2018

The African Open Science Platform: The Future of Science and Science for the Future

African Academy of Sciences

Academy of Science of South Africa

Committee on Data for Science and Technology

International Council for Science

National Research and Education Networks

See next page for additional authors

Authors

African Academy of Sciences, Academy of Science of South Africa, Committee on Data for Science and Technology, International Council for Science, National Research and Education Networks, Research Data Alliance, South African Department of Science & Technology, National Research Foundation, Square Kilometre Array, and UNESCO

THE AFRICAN OPEN SCIENCE PLATFORM
THE FUTURE OF SCIENCE AND SCIENCE FOR THE FUTURE

The African Open Science Platform
The Future of Science and Science for the Future

LAYING THE FOUNDATIONS FOR THE AFRICAN OPEN SCIENCE PLATFORM

In late 2016, a preparatory pilot for the Platform was begun, funded by the South African government's department, Science and Technology (DST), through the National Research Foundation (NRF), the International Science Council (ISC) and its Committee on Data for Science and Technology (CODATA); and managed by an African Open Science Platform (AOSP) Office hosted by the Academy of Science of South Africa (ASSAf) under direction of CODATA. Its priorities are to:

a) Put in place the foundations for the AOSP, comprising:
   • guidance on data policies;
   • guidance on incentivising Open Science and overcoming barriers;
   • a curriculum framework for data skills and competencies, tested and adapted to African circumstances; and a roadmap for technical requirements and infrastructure.

b) Map the current landscape of data/science initiatives in Africa.

c) Build a pan-African Open Science community and encouraging the formation of national Open Science fora.   A notable and crucial success of the pilot in this regard has been the development of an African community of practice and support. The strategy has addressed both national entities (e.g. Ministries of Science, ITCs or Environment, national councils for science and technology, data-gathering initiatives and universities) and pan-African or regional bodies, including the African Academy of Sciences (AAS), the Network of African Science Academies (NASAC), the Association of African Universities (AAU), the Regional Universities Forum for Capacity Building in Agriculture (RUFORUM), the West and Central African Research and Education Network (WACREN) and the UbuntuNet Alliance, a regional association of National Research and Education Networks (NRENs) in Africa. Meetings have included participants from Botswana, Burkina Faso, Cameroon, Côte d'Ivoire, Ethiopia, Ghana, Kenya, Madagascar, Morocco, Nigeria, South Africa, Tanzania, Uganda and Zimbabwe. Institutions involved in discussions include AAS, the AAU, and UbuntuNet Alliance and the NRENs.

ADVISORY COUNCIL

Dr Khotso Mokhele (Chair)
Chancellor of the University of the Free State, South Africa

Dr Ismail Serageldin
Founding Director of the Bibliotheca Alexandrina, Egypt

Prof Ogunlade Davidson
Former Minister of Electricity and Water Resources, Sierra Leone

Dr Heide Hackmann
Executive Director, International Council for Science, France

Ex Officio
Prof Geoffrey Boulton (President, CODATA)

Dr Simon Hodson (Managing Director, CODATA)

Mrs Susan Veldsman (Director Scholarly Publishing Program, ASSAf) – AOSP Project Manager

Prof Joseph Wafula (Director: ICT Centre of Excellence and Open Data – iCEOD, JKUAT) – Chair: AOSP Technical Advisory Board (TAB)

TECHNICAL ADVISORY BOARD

Prof Joseph Muliaro Wafula (Chairperson)
Director: ICT Centre of Excellence and Open Data – iCEOD, JKUAT (Kenya)

Mrs Susan Veldsman (Co-Chair)
Director: Scholarly Publishing Program, Academy of Science of South Africa (SA)

Dr Anwar Vahed
Director: DIRISA (SA)

Dr Audrey Masizana
Head of Department Computer Science/ Senior Lecturer/ Researcher, University of Botswana (Botswana)

Mrs Nodumo Dhlamini
Director: ICT Services and Knowledge Management, Association of African Universities (AAU) (Ghana)

Prof Benjamin Aribisala
Dean, Faculty of Science, Lagos State University (Nigeria)

Mr Lazarus Matizirofa
Acting Executive Director, Knowledge Management Corporate (KMC), National Research Foundation (NRF)(SA/Zimbabwe)

Dr Daniel Nyanganyura
Director: ISC Regional Office Africa (SA/Zimbabwe)

Dr Joseph Mwelwa
Founder and Director, Joint Minds Consult (Botswana/Zambia)

Dr Arona Diedhiou
Research Director, Institute of Research for Development at Joseph Fourier University of Grenoble (Côte d'Ivoire/France)

Ms Noasilalaonomenjanahary Ambinintsoa Lucie
Ministry of Environment, Ecology and Forest (Madagascar)

Ms Lynn Woolfrey
Manager, DataFirst (South Africa)

Ex Officio:
Ina Smith (ASSAf), Simon Hodson (CODATA), Nozuko Hlwatika (ASSAf)

Table of Contents

Preface: Status of the Document

This document presents a draft strategy and makes the scientific case for the African Open Science Platform (AOSP). It is based on an expert group meeting held in Pretoria on 27-28 March 2018. Its purpose is to act as a framework for detailed, work on the creation of the Platform and as a basis for discussion at a stakeholder meeting to be held on 3-4 September 2018, which will lead to a definitive strategy for implementation from 2019. Expert group members at the March meeting were drawn from the following organisations: African Academy of Sciences (AAS), Academy of Science of South Africa (ASSAf), Committee on Data for Science and Technology (CODATA), International Council for Science (ICSU), National Research and Education Networks (NRENS), Research Data Alliance (RDA), South African Department of Science & Technology (DST) and National Research Foundation (NRF), Square Kilometre Array (SKA), UNESCO.

Summary

__The African Open Science Platform__. The Platform's mission is to put African scientists at the cutting edge of contemporary, data-intensive science as a fundamental resource for a modern society. Its building blocks are:

- *a federated hardware, communications and software infrastructure, including policies and enabling practices, to support Open Science in the digital era;*
- *a network of excellence in Open Science that supports scientists & other societal actors in accumulating and using modern data resources to maximise scientific, social and economic benefit.*

These objectives will be realised through seven related strands of activity:

__Strand 0:__ Register & portal for African & related international data collections & services.
__Strand 1:__ A federated network of computational facilities and services.
__Strand 2__: Software tools & advice on policies & practices of research data management.
__Strand 3:__ A Data Science Institute at the cutting edge of data analytics and AI.
__Strand 4:__ Priority application programmes: e.g. cities, disease, biosphere, agriculture.
__Strand 5:__ A Network for Education & Skills in data & information.
__Strand 6:__ A Network for Open Science Access and Dialogue.

The document also outlines the proposed governance, membership and management structure of the Platform, the approach to initial funding and the milestones in building up to the launch.
The case for Open Science is based on the profound implications for society and for science, of the digital revolution and of the storm of data that it has unleashed and of the pervasive and novel means of communication that it has enabled. No state should fail to recognise this potential or to adapt their national intellectual infrastructure in exploiting benefits and minimising risks. Open Science is a vital enabler in maintaining the rigour and reliability of science; in creatively integrating diverse data resources to address complex modern challenges; in open innovation and in engaging with other societal actors as knowledge partners in tackling shared problems. It is fundamental to realisation of the Sustainable Development Goals. National science systems worldwide are struggling to adapt to this new paradigm. The alternatives are to do so or risk stagnating in a scientific backwater, isolated from creative streams of social, cultural and economic opportunity. Africa should adapt and capitalise on the opportunities, but in its own way, and as a leader not a follower, with broader, more societally-engaged priorities. It should seize the challenge with boldness and resolution.

Draft Strategy for the African Open Science Platform

1. Vision

African scientists are at the cutting-edge of contemporary, data-intensive science as a fundamental resource for a modern society. They are innovative global exponents and advocates of Open Science, and leaders in addressing African and Global Challenges.

2. Mission

The African Open Science Platform will convene and coordinate the interests, ideas, people, institutions and resources needed to advocate and to advance Open Science in and for Africa. The Platform is:

- a federated system that provides scientists and other societal actors with the means to find, deposit, manage, share and reuse data, software and metadata in pursuing their interests;
- a network providing connective tissue between dispersed actors in pursuit of shared and overlapping Open Science goals by:
  - o supporting scientists in pursuit of the highest levels of excellence;
  - o enabling consortia that wish to utilise powerful digital tools and cutting-edge data science to address important scientific problems;
  - o developing Open Science capacities in individuals and institutions at all levels in both public and private domains;
  - o creating and supporting networks of engagement between scientists and other societal actors in open innovation and in addressing local, national and international issues of major public concern.

3. Activity Strands

There will be two complementary sets of activity: **enabling activities** (strands 0–2) that provide and manage access to data, computational hardware, connectivity and the tools and concepts required for effective Open Science policies and practice; and **application activities** (strands 3–6), directed towards scientific and societal objectives.

ENABLING

3.1   Strand 0: Access to scientific data from both within and beyond Africa, with a minimum of barriers to access.

Activity: The preparatory, pilot study for the Platform has been building up an analysis of data-intensive activity in Africa including data resources. This will be developed into a registry of African data collections and services, as a basis for early partnership development and as content for cloud technologies. Some collections will be ones where Africa could be at the forefront through appropriate focus and funding. Links will also be made to related collections that are inherently international (e.g. environmental sciences, astronomy). This should also be associated with the development of internationally significant, registered and accredited databases (there is currently only one database/data repository accredited with the Data Seal of Approval standard in the whole of Africa).

3.2   Strand 1: Provision of cloud computing facilities that provide networked computation, data access and analysis tools for African science.

Activity: Promoting and realising, together with partners, the creation of a federated network of existing and new facilities as the basis of an African network, and a research cloud offering state-of-the-art, flexible and economically efficient computational capacity, data stewardship and networking services to users.

3.3 Strand 2: Provision of software tools, experience-based advice on research data management and on Open Science policies and practice.

Activity: The principles, policies, practices and data and publication tools that have been found to be of value in maintaining an Open Science environment have been collated during the Platform pilot phase, and will be maintained by access to the CODATA network as international experience evolves. Adaptation to African needs, where necessary, will be managed through a 'Tools Network' which will also organise regional meetings to create awareness of available tools. The national fora currently being created will also be a communication route for the maintenance of best practice in national systems.

DELIVERING

3.4 Strand 3: Create and sustain high level, internationally competitive research capacity in data analytics and artificial intelligence in support of platform science priorities.

Activity: An African Data Science Institute will be created, with a two-fold function:
- to develop African capacities at the international cutting-edge of research in data analytics, artificial intelligence and machine learning, and to develop the careers of individuals who embody them;
- to provide support in the application of these technologies to key research problems that have a global scope but particular relevance to Africa? (See strand 4).
  Box 1 suggests how these two roles might most efficiently be combined. The Institute will play a role of fundamental importance in helping to inform and contribute to the priorities of strands 4, 5 and 6.

BOX 1. Combining cutting-edge science and service delivery

It is notoriously difficult to combine these roles in one institution, involving both a service ethos and a more individualistic, creative ethos. This problem is compounded by the tension between co-location of staff for creative efficiency and their dispersal for service efficiency. An efficient model to overcome these problems could be:
a) For creative efficiency, locate a single Institute hub close to or co-located with a group working in a cognate field. For example, we could approach the leadership of the African Institute for Mathematical Sciences (AIMS) centre about the possibility of co-location.
b) Resolve the science versus service tension by having appropriately skilled members of the Institute co-located with groups working on the thematic programmes described in strand 4 (Box 2). They would couple the expertise of the Institute hub with the questions and problems generated by thematic experts.

3.5 Strand 4: Place African scientists at the international forefront in the application of cutting-edge data technologies to major research domains, and as a fundamental resource for a modern society.

Activity: Create major data-intensive programmes in science areas where Africa is data-asset rich, where domain scientists are active, which are important to African states, and that have the potential to attract collaboration with the best scientific institutions world-wide, a potentially powerful route to development of the African research base. The examples of such programmes as shown in Box 2 would also provide essential underpinnings for several of the SDGs.

BOX 2. Strand 4: Possible priority programmes

The following are examples of potential data-intensive programmes that could fulfill the criteria for strand 4 and for which the Platform could provide a powerful frame for policymakers and public and private users:

Biodiversity

Africa is one of the global hot-spots for biodiversity. Increased understanding of biosphere dynamics in such a setting is a route to understanding how biodiversity might be preserved in the face of human inroads. It could be globally significant. Eleven African countries are voting members with managed nodes in the Global Biodiversity Information Facility (Mauritania, Guinea, Ghana, Togo, Benin, CAR, Uganda, Kenya, Tanzania, Madagascar, South Africa).

Infectious Disease (and non-communicable disease)

Misjudgements in predicting, detecting and responding to infectious disease outbreaks mean that epidemics and pandemics remain among Africa's and humanity's greatest threats. Relevant data remain isolated in silos within the various domain-specific communities and projects. Integrating them has great potential to identify underlying relationships that determine the trajectory of an outbreak and maximise the effectiveness of a response.

Resilient Cities

Rapid population growth and urbanisation continue to diminish the capability of governments to tackle urban development challenges, which are particularly exacerbated by inequality, poverty and uncontrolled development of urban slums. Many cities in Africa are confronted with water shortages, inadequate housing, and increasing rates of air pollution, sanitation, traffic and outbreaks of diseases. Reliable and accurate development data on the underlying distributions, patterns, trends or disparities inherent, are critical for urban planners and governments to effectively make decisions.

Precision Medicine

Targeting disabling diseases which have the greatest individual and societal impacts together with remote diagnosis and the development of local treatment capacities that are well adapted to need are fundamental to the most efficient use of national resources. Appropriate data collection and analysis and the application of learning algorithms to maximise precision offer powerful potentials for the improvement of public health.

Disaster Risk Reduction

Climate-related disasters, such as desertification, famine and flooding have been particularly common in Africa. They have significantly set back progress towards sustainable development in many African countries. Increased incidents of disasters involve new risks and a steady rise over time in disaster-related losses, with a significant economic, social, health, cultural and environmental impact in the short, medium and long term, especially at the local and community levels.

Open Innovation

There is great potential in Africa for open innovation that utilises digitally-intensive tools such as big and broad data, 3D printing, where production is in the hands of the many and not the few, AI technologies and robotics. The Platform could be a means of bringing these tools into the mainstream, so that they can be seized upon by policymakers and businesses to develop distinctively innovative solutions.

3.6 Strand 5: Create understanding and awareness in citizens and professionals about the dynamics of a data – and information-intensive world.

.

Activity: A Network for Education and Skills in Data and Information will be created involving three foci as described in Box 3.

3.7 Strand 6: Develop a community where access to data, information and scientific expertise is enabled and where scientists are engaged in cross-societal collaborations that jointly frame and jointly seek solutions to significant problems.

Activities: Create a Network for Open Science Access and Dialogue and an access portal that stimulates and supports engagement and joint action between scientists and stakeholders in business, government, non-governmental organisations (NGOs), community groups and citizens, open innovation, public policy and data diplomacy. For these purposes it is important that the network adapts to the needs of the communities involved, and is managed to do so.

BOX 3. Network for Education and Skills with focus on:

Schools. Children are growing into an increasingly complex world of data and information. The capacity to negotiate that complexity and to avoid its pitfalls is increasingly important for individual well-being and for development of the judgements required of responsible citizens. The Network will be a source of advice and curricula development in this domain.

Universities. In collaboration with other partners (OECD, Research Data Alliance) CODATA delivers training workshops in data science, which have well-developed curricula. It is important that universities increasingly take over this role by providing data science education as an integral part of Bachelor's programmes as well as at more advanced levels for researchers and research trainees.

Short courses for professionals. There is an increasing demand for in-service short courses in data science and management for professionals from a wide variety of sectors. Such courses will build on the existing CODATA curriculum and could be a significant source of income that would contribute to self-sustenance of the Network.

---

4. Benefits to National Science Systems

Science is an international enterprise in which the major problems that test scientific understanding are shared internationally. However, scientific research is conducted within national systems that have their own policies, priorities and ways of working that depend on national history, culture, geography and economy. The Platform as described above is designed to support both the bedrock of scientific understanding through access to modern data capabilities and their application to national needs and priorities, and to particularly priorities that are shared across Africa (See Box 2). It is conceived as a means of enhancing research support to individuals and institutions at national levels through pan-African collaboration. Although national funding for science varies considerably across Africa, levels are generally very low compared with those in the global north, it is often difficult to generate a critical mass of scientists in any one field, research budgets in universities tend to be relatively small and computational networks are not always adequate for the tasks in hand. The Platform is designed to counteract these tendencies by:

- supporting the creation of a virtual critical mass for groups of scientists with common interests, irrespective of their individual locations;
- federating existing computational resources and networks in financially efficient ways that enhance connectivity and access to computing resources and cloud-hosted data;
- enabling the application of cutting-edge data science to scientific problems through support from an international network of excellence;
- developing trusted data repositories in areas where Africa is data asset-rich;

- promoting major data-rich science programmes in key areas of African concern, in parallel with the development of data repositories as above;
- evolving as a major target for beneficial international collaboration, including support from international foundations and agencies in supporting research and the careers of young scientists;
- recognising that most governments look for enhanced translational value in their research investments, Open Science has a high potential to deliver such value as it is not locked into a small cohort; whether it be in improved public health policy, new business processes, or better community engagement with outcomes;
- recognising that openness buys partnerships, and partnerships deliver more value than closed, local exploitation.

Some scientists and some states are nervous that they will lose more than is gained through openness of the type described here. We argue that a bold and ambitious programme, carried through with energy and purpose, will be a stimulus to national scientific efforts, a basis for beneficial collaboration, greater translational efficiency and a vector for a more powerful African contribution to the international scientific enterprise.

5. Governance and Management

5.1 Governance

The current pilot phase of the AOSP is governed by an Advisory Council responsible for strategic development of the Platform and a Technical Advisory Board responsible for identifying and addressing technical priorities (See page 2). Management of the process is led by ASSAf supported by ISC-CODATA. The membership of these bodies was determined by agreement between DST as funder of the pilot phase, ISC-CODATA and ASSAf.

Moving into the operational phase, the Advisory Council will evolve into a Governing Council with a composition determined by the membership. Rather than a single Technical Advisory Board, it may be preferable to have Advisory Boards or working groups for each of the activity strands, all reporting to the Governing Council and possibly comprising representatives of members that take the lead on developing/implementing the work strands. The Governing Council will appoint a team responsible for management of the Platform and its activities.

5.2 Members

Members of the Platform should be institutions with a pan-African, regional or national remit, interest or responsibility for the design and delivery of the Platform and its activities as described in 3.0. They will invest resources (financial and/or in-kind) and will be engaged in the implementation of particular work strands. Members will determine the governance of the Platform, which would be accountable to them. Members should commit to self-funding participation in annual AOSP membership meetings. They should also do so if elected/appointed to work strand groups or the Governing Council. Examples of relevant members are the Association of African Universities, the African Research Universities Alliance (ARUA), individual research-intensive universities, the African Academy of Sciences (AAS), the Network of African Academies of Science (NASAC), National Research and Education Networks (NRENs), the Square Kilometre Array (SKA) partnership in Africa, the African Institute for Mathematical Sciences (AIMS), Data Intensive Research Initiative of South Africa (DIRISA), the (South) African Research Cloud of the Inter-University Institute for Data Intensive Astronomy (IDIA), the African Observatory for Science, Technology and Innovation (AOSTI) and individual African national academies. Whether individual universities/institutes should be eligible for membership should be discussed further.

5.3 National Forums

These should be a special class of member. They are currently being developed as part of phase 1 as a basis for articulating national priorities and ensuring that national communities are aware of the opportunities offered by the developing

Platform and that they exploit them. They are an important output of the Platform in stimulating creation of national communities and are a key to realising Platform goals at the national level.

5.4   Knowledge Partners

Knowledge partners will be activity-based users of the work strands described in section 3.0. It is a category that will develop as these evolve. For example, the Network on Open Science Access and Dialogue could reach out to NGOs, and the Network on Education and Skills could engage with national educational advisory councils, or even to individual schools, in the provision of curricula. Knowledge partners might also be members if their activities justify this status, and would play a role in discussions about creative development of the Platform.

5.5   Sponsors

The Platform will seek to attract African and international sponsors to:
•   support resource mobilisation;
•   facilitate global networking and profiling of the AOSP;
•   provide advice, and participate in strategy development.
Potential sponsors include the African Union, appropriate regional bodies, the International Science Council (ISC), ISC-CODATA, ISC-WDS (World Data System), RDA (Research Data Alliance), UNESCO, INGSA (International Network for Government Scientific Advice) and the United Nations University.

5.6   Start-up Management

An initial, necessary hub management capacity is estimated as:
•   A Director
•   Four Platform Officers
    o   Data science
    o   Data stewardship
    o   Training and skills
    o   Network building, communications and outreach
•   An administrative officer

The AOSP Office with this staff complement might best be hosted within the International Council for Science Regional Office for Africa (ROA) which could also host a regional CODATA officer.

5.7   Funding

Support for the key strands of the Platform's activities during the start-up and early growth phases will be sought from African and international bodies concerned with capacity development in Africa. They would include the African Development Bank, the World Bank, Donor Aid Agencies and Foundations and Philanthropic bodies. Though some such bodies might be prepared to consider longer-term, activity-based funding for specific strands, it is important that longer-term funding is targeted for the core management function of the platform, infrastructural strands 0–2, and with a combination of core funding and project-based funding for strands 3–6. Discussions should take place with science granting councils in Africa and African Ministries of Science and Education in relation to this longer term.

6. Build-up to Launch

6.1 Principles

Although it is conceived that optimal benefit from the Platform will be delivered when the strands interact strongly, they are also able to act independently, and can therefore be individually launched as and when planning, funding and recruitment are complete. It is however expected that the full range of activities will be launched by early 2020 at the latest.

Although the long-term objective is for the Platform to have pan-African scope, it is anticipated that its initial growth will take place by:
- coordination of effort between institutions in a smaller number of states able and willing to align their scientific data-related policies and practices to create mutual benefit through the functions identified in 3.0;
- agreements for collaboration between existing service providers and coordination their activities to maximise benefit to the African science community.

It is important to recognise that developments can be rapid, and should be wherever possible. India leapt to prominence in IT in the 90s through strong, concerted efforts. There is a great opportunity for Africa to emulate this, as the fixed infrastructure is a comparatively small cost compared to the policy/workforce changes that are necessary for a successful Open Science enterprise.

6.2 Milestones to End of 2018

The roadmap for build-up to the operational launch has the following milestones:

Immediate:
a) Confirmation of the current document as a strategic basis for action.
b) Identification of task groups to further develop the concepts for strands 0–6: function; key institutions and collaborations; funding; planning for operational launch.

**September 3–4, 2018, Stakeholder Meeting to:**
- secure engagement, enthusiasm and commitment from African national and pan-African representative bodies;
- create awareness of the Platform and seek synergy/collaboration/coordination with cognate initiatives;
- stimulate engagement and support from relevant international scientific and inter-governmental bodies and potential funders.

September–November, 2018: Development of a full proposal including agreed collaborations with other providers and determination of the initial group of platform members.

November 5–10, 2018: Presentation of the Platform at International Data Week, Gabarone, Botswana.

December 5–7, 2018: Formal announcement of the operational Platform at Science Forum South Africa 2018.

Through 2019: Intensive work for operational launch of Platform strands.
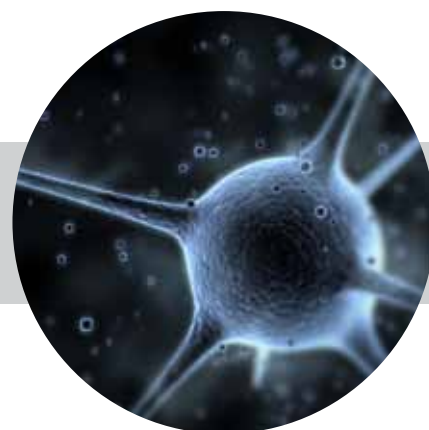
The Case For Open Science

7.   The Digital Revolution

The means by which information and knowledge are acquired, stored and communicated have always been powerful drivers of human material and social progress. The development of moveable type by Gutenberg in the 15th century dramatically reduced the cost of printing and led to the spread of literacy, development of education, stimulation of trade, changes in the governance of states and even influenced religious belief. We are now in the throes of a similar revolution, in which acquisition, manipulation, storage and ubiquitous communication of information using digital rather than analogue technologies has vastly increased the volume and diversity and decreased the cost of acquiring significant data about nature and society. In 2003, it was announced that the human genome had been sequenced for the first time. It had taken ten years and cost $10 billion. Today it takes two days and costs $1 000.

8.   Implications for Society

The power of the digital revolution lies in its potential to uncover hitherto inaccessible patterns in nature and society and to deduce profound relationships, on all scales from the molecular to the cosmic. It enables unprecedented access to data and information by a great diversity of public and private actors in the creation and use of knowledge, and in all areas of human concern, from local health systems to global sustainability. There are few areas of individual, social or political action to which it is not relevant. It offers science the potential to unravel the complexity inherent in many global challenges that have hitherto been beyond our grasp. It prompts new questions, which have always been starting points for new ideas, and elicits new answers. The explosion of digital data has also brought learning algorithms developed in AI laboratories into their own, with the potential to super-charge the digital revolution. Many were created decades ago but could not be fed with sufficient data to realise their potential. But, as with all technologies, there is a down side, in providing tools for cyber-crime, cyber-warfare and political manipulation.

Such are the implications of these developments that no state should be unaware of them, fail to develop relevant individual and institutional capacities in them and to embed them in their national intellectual infrastructure: in education, in the national science base, in public administration and in the private sector. Failure to do so risks falling behind in these domains and in the national economies that depend upon them. A country that fails to develop its own capacities will inevitably become dependent upon skills bought in from elsewhere as a passive and ill-informed consumer of expensive data services.
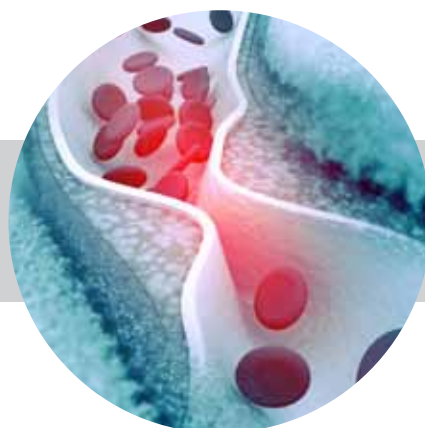
BOX 4. Reality and potential of the digital revolution

In the early stages of application of new technologies it is often difficult to distinguish reality from hype, and it has often taken decades before the society-changing potential of many scientific discoveries has been recognised and realised (electricity, radio, nuclear energy, computers, superconductivity, etc). The same is likely to be true of the technologies of the digital revolution, whose tools include:

- Tools of understanding and analysis: massive fluxes of 'big data'; 'broad data', comprising a diversity of data types that bear on the same phenomenon; and intelligent machines that feed on such data to characterise their environment and act to maximise goals.
- Tools for transmission and communication of data and information.

Powerful applications are being realised and further potentials are being actively explored using these tools, for example:

In Science. They have transformed physics and genomics, are integral to much astronomical research (e.g. the SKA) and are being applied to the discovery and design of new materials. They are being applied to neuroscience, to our understanding of the brain and to modes of cognition. They are essential to modern weather forecasting, to the application of satellite Earth observation in real-time monitoring of many of the key changes in the planetary system and provide GPS facilities for a myriad of applications. They increasingly provide evidence of patterns of human behaviour, with enormous potential for the social and behavioural sciences.

In Public Policy. They have increasing applications in national health systems in understanding patterns of need, resource allocation, identification of risk and in early diagnosis. They are increasingly used in analysing the efficiency with which wide ranges of public services are delivered, including policing, patterns of criminal activity, the operation of energy/water utilities and in efficient administrative procedures.

In Business. They are increasingly important in ensuring efficient business processes, in analysing and addressing markets, in planning distribution, and in the application of scientific discovery to new products and services. There are new, data driven analyses of agricultural planning, whilst the mobile phone has transformed agricultural market places in Africa leading to major changes in agricultural practice.

In Education. This is one of the areas where relatively little change to traditional modes has occurred, though the potential is great to more effectively address the needs of individual students, to provide lifelong learning and to understand and plan for high levels of efficiency and effectiveness.

9. The Open Science Imperative

The digital revolution offers profound opportunities for scientific discovery and for the application of new understanding, for which Open Science is a crucial enabler. This is a powerful paradigm that combines the historical imperative for scientists not only to expose their working, evidence and results to scrutiny by their peers, with the power of digital communication to open them to public access and scrutiny, but also to facilitate engagement with other public actors in collaborative learning and problem-solving. It embraces three fundamental priorities:

9.1 Maintaining Rigour

The edifice of modern science has been built on openness. The requirement that published scientific hypotheses should be accompanied by the evidence (the data) on which they are based has permitted others to scrutinise the logic of the relationship and to replicate the results of experiments or observations. Failure on either count is taken as failure of the hypothesis, and characterised as "scientific self-correction". However, the volume and diversity of modern data

resources, and the complex computational strategies often required to recover deep patterns from them, challenge these principles. First, because of the complexity of making all the data, metadata and computer code available for interrogation. Second, because of the difficulty of tracing the logic of the processing to which raw data is subject in the inference of underlying patterns, with the potential for opaque scientific reasoning, hidden within a black box of machine learning. The importance of these issues for the credibility of the modern scientific enterprise has been highlighted in recent years by demonstrations that the results of otherwise highly regarded papers in several fields can be replicated in only a small number of cases, often because of inadequate metadata or the code used in analysis is absent, unstable or absent. It is vital that scientists invariably publish the data on which a scientific claim is based concurrently with that claim[1]. To do otherwise is malpractice. These are crucial issues for modern science, which must be addressed internationally in developing new, open standards for replication in a data-rich age. A national science system that is unable to achieve these norms will lose credibility and, quite properly, fail to attract investment.

9.2   The Power of Sharing

The digital revolution is not only characterised by formidable fluxes of data but also by a very wide diversity of data streams. Integrating diverse data streams permits us to characterise the behaviour of complex multi-dimensional phenomena, with high resolution in both space and time, with the potential for unprecedented insights into such complex issues such as infectious disease, urban systems and ecological dynamics. They are approaches that will be essential to the success of the sustainable development goals. However, the integration of diverse data that this implies requires easy access to data from a diversity of sources and their re-use. Data should therefore be readily shared using the FAIR standard[2] and as a new universal norm for modern science that requires individual scientists to accept new responsibilities[3].

9.3   Open to All and Societally Engaged

The impacts of modern science and technology on individuals and societies, and in particular the emerging potential of data-enabled artificial intelligence to intensify such impacts, are such that citizens and other societal actors need to be engaged with science in considering societal futures. Democratic societies in particular are best able to govern themselves when individuals and groups within them are equipped with the knowledge required to exercise their responsibilities as citizens in making informed choices. Science now needs to be a public enterprise rather than one conducted behind closed laboratory and library doors, one that engages actively with business, policymakers, governments, communities and citizens. Open Science is not simply a matter of open data and open access publishing, that merely represents science talking to itself, albeit more efficiently. It is also about engaging with other societal actors as knowledge partners in jointly framing questions and jointly seeking solutions. They are processes that are greatly enhanced by the ubiquitous nature and varied modes of modern digital communication, which lend themselves to unprecedented inter-sectoral communication.

The Open Science paradigm also meshes with and is integral to concepts of open innovation, in both society and industry, which are most productive, in a data-intensive era, when there is open access to data and information. At the most fundamental level however, Open Science is a product of and provides inspiration for an Open Society, and is essential to the future of both science and society.

---

[1] Note that relevant data may be that acquired through precisely planned experiments and observations. It may be data collected routinely by government agencies such as weather services or taxation services. It may be derived from the "internet of things", the myriad data generated and collected by a large number of private and public sources, which is of particular potential importance to the social sciences.

[2] Note The developing standard of data publication is FAIR - Findable, Accessible, Interoperable and Re-usable – demanding requirements that are internationally formulated, but that require, if they are to be operationally effective, strong support from institutions, scientific journals and funders.

[3] Note the first principle set out in the Science International Accord, Open Data in Big Data World, endorsed by over 120 major scientific bodies worldwide: "Publicly funded scientists have a responsibility to contribute to the public good through the creation and communication of new knowledge, of which associated data are intrinsic parts. They should make such data openly available to others as soon as possible after their production in ways that permit them to be re-used and re-purposed."

10. Challenges at the Cutting Edge

10.1 Data Integration for Grand Interdisciplinary Challenges

Global society is confronted by multiple, intersecting sets of converging environmental, socio-economic, political and cultural problems, many as consequences of complex coupling between social and bio-geophysical processes. These dynamics that have re-configured the global ecology to produce one which is novel to the Earth and to which social dimensions such as poverty, inequality and conflict are integral as the bio-geophysical. They lie at the root of the sustainable development challenges, which are not only challenges for society, but also profound challenges for science. It is such a nexus of common challenges to which the Open Science paradigm is a response. Such challenges are not posed as problems for a discipline, for physics, psychology or demography, but as ones that can only be resolved through deep interdisciplinary engagement with societal actors. They are an important rationale for the creation of the International Science Council[4].

The above challenges pose problems in which the constituent variables interact in complex ways such that the 'system' has emergent properties and behaviours that cannot be predicted by considering the variables separately. Although we can computationally simulate the behaviour of such complex, coupled systems, we are generally unable to match simulation to any deep empirical patterns that might extend across the relevant data derived from different disciplines. This arises because of the varying data standards and vocabularies that are used across the disciplines, so that online Integration of diverse data from globally distributed sources can often only be achieved within and between closely allied fields. Thus, although we frequently create broad, integrated, interdisciplinary concepts, we are generally unable to support them with integrated, multi-disciplinary data. A major, long-term ISC-CODATA programme has been launched that is designed to overcome this barrier to interdisciplinarity, with the potential to support the research priorities referred to in Box 2.

10.2 Knowledge without Understanding: Data-driven Science and Machine Learning

It is broadly accepted that science seeks understanding of the causes of things that are empirically known through transmitted information, sensory awareness or experience. We are now faced with a novel juxtaposition, where the cause

---

of a known outcome cannot be established. The large and complex datasets created as a consequence of the digital revolution have brought the learning algorithms produced in recent decades by the artificial intelligence community into their own. They have become powerful means of deriving deep empirical patterns from the high volume and high velocity data streams now available, through their capacity to progressively learn and improve performance in deriving patterns in data without being explicitly programmed to seek for a particular pattern. Such a learning process takes place without any knowledge of the underlying science. A learning machine merely applies pattern-learning algorithms to the data fed through it, but it is a process that poses a profound enigma for traditional concepts of scientific method. The outputs of machine learning algorithms can neither be predicted nor explained. They create knowledge without understanding. It is neither clear how scientists will address this, nor how citizens will react to it. An Open Science dialogue could be critical in addressing this dilemma.
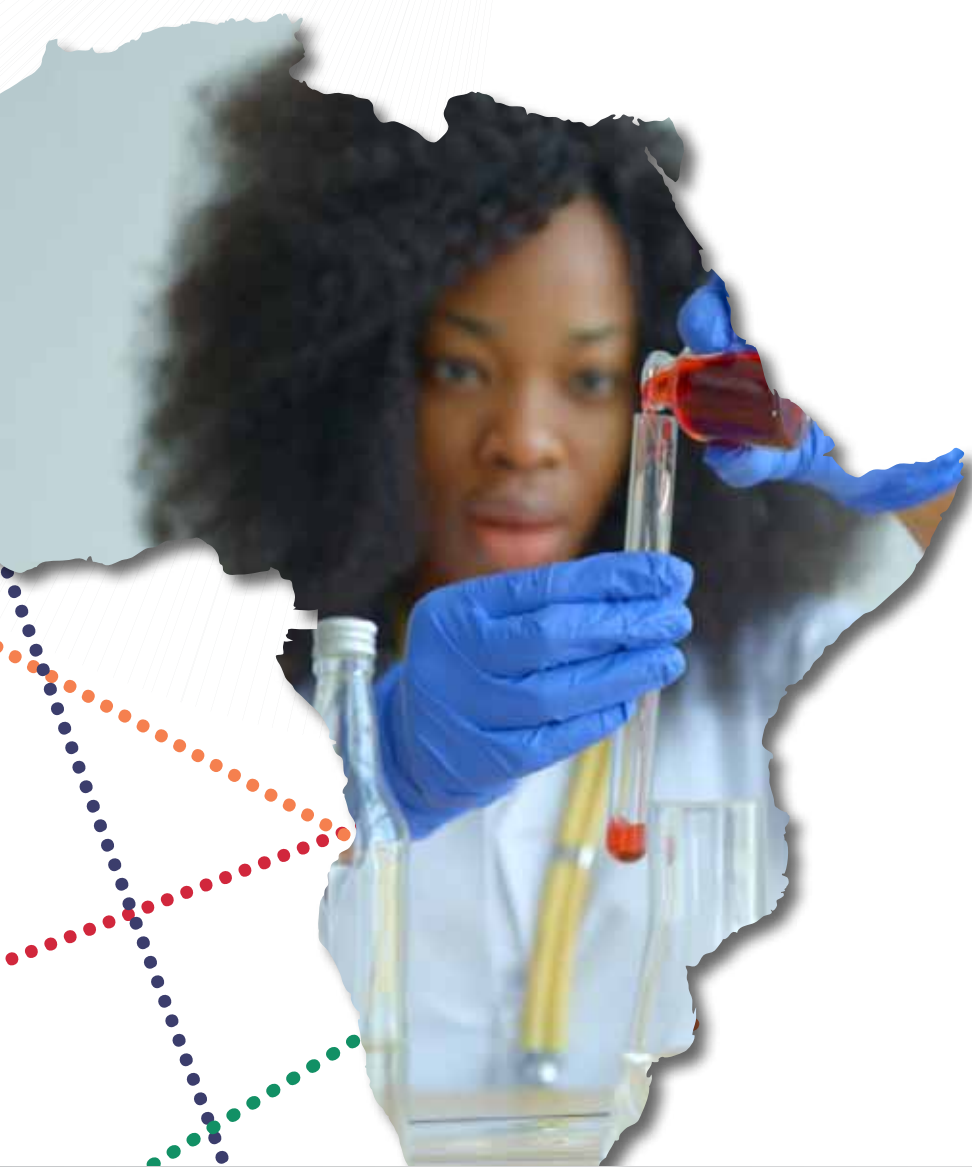
Seizing The Opportunity: The Case For Africa

Open science offers great opportunities for scientific discovery and for social benefit. But this new paradigm also poses challenges for national science systems, for although science is an international enterprise, it is largely done within national institutional ecologies, and with normative habits, incentives and priorities that are adapted to the paper/print technologies of the Gutenberg era, even though modern science predominantly operates with digital tools. Adapting incentives, performance indices, processes of communication, of logic, of scientific analysis, of data and of information management and storage to this new mode makes major demands on scientists, their institutions, managers, funders and science policymakers. Deeper interactions with other societal actors will require changes in outlook and habit, potentially driven by incentives, new funding practices and systems of training. It is also important to recognise and confront the ethical issues that are at stake in the open exchange and use of data, and in respecting the privacy of potential data subjects and observing norms of fairness and attribution of data originators.

Almost all national science systems are struggling to adapt to this new paradigm, with varying levels of investment and energy, and with varying definitions of its scope. Although the emphasis is beginning to change, interventions in Europe and elsewhere have focussed primarily on open data and open-access publishing. In contrast, we take the broader view Open Science presented in paragraph 9.3 as a distinctly African approach. The argument of this paper is that there is no alternative for Africa but to adapt to this new paradigm, but in its own, unique way, as a leader and not a follower, with broader, more societally-engaged priorities, and to seize the challenge with boldness and resolution.

There will of course be fear that an Open Science system will encourage researchers from better-funded research systems to swoop on Africa's open data to scoop the benefit. But we see no other option. For a national research system to set itself against the trend towards Open Science, inevitably limits the resources available to its scientists, shuts them out of international collaboration and loses the potential that strong engagement with an active international science community confers on its members. Indeed, working with the best scientists internationally is probably the most efficient way of improving the excellence of a national science base[5].

In an international setting where Open Science opportunities and priorities are being pursued with urgency in many well-funded science systems, the danger to Africa is that yet another knowledge divide develops, in which it is a relatively weak recipient of innovations forged elsewhere, to its great cost. The proposals for a pan-African Open Science Platform are designed to avoid this outcome, and moreover, not simply to play catch-up, but to leap-frog and take the lead in key areas of development, through a more ambitious approach to Open Science and by exploiting existing areas of excellence and the globally dispersed diaspora of excellent African scientists and a cultural disposition towards greater openness. There are two fundamental building blocks in this pan-African strategy for achieving these objectives:
- implementing the policies and practices and creating the infrastructure to support Open Science in the digital era;
- developing the capacities of individuals and institutions in ways that enable effective exploitation of the potential of Open Science by all sectors of society.

[5] Note Open science approaches also include more open ways of working including the open availability of pre-prints and data papers which in some fields of science have become important ways of establishing primacy and inviting collaboration at earlier points in the research process.

Notes