

# Concept Of The African Open Data/Open Science Platform

## 1. Introduction

This document is a brief summary of the rationale, objectives and specific actions involved in the creation of an African Open Data/Open Science Platform, based on agreements between and initial funding from the International Council for Science (ICSU), its Committee on data for Science and Technology (CODATA) and the South African Government's Department of Science and Technology (DST), as well as the support and commitment of Science International<sup>1</sup>. It represents the first substantive step in a major initiative of ICSU and CODATA to create regional open data/open science platforms designed to maximise the benefits of the digital revolution for science, societies and economies.

## 2. Opportunities and Challenges of the Digital Revolution - the Rationale for Action

The digital revolution that kicked off in the first decade of the 21<sup>st</sup> century is a world-historical event that is at least as important, and more pervasive, than the invention of printing. It is based on the enormous growth of the capacity of digital systems to acquire, store, analyse and instantaneously communicate data and information, and to do so at continuously reducing cost (For example, the human genome was first sequenced in 2003. It had taken 10 years and cost \$4 billion. It now takes 3 days and costs \$1,000). It has created enormous potential for scientific discovery, for reducing costs and increasing efficiency in business and public administration, and new and powerful ways of addressing issues such as agricultural productivity, disease tracking and control, and the production of bio-fuels etc. These opportunities are not primarily a consequence of the availability of a great quantity and diversity of digital data, but lie in their creative use in ways that are statistically valid and utilise high levels of scientific comprehension and competence. More specifically, they lie in:

- Exploiting 'Big Data', where enormous fluxes of data stream into computational devices from a diversity of sensors and sources; '
- Linked Data', where semantic links between datasets open opportunities to eliciting deeper meaning;
- The myriad opportunities that arise from blending the physical and digital realms in the 'Internet of things';
- In text and data mining as means of accessing the whole historical record of scientific discovery, exemplified by current exploration of prior knowledge of the Ebola virus; and
- In the capacity of "Learning Machines" to undertake complex tasks, and the ways that are changing the face of scientific discovery and what it means to be a researcher in the 21<sup>st</sup> century.

But the digital revolution not only opens novel opportunities for benefit, it also challenges the fundamental process that has made science the most effective way of generating reliable knowledge. Since the 17<sup>th</sup> century, a fundamental axiom of science has been that the evidence (the data) that underpins a published scientific claim must be made concurrently open. This permits others to scrutinise the logic of the claimed relationship between evidence and concept, to repeat observations and experiments and to replicate results. It is the basis of "self-correction" that exposes error. The digital data deluge challenges this principle through the technical difficulties of making large and complex data volumes effectively scrutinisable though "intelligent openness"<sup>2</sup>, together with the metadata and code. Unless national science systems support and enable their communities to achieve re-defined standards of reproducibility, the work done by their scientists will fall far short of acceptable international

---

<sup>1</sup> Science International is an annual action-oriented meeting of the four top-level representatives of international science (the International Council for Science – ICSU, the Inter Academy Partnership – IAP, The World Academy of Sciences – TWAS and the International Social Science Council – ISSC) that is designed to address major global issues of policy for science.

<sup>2</sup> Royal Society, 2012. Science as an Open Enterprise. <https://royalsociety.org/topics-policy/projects/science-public-enterprise/report/>

standards.

Open data is therefore not only a fundamental enabler in exploiting opportunities whose potential benefits to economies and societies, both nationally and internationally, are no longer in doubt<sup>3</sup>, but it is also a basic requirement if acceptable standards of scientific rigour and quality are to be maintained.

### 3. The Priorities for Action: From Open Data to Open Science

The route to maximal exploitation of the benefits of an open data regime depend therefore on two fundamental pillars: having open access to data, and in having the competence and capacity to use them through the approaches noted above. However, scientific, economic and social exploitation of the digital revolution depend not only on developing the technical capacities of scientists and practitioners across all sectors, but also upon the extent to which national systems of scientific priority setting, funding and institutional research management respond in a concerted way to the challenge. Although many countries are now responding to these opportunities, it is essential that developing and least developed countries, which have much to gain from the data revolution but which typically have poorly resourced national research systems, do not fail to respond to this imperative, which is also vital for the attainment of the UN Sustainable Development Goals. If they cannot participate in research based on big, linked and open data, the gap could grow exponentially in coming years. They will be unable to collect, store and share data effectively, unable to participate in the global research enterprise, unable to contribute as full partners to global efforts on climate change, health care, and resource protection, and unable fully to benefit from such efforts, where global solutions will only be achieved if there is global participation. Thus, both emerging and developed countries have a clear and direct interest in helping to fully mobilize the science potential of least developed countries and thereby to contribute to achievement of the UN Sustainable Development Goals.

However, the practice of open data is not only of great benefit in itself, but when coupled with open access publication and ubiquitous communication, science is enabled to reach out to other public and private stakeholders, citizens and political actors in the joint, trans-disciplinary, creation of actionable knowledge, in a process that we term “open science”<sup>4</sup>. It is a necessary contribution to linking and integrating the perspectives of different knowledge subcultures in addressing social complexity and supporting collective problem-solving. In such trans-disciplinary research, scientific knowledge ‘producers’ cease to think of knowledge ‘users’ as passive information receivers, or at best as contributors of data to analyses framed by scientists. Instead, scientists integrate the concerns, values, and worldviews of policymakers and practitioners, of entrepreneurs, activists and citizens, giving them a voice in developing research that is compatible with their needs and aspirations. Open data is a necessary, but not sufficient basis for this development.

### 4. The function of an African Open Data/Open Science Platform

An effective and efficient open data capacity depends not only on the hard, physical infrastructure for acquiring, storing, manipulating, communicating and sharing data and information. It also depends, crucially, on the ‘soft’ infrastructure of protocols and procedures at all levels in the science system that provide support, advice, management and incentives to ensure that the research system operates in a way that permits it to achieve the objectives described above. The African Open Data Platform is conceived as an integrated set of arrangements that provide this hard and soft infrastructural frame.

Such a platform must be adapted to the fact that science is an international enterprise that requires common principles and standards, such as those recently advocated in Science International’s *Accord on Open Data in a Big Data World*<sup>2</sup>. It must also be adapted to the reality that science is done within

---

<sup>3</sup> Open Data in a Big Data World. Science International, 2015. Extended and Abbreviated Versions. <http://www.icsu.org/science-international/accord>

<sup>4</sup> e.g. Hackmann, H. and Boulton, G.S. 2015. Science for a sustainable and just world: a new framework for global science policy? UNESCO Science Report 2015 – Towards 2030.

national systems of priority setting, funding and institutional ecology, involving government, the funders of research, universities and institutes that perform research and scientists themselves.

The multi-state, regional rationale arises from the experience of increasing collaboration between African states through the African Union and other institutions, and in many common problems and priorities. In this context, the African Platform will be a basis for shared investment in infrastructure. It will harvest and circulate good ideas, spread and support good practice and develop the capacities of individuals and institutions. It will promote key applications of relevance to African economies and societies. It will also act as a conduit for links with international open data and open science programmes and standards that will be vital if it is to flourish.

A number of open data activities are currently underway across Africa, including major science capacity-building efforts of CODATA and collaborators such as the Research Data Alliance (RDA) and TWAS, but a great deal is to be gained if, in the context of developing inter-regional links, these activities were to be coordinated and developed in ways that enhance capacity at the level of individuals, institutions and national systems. The purpose is to put the national partners in the African Platform in positions where major tangible gains can be realised from enhanced capacities, which in turn will determine where increased infrastructural investments can be most profitably made.

However, an important longer-term objective of the Platform is to create the basis for what is referred to above as Open Science. It is anticipated that this will become a priority as the Open Data Platform is robustly established, and thereafter also develops as the basis for Open Science in Africa. Such a development will also require convergence and coordination with processes of open government data, such that processes of interoperability will need to be considered at an early stage.

## **5. Embedding an Open Data/Open Science Platform in a global context**

The platform will also act as a conduit for links with international open data and open science programmes and standards, particularly through ICSU and CODATA and in collaboration with the RDA. A successful initiative would be part of an emerging global network of open data/open science platforms. An effective Open Data Platform has been operating for several years in Australia. A coordinated Open Science system is being planned by the European Union. A National Open Research Data Forum is operating as a coordinating body in the UK. Discussions are under way with South American and South East Asian bodies about the creation of regional Open Data/Open Science Platforms based on the model described above. The African initiative is strongly supported by the international CODATA community, CODATA national committees, ICSU's Regional Office for Africa, as well as the networks of other Science International partner organisations, the ISSC, IAP and TWAS.

## **6. Governance of the African Open Science Platform**

The governance of the platform is through:

- a) A high level, representative Advisory Council to advise on the trajectory and priorities of the Platform and the development of common, platform-wide priorities from national objectives and needs.
- b) A Technical Advisory Group to advise on technical priorities and processes, comprising both regional and non-African experts.
- c) Management of processes will be in the hands of a Platform Office located in the Academy of Science of South Africa (ASSAf), coordinated with the ICSU/CODATA Office in Paris.

## **7. Phased Programme**

### **2016**

1. Creation of an Advisory Council and Technical Advisory Group.
2. Mapping the African Open Data scientific landscape of policies, projects, institutions and individuals.

3. Formal consortium agreements, including with the African Union (AU) and the New Partnership for Africa's Development (NEPAD).
4. Launch of the Platform at the 2016 South African Science Forum (December 8).

**2017**

5. Agreement on regional open data policies that are consistent with the needs of its partners.
6. Coordination of frameworks of incentives for open data.
7. Planning and implementing regional and national open data training and skills initiatives that enhance, coordinate and build on existing CODATA/RDA initiatives.

**2018**

8. Further development of training initiatives, including at all levels of national science systems.
9. Collaborate with international scientific unions and their regional chapters in developing discipline specific training.
10. Provision of fellowships and scholarships through international partnership and through TWAS schemes.
11. Development of an open science concept that builds on open data in a way that is adapted to African needs and perspectives in collaboration with other stakeholder groups.
12. Creating a co-designed, regional open data infrastructure roadmap.

**2019**

13. Rolling out an African open data infrastructure initiative based on the roadmap.
14. Co-Design and implementation of flagship data intensive research projects that are highly Africa-relevant to test and demonstrate the benefits of open data in an open science context and the impact of the platform in the regional context.
15. Implementation of Open Science networks.