evaluators to create user friendly tools to clearly document the evaluation process and archive it in ways that can be beneficial to others. At organizational level, data management and sharing policies need to be upgraded mandating that data is shared with the community of practices such as the AfCoP.

Further, to ensure that development actors are fully empowered and fully equipped to apply the concept of reproducibility and use the findings from such research oriented development evaluations, there will be the need to build their capacity in terms of skills and competencies in data collection, analysis and use of evaluation findings.

It is also essential to advocate and sensitize critical mass of development actors on the relevance and value of reproducible research oriented development evaluations, especially as the continent accelerates the implementation of the Sustainable Development Goals as well as other country specific strategies.

Last but not the least, AfCoP/AfriK4R through its annual conferences and knowledge sharing engagements are called to educate its members on how to explore, adopt and use such approaches and methodologies in enhancing the conduct and use of evaluations for decision making in their respective countries.

## References


Chan, Margaret, Michel Kazatchkine, Julian Lob-Levyt, Thoraya Obaid, Julian Schweizer, Michel Sidibe, Ann Veneman, and Tadataka Yamada. 2010. "Meeting the Demand for Results and Accountability: A Call for Action on Health Data from Eight Global Health Agencies." PLoS Med 7 (1): e1000223. http://dx.plos.org/10.1371/journal.pmed.1000223.

Collaboration, Open Science. 2012. "An Open, Large-Scale, Collaborative Effort to Estimate the Reproducibility of Psychological Science." Perspectives on Psychological Science 7 (6): 657–60. doi:10.1177/1745691612462588.

Dereck Osborn, Cutter Amy, and Ullah Farooq. 2015. "UNIVERSAL SUSTAINABLE DEVELOPMENT GOALS Understanding the Transformational Challenge for Developed Countries." Stakeholder Forum. https://sustainabledevelopment.un.org/content/documents/1684SF_-_SDG_Universality_Report_-_May_2015.pdf.

Gandrud, Christopher. 2013. Reproducible Research with R and R Studio. CRC Press.

Hampton, Stephanie E., Sean S. Anderson, Sarah C. Bagby, Corinna Gries, Xueying Han, Edmund M. Hart, Matthew B. Jones, et al. 2015. "The Tao of Open Science for Ecology." Ecosphere 6 (7): 1–13. http://onlinelibrary.wiley.com/doi/10.1890/ES14-00402.1/full.

Kipf, Andreas, Waylon Brunette, Jordan Kellerstrass, Matthew Podolsky, Javier Rosa, Mitchell Sundt, Daniel Wilson, Gaetano Borriello, Eric Brewer, and Evan Thomas. 2016. "A Proposed Integrated Data Collection, Analysis and Sharing Platform for Impact Evaluation." Development Engineering. Accessed March 16. doi:10.1016/j.deveng.2015.12.002.

Koop, David, Juliana Freire, and Claudio T. Silva. 2013. "Enabling Reproducible Science with VisTrails." arXiv: 1309.1784 [cs], September. http://arxiv.org/abs/1309.1784.

Martinez, Andreas, and Michael Clark. 2014. "Literate Statistical Programming: An Introduction Using R and RStudio." http://www3.nd.edu/~mclark19/learn/RDA_intro.pdf.

McCall, Robert B. 2009. "Evidence-Based Programming in the Context of Practice and Policy." Social Policy Report 23 (3): 3–18. https://sci-hub.io/http://www.srcd.org/sites/default/files/documents/23-3_mccall.pdf.

Monya, baker. 2016. "Over Half of Psychology Studies Fail Reproducibility Test : Nature News & Comment." Accessed March 10. http://www.nature.com/news/over-half-of-psychology-studies-fail-reproducibility-test-1.18248.

Peng, Roger D. 2011. "Reproducible Research in Computational Science." Science (New York, NY) 334 (6060): 1226. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3383002/.

Sanga, Dimitri. 2011. "The Challenges of Monitoring and Reporting on the Millennium Development Goals in Africa by 2015 and beyond." African Statistical Journal 12. http://www.afdb.org/fileadmin/uploads/afdb/Documents/Publications/The%20Challenges%20of%20Monitoring%20and%20Reporting%20on%20the%20Millennium%20Development%20Goals%20in%20Africa%20by%202015%20and%20Beyond%20St12.pdf.

Vines, Timothy H., Rose L. Andrew, Dan G. Bock, Michelle T. Franklin, Kimberly J. Gilbert, Nolan C. Kane, Jean-Sébastien Moore, et al. 2013. "Mandated Data Archiving Greatly Improves Access to Research Data." The FASEB Journal 27 (4): 1304–8. doi:10.1096/fj.12-218164.

Wadern, Schloss. 2016. Reproducibility of Data-Oriented Experiments in e-Science." Accessed March 21. http://www.dagstuhl.de/en/program/calendar/semhp/?semnr=16041.

Xie, Yihui. 2016. Knitr: A General-Purpose Package for Dynamic Report Generation in R. https://CRAN.R-project.org/package=knitr.

## Acknowledgement

This knowledge series is intended to summarize good practices and key policy findings on managing for development results. AfCoP-AfriK4R Knowledge products are widely disseminated and are available on the website of the Africa for Results initiative (AfriK4R), at: www.afrik4r.org/page/resources .

This knowledge product is a joint work by the African Capacity Building Foundation (ACBF) and the African Development Bank (AfDB). This is part of the knowledge products produced by ACBF under the leadership of its Executive Secretary, Professor Emmanuel Nnadozie.

The product was prepared by a team led by the ACBF's Knowledge, Monitoring, and Evaluation Department (KME), under the overall supervision of its Director, Dr. Thomas Munthali. Within the KME Department, Ms. Aimtonga Makawia coordinated and managed production of the knowledge product while Dr. Barassou Diawara, Mr. Kwabena Boakye, Mr. F.Thoto and other colleagues provided support with initial reviews of the manuscripts. Special thanks to colleagues from other departments of the Foundation who also supported and contributed to the production of this paper. The ACBF is grateful to the Africa Development Bank which supported production of these guidelines under grant number 2100150023544.

ACBF is also immensely grateful to Mr. Shylock Muyengwa, as the main contributor, for developing this publication. We also thank Prof G. Nhamo, Dr. Massoud Moussavi and Dr. Kirenga whose insightful external reviews enriched this product. The Foundation also wishes to express its appreciation to AfCoP members, ACBF partner institutions, and all individuals who provided inputs critical to completing this product. The views and opinions expressed in this publication do not necessarily reflect the official position of the ACBF, its Board of Governors, its Executive Board, or that of the AfDB management nor its board.