ACADEMY OF SCIENCE OF SOUTH AFRICA (ASSAf)
COMMITTEE ON DATA FOR SCIENCE AND TECHNOLOGY (CODATA)

Report on the African Open Science Platform

Side Event

WEDNESDAY, 7 DECEMBER 2016

CSIR CONVENTION CENTRE, PRETORIA, SOUTH AFRICA

Table of Contents

Welcome

Ms Susan Veldsman (*Director: Scholarly Publishing Unit, Academy of Science of South Africa (ASSAf)*) opened the proceedings and introduced herself. She reported that the workshop host, ASSAf, was grateful for the participation of all stakeholders in the process of scheduling, planning and road mapping the African Open Science Platform (AOSP). She welcomed all attendees and specifically Prof Geoffrey Boulton, Dr Heide Hackman, Dr Simon Hodson and other CODATA EXCO members who had taken the time to attend the workshop and thanked them for their support, interest in the project and for encouraging the project team to make a success of the project. It was also noted that the 55 delegates in attendance represented countries from all over the world, mainly from Africa (Ghana, Tanzania, Kenya, Zimbabwe, Botswana, Uganda, Zambia, Lesotho and South Africa) but also from India, UK, USA, and Canada. Additional delegates joined the workshop later in the day - from Russia, Israel, China, Nigeria and Liberia.

Introduction

Science International Accord on Open Data in a Big Data World: introduction and context

View presentation: http://www.slideshare.net/AfricanOpenSciencePlatform/african-open-science-platform

Prof Geoffrey Boulton (*President CODATA & University of Edinburgh, UK*) indicated that he would introduce the fundamental concepts of the African Open Science Platform (AOSP) in an international context.

He started off by demonstrating the fundamental reasons for engaging on the issue of Open Science and Open Data (OSOD). In about 2000, digital storage overtook analogue storage, and it has expanded massively since then. One reason for this is the highly flexible way in which the data is contained and in which it can be manipulated, since it is now more affordable than before. For example, the Human Genome Project (HGP) was an international scientific research project with the goal of determining the sequence of nucleotide base pairs that make up human DNA, and of identifying and mapping all of the genes of the human genome from both a physical and a functional standpoint. It remains the world's largest collaborative biological project. After the idea was picked up in 1984 by the US government when the planning started, the project formally launched in 1990 and was declared complete in 2003. At the time the project cost ZAR10 billion and required enormous time (13 years); now it can be done in 3 days for US$10. What is crucial, is that there are now opportunities to understand phenomena in ways they could not be understood before. This has implications for science and government, which supports populations and how the economy operates, etc.

Key issues addressed during the presentation: Why Open Science/Open Data; the international context; what is a platform—how should it be structured and what is its value, how is it governed; the role of CODATA and ICSU. These talk to the issues of: is science/data accessible; can people find it; can people understand it/ is it comprehensible; can people use it; can it be re-used; etc. This is important, as the ability to integrate data from a variety of sources means semantic linking is possible, and therefore also acquiring much deeper learning than before. The scenario will no longer be one of a spreadsheet containing a list of data; instead it will be possible to combine and interlink unlimited data sets.

Society is faced by numerous phenomena related to infectious diseases and climate, environment, mobility, etc. If these various aspects are combined, patterns and relationships will be seen that were

not possible before, and it will lead to a revolution in science. Consider the extraordinary information about the world and that 600 years ago movable type was invented and drove many innovations; now another revolution is possible, i.e. digital storage, acquisition, manipulation, etc. of data. Data must be open – if not, then no-one else can re-use and combine data sets for new learning to occur.

Open data has also created new commercial sectors, e.g. US meteorological data was released for free and young people could use it to support a number of things, e.g. evaluation of crop fertility and where to go on holiday to get some sun. Whole sectors have grown up around open data and the open data sector is growing at a phenomenal rate.

For over ten years, a number of international bodies have argued for data on important international and national issues to be opened, in order to address challenges faced. The *Science International* Accord *on Open Data in a Big Data World* was launched at the Science Forum SA in 2015. This *Accord* provides guiding principles to abide by when using open data. It has been endorsed by over 110 scientific and data bodies worldwide and will increasingly form the framework for programmes such as AOSP. The drivers are either individual disciplines that see opportunities or government/ government agencies.

Examples of discipline-driven data sharing platforms include the European Bioinformatics Institute (EMBL-EBI), International Union of Crystallography, and the work done by ICSU CODATA (ontologies, standards etc.). Examples of government-driven data sharing platforms include the European Science Cloud, the National Cancer Institute Genomic Data Commons, and the Open Data Transition Report. CODATA is considering promoting OSOD at UN level, as it is so important for human society in this decade of data.

Collaboration and combining capacity and funding can speed up discovery and new ideas. Mobility is essential (people and ideas) – joint enterprise means ideas flow easily. We need to keep discovering new ways of doing things, otherwise we become trapped inside our own approach to doing things. This also applies to AOSP. AOSP would function in a similar way than Uber, the worldwide online transportation network, which does not own any taxis and have no employees. Its main function is to connect users with procurers in efficient ways. AOSP will – in a similar way - serve as a platform for use by those who wish to exploit data in effective ways to be put in touch with others; it should be used to maximise data sharing, which is a fundamental target, because of the power of shared data. But the community is now discovering how these things might work and therefore there is no template to use and no rules to follow – these must be invented by the users.

AOSP is not an attempt to displace an existing programme or projects, but an attempt to find the common denominator for effective operations and capacity enhancement of national policies and frameworks. Science is international, but it is done in the context of a national state, which has its own approach, culture, etc. So this adds value and does not displace.

An iceberg in the ocean provides a good metaphor to demonstrate the challenges faced in terms of OSOD: what one can see is the smaller, 'easier' bit above the waterline. This includes technical and consent challenges. The difficult aspects are below the surface and far bigger: processes, organisations and people. Challenges below the surface include:

- Ecosystem challenge: how organisations work together; national policies enunciated by government.
- Funding challenge: funding is required to ensure good science (incl. research & data) practice. If researchers don't bother with data, the result is bad research. Funding is needed to make data accessible. At funding body level, universities and institutes, etc. should orchestrate efforts to ensure they do better together than separately.

- Support challenge: if there is support from other domains, then not everyone need to be a data expert.
- Skills challenge: a challenge for universities, as statistical excellence is less than it was forty years ago. But old statistics are not adequate for the new data regime and there is a need to enhance skills in this regard.
- Incentives challenge: more immediate incentives are required to encourage researchers to openly share data.
- Mind-set challenge: mind-sets must change. Data collected in Antarctica and funded by government should be accessible to experts from all over to use. Not all can go to Antarctica to collect data due to challenges, and the data should be shared with a wider group. Open data is therefore what is recommended. Open science suggests there is a high wall around science that separates it from the rest of the community and that the wall must come down and there must be engagement with society. But open science needs open data.

A platform such as AOSP is therefore much needed to address the mentioned challenges faced on the African continent, and to contribute to international science. The purpose of the platform will be to provide a federated space for scientists to find, deposit, manage, share and reuse data, software and metadata. The platform will have four main functions:

- Establishing common shared open principles (*Accord*), policies and practices for data acquisition and use, and providing the facilitating tools in ways that are adapted to varying national, disciplinary and application priorities and approaches. Novel mathematical and computational technology should be utilised to extract meaning. Machine learning is now being utilised, as machines are able to learn and extract meaning, and the best technology and techniques must be available to users.
- Recognising the roles and developing responsibilities of different actors at all levels in national scientific ecosystems. This includes government, funders, universities and institutes, publishers and the researchers.
- Developing the technical capacities of researchers and data professionals. Programmes and capacity-building workshops will be presented in Africa and elsewhere in collaboration with CODATA and the RDA.
- Creating meaning from data: using and applying analytical tools, machine learning, semantic methods to data integration, ontologies and vocabularies for discovery and integration.

It is important to recognise that collaboration with existing bodies is key to the success of this project. We also need to bear in mind that we should not use 20th century methods to solve 21st century problems.

The SA Dept. of Science and Technology (DST) is funding AOSP through the National Research Foundation (NRF). The project will be managed by ASSAf, and directed by CODATA. The pilot phase priority is twofold: 1) setting up a high-level advisory council, comprising senior people who are able to advise in an Africa setting and how to enlarge the enterprise in an Africa setting; 2) setting up a technically proficient body to advise on how the platform should navigate and what the priorities should be. A key role is to develop partnerships, as indicated. This is to ensure maximum support is delivered to CODATA.

There is a need to support international science and a need for the use of data from Africa by the international scientific community. AOSP will therefore play an international role. Already ideas from the Africa platform have been picked up by the Caribbean and Latin America. This will help push through analogue matters and support development of a powerful South-South relationship.

In closing, Prof Boulton advised that the workshop was intended to obtain input from the participants for the initial phase and how to expand the partnership and deliver benefits.

Questions: None

Before calling on the next speaker, the Chairperson introduced **Dr Kghotso Mokhele**, who will be the Chair of the AOSP Advisory Council. **Dr Mokhele** was also thanked for the support that was being provided. **Ms Ina Smith** was also introduced; she will be responsible for implementing the AOSP.

Data Coordination and the Role of the Research Data Alliance (RDA)

View presentation: http://www.slideshare.net/AfricanOpenSciencePlatform/data-coordination-and-the-role-of-rda

Dr Devika Madalli (*Indian Statistical Institute, Bangalore, India*) introduced the RDA as an international member based organization focused on the development of infrastructure and community activities that reduce barriers to data sharing and exchange, and the acceleration of data driven innovation worldwide. With more than 4,500 members globally representing 115 countries, RDA includes **data science professionals** from multiple disciplines, including but not limited to academia, library sciences, earth science, astronomy and meteorology.

RDA is building the social and technical bridges that enable open sharing of data to achieve its vision of researchers and innovators openly sharing data across technologies, disciplines, and countries to address the grand challenges of society.

There are many working groups and many of these are domain based. Groups include: chemical working group, infrastructure, policy working group and data sharing. The group was alerted that there is overlap with what RDA is doing and what this group intends to do.

RDA functions through two mechanisms: working groups and interest groups. Interest groups are formed when a number of researchers get together and show enough interest to form a working group. They discuss the work that is to be done. An interest group is formed and this spawns a working group, which usually works under the interest group. The working group works on a problem, with people working on common problems of interest and deliverables, e.g. programmes, suggestions. There is always a deliverable and there are already a number of recommendations and products that have been produced by RDA. The website provides data, presentations and more.

Work is done in various areas, including: data preservation, best practice for domain repositories, data citations, etc. These are just some examples of areas engaged in and many more are detailed on the website.

The following principles guide the work that is done and there is consensus on what is done and approved: *Openness, Consensus, Balance, Harmonization, Community-driven, Non-profit and technology-neutral.*

RDA does not fund anyone – all participants pay for their own cost to participate.

It is interesting to note Africa's participation (3% is not tiny) and it should soon surpass Asia (9%), which has participated in all events.

Data Challenges of the developing and third countries

- Lack of scientific data
- Lack of infrastructure
- Knowledge and skill deficit
- Lack of awareness

- Lagging in standards adoption
- IT gaps
- Lack of FUNDING!

The lack of scientific data is a particular challenge when compared to developing countries.

While Bangalore has an IT ambiance, if one excludes the three big cities, the repositories in India – a fast-developing country - are finding it difficult to cope, as there is a problem with infrastructure.

The knowledge and skill deficit shows a particular lack among those who must handle data repositories and this relates to the previously explained iceberg analogy. Third-world countries can relate to this, but the first step to overcoming challenges is to acknowledge that they exist.

Standards are being spoken about at many conferences, but adoption is slow. This is a problem, as people nod their heads at conferences, but when the surveys are done a few months later, the adoption levels are very low.

The IT gap is at the tip of the iceberg, but in India (and other parts of the world), IT reaches people in bits and pieces, sometimes because of the diversity of technology that is available and the lack of people who are able to use it. Here the problem is to find the easiest solution, rather than it being a problem of absent technology.

There is a need for a centrally-managed platform in order to better answer problems.

Collaboration will assist in these problems: Prevent re-inventing the wheel; Promote reuse; Progress together.

An advantage of being a late-starter is that one stand on the shoulders of giants. The 3Ps will help in this regard and prevent re-inventing the wheel. The programme should look at what is already available and collaborate. There is no way that data from Antarctica can stay in one lab, and the value of that data increases when it is used everywhere else – so it cannot be boxed.

Several recommendations and products can be adopted and it is an opportune time to share these with this group, as you begin your work. Examples to illustrate what RDA already offers were shared, but the website needs to be consulted for a complete overview. Some examples: Wheat Data Interoperability: the group came up with a cookbook that describes how they created interoperability between wheat datasets around the world. Using this as the basis, a rice data project has been initiated and a rice dataset interest group is now functioning. In addition, a soil data group has joined. The wheat data interoperability book did not work for them, but they found a group elsewhere in the world. Billions of dollars have been pumped into soil data, but RDA still invited them to participate and then realised how disparate the dataset was: the database could be put to good use nationally, but not internationally. So a soil interoperability data group now exists as well.

An example of how the RDA can complement the work done by AOSP was shared:

Africa Open Science Forum	Research Data Alliance
Open Data and Policy Developments National data for a coordinated policy development	▸ Guidance on Legal Interoperability (how to license research data for widest reuse and principles that underpin this) - joint CODATA-RDA work ▸ RDA/NISO Privacy Implications of Research Data Sets WG

Data and communication of research Incentives and benefit	▸ International Collaborative platform for information dissemination, plenary for interactions and networking ▸ Standards for Data: Agricultural data standards
Data Curriculum materials for foundational data skills (CODATA-RDA training schools)	Curriculum materials for foundational data skills (CODATA-RDA training schools)

Data infrastructure and research strategy priorities
    Infrastructure roadmaps
	Data Interoperability frameworks (Wheat WG)
	    PID Information Types WG
	    Data Fabric IG

Questions: None

African Open Science Platform: Introduction, Large Scale Vision, Mission and Objectives of the DST Project; Draft Framework of Activities and Discussion

View presentation: http://www.slideshare.net/AfricanOpenSciencePlatform/african-open-science-platform-pilot-phase

Dr Simon Hodson (*Director CODATA, France*) thanked the previous speakers for providing an overview of RDA and for indicating how international collaboration was functioning, what the activities of other organisations are, and how these could slot into AOSP. He indicated that work was already being done with RDA and they needed to be involved, but there were other activities and initiatives represented that was important to engage with as well. He reiterated that this was the ethos for the workshop and for the initiative as a whole.

ICSU's mission is to strengthen international science for the benefit of society and CODATA's mission is to support the mission of ICSU and promote data for collaboration. Dr Hodson referred participants to the CODATA information packs that had been made available and which details activities undertaken and outputs delivered. He clarified that certain outputs had been pursued in association with RDA.

The CODATA strategic priority areas are as follow:

1) Underlying principles, policies and practices, e.g. The Value of Open Data Sharing. This was led by the working group on open data in the world. Work was done through working groups, task groups. 2) Frontiers of Data Science: The Data Science Journal which is at the frontier of data science, communicating on making data more useable and available, and how to best extract meaning from data. 3) Capacity Building: a great deal of training and capacity building is done to guide the activities and see them grow. At the Beijing workshop, Prof Madalli had invited CODATA to collaborate in a similar initiative in Bangalore. This had been a rewarding experience, with value being delivered. An attempt had been made in the last two years to make the activities as effective as possible, by partnering with Data Carpentry and Software Carpentry partners and utilising materials online and adding additional materials to develop a curriculum of core foundation skills. There is now a broad curriculum that can be reproduced easily and the idea is to now explore how to benefit and network to build skills networks and materials to benefit the maximum number of people using the investment available.

The AOSP will build on the **principles, responsibilities and enabling practices** laid out in the *Accord*. The platform is an ambitious and far reaching initiative to encourage the development of national science ecosystems that are equipped for open science. It is conceived as a broad African initiative, directed towards engagement with other African countries. Funding for a pilot phase was made possible through the Dept. of Science and Technology South Africa and the NRF. It will be managed by ASSAf with direction provided by CODATA International. A high level Advisory Council and Technical Advisory Board will be established to support the development of the platform.

This project will: 1) encourage open data fora for development of strategies and policies; 2) review incentives and disincentives and make recommendations; 3) roll out research data training activities; 4) undertake a research data infrastructure roadmap exercise.

The immediate task is high level and the activity envisaged for the pilot initiative is to undertake a research data infrastructure roadmap exercise, to determine the key data initiatives - national and regional – on the African continent. The survey can be accessed at http://bit.ly/African_Data_Initiatives_Survey. The survey results will inform the following:

- **Policy Development:** what policy frameworks are required to help take advantage of the data revolution and Open Data?

- **Incentives:** what are the benefits of Open Data for African research institutions? For researchers and research groups? What are the disincentives for data sharing? How can these be mitigated?

- **Data Skills and Training:** what are the foundational data skills we need to develop? How can data training be achieved at scale? Can we use the approaches of Data Carpentry, Software Carpentry and the CODATA-RDA School of Research Data Science? What other initiatives can we learn from?

- **Data Infrastructure Roadmap:** what are the priorities in terms of infrastructure? How can we ensure data infrastructure benefits African researchers and research institutions? How can we build on existing work in South African and SADC?

Dr Hodson informed the meeting that CODATA already has very strong engagement with Botswana. Meetings were arranged by *Joint Minds Consult* with key Botswana research institutes. He referred to workshops and discussions on research priorities, the data revolution and the benefits of open science that were held and also planned. Building expertise in data, and the development of open data platforms can benefit strategically important areas of research. It was felt that a vision of excellence, with a strong focus on open data and open science, will bring funding and prestige to the open data initiatives.

With input from colleagues, *Joint Minds Consult (JMC)* has developed a White Paper to establish a National Data Forum and a process for Data Strategy and Data Policy. A paper has been presented to the Ministry of Tertiary Education, Research, Science and Technology. It is expected that this will lead to a National Data Forum, planned for March or April 2017, to continue the conversation on data strategies. The discussions had been fertile and it was hoped this approach could be readily reproduced across the region, as it is believed that this approach would be of benefit. Further discussions are being held with Kenya (Prof Joseph Muliaro Wafula) and Madagascar, although these are only two examples and further initiatives will be pursued. A CODATA Task Group is planning a workshop in Madagascar in May 2017.

This session will be followed by five presentations from experts involved with important initiatives. Each relates to one of the proposed themes of the pilot project. Thereafter a discussion will take place whereby the following questions need to be answered:

- Understanding the Landscape: Of what other initiatives should we be aware? Tell us about them http://bit.ly/African_Data_Initiatives_Survey

- What are the key lessons from these initiatives?

- How should the platform engage and collaborate with these initiatives?

- Recommendations: for each group and theme (national strategy forums, policies, incentives, training, roadmaps), please agree 5 recommendations for the pilot.

- What are the key issues and challenges?

- What do you want to achieve under these headings?

- What are the most effective and realistic steps for the platform to take?

These answers will assist the team with the framework as it is being developed, but direction and input is required from partners and colleagues around the world. In particular, feedback is required i.t.o. are we addressing the right issues. This is to ensure that proceeding with the initiative is done in the most effective way.

Discussion

Dr Kghotso Mokhele indicated that there is a perverse potential from the point of view of how politicians and those with resources in developed countries may use the open data initiatives from developing countries to worsen problems: if data is open, why would many developed countries see the need to fund activities that are primary generators of data? It is perverse, as when they don't fund, existing infrastructure and skills issues worsen. So as we think about how to progress this, we must be mindful of this unintended consequence, i.e. funding in developed countries can make it harder to obtain limited resources for primary generation of science and data because of logic that says data is now open so there is no need to generate data. He has experiences in this type of conversations – South Africa has built the telescope and will build SKA, which will generate enormous amounts of data in this country. So the issue arises in SA that, while we may want to build another telescope, the response is no, because with data being open and astronomical data being open it is not necessary to fund this in SA any more. This is an unintended consequence that is real; so how do we deal with it?

Prof Geoffrey Boulton replied that there is an interesting parallel elsewhere. One argument posed by many is that if we have enormously varied and rich sources of data in the world it undermines the traditional approach in science, which includes hypothesis and test. But most scientific ideas are wrong. So the idea is that you have a lot of data about phenomena and you don't need to develop Newton's Law or develop theories – you can simply make direct observations and have machines that learn and deduce. But Newton's Law can be undermined by data. The error is to suppose that all the data you want is available. But it is not. For example, if one wants to explore soil formation in Africa, all the data for that is not available. So when you formulate data, you use open data and don't have to replicate some, but you still have to collect some new data. But you need a theory in your head as to what data you need. The implicit assumption, however, is that the data exists and you can relax and have a North American company come in and apply it. This is not a good example of the way data exists. What is proposed for this platform for the three-year period is that in year 3 we would focus on strategic research that is highly relevant and which requires large volumes of data. Dr Hodson has already been promoting the idea that we can have that demonstratively in year 1. So, in a way, we have to work to avoid your argument taking root by demonstrating utility and sharing data. There is always the danger of excuses for not doing something and not spending money and this is crucial. In the UK they have a problem with an appalling national debt to income ratio and our grandchildren will be paying off the current debt incurred. Science funding needs to be uphold by demonstrating the

direct impact on social and government priorities. If you are able to demonstrate value, government will not take the route implied; but this must be guarded against.

Prof Agnes Mwang'ombe indicated that we have to be clear on what has been presented as participants have been asked to carry the message home. We have open data, but what is our contribution to that in terms of local content for that platform. Her specialised field is agriculture and with challenges in Kenya, they have totally different information, as crops grown in other countries don't have the same problems, e.g. maize disease in Kenya is spreading to other African countries and causing a catastrophe here, but not in other areas or regions. So we need to look at local data and we need to contribute to the open data platform. We now have a new challenge with potatoes and are trying to understand the climatic changes underlying these challenges. We have a role to play in explaining the need for data to government and colleagues. Initially there was a lot of resistance from researchers, but they now understand the matter and must carry the message home to support this. In plant pathology, there are many new challenges not seen previously. So the mind set challenge is an important and real one.

Prof Devika Madalli responded that all publicly funded data should find its way to open data but agencies are resisting this. RDA made open data available to show an example of the value.

Mr Harris Andoh commented that throughout the discussions, we have been talking about access to data to improve science and society. But one thing we haven't touched on is the quality of the data generated. Quality of open data is relevant: if we input bad data, science will be affected. If we look at the quality of universities, it relates to the quality of data. Data generated needs parity data to help improve inequality and we need data to improve science and research.

Dr Patrick Molotsi commented that a good point was made by Dr Mokhele and the more we raise the counter-arguments, the better, as this will enrich discussion. He often heard at workshops about the threat to IP and losing IP and the indigenous knowledge used in the developing world. It is good that Prof Boulton raised the mind set and ecosystem challenge, as these matters speak to critical challenges and counter-arguments encountered. In countries in the developing world huge amounts of data are not utilised. A counter-argument we hear in Botswana is the loss of government data, i.e. researchers access data they should not, e.g. private or personal data. So there will be challenges and there is more to be heard on this subject, but the critical issue is whether the benefits outweigh the challenges. And the messengers should be strong enough to show the importance of the dialogue.

Andiswa Mlisa (GEO / AfriGEOSS): The point raised is a critical one. With open access to earth imagery, we need to look at how open access affects business models. But we still need authoritative datasets. Perhaps the way maps are created will change, but with open access to earth imagery taken by other players and maps being developed that were previously done by the national mapping agency. So we need to look at the changing business models at institutes.

Prof Boulton responded that his view is that the principle output of research is people, not theory or gizmos. It is people who do research and go into society in all sorts of positions in government and private institutions. Here we have the technology to understand technology and the dangers; and to help at societal level with sustainability. The whole enterprise is about capacity building nationally. If one stands back from the immediate problem of making budgets work, it is a no-brainer.

The Chairperson thanked all participants for the questions posed and the thought-provoking discussion.

Expert Presentations on issues relating to the themes of the Open Science Platform Initiative

Mrs Ina Smith (*Planning Manager, ASSAf*) introduced the next session and commented that, as scientists, all were attending because of a passion for science and the continent and the effort was intended to better the continent, through finding solutions to opening up data, which in turn can speed up innovations and discoveries. Open access has been seen in SA for over ten years, starting with open access repositories and journals; the *Homo Naledi* project is an excellent example of open science, making everything openly accessible from the start of the research process, i.e. data, journal articles etc.

She also advised that the session would focus on the work done on the continent to date, using just a few examples, with speakers having been brought from Botswana, Ghana, Kenya and South Africa. Finally, she indicated that it was hoped that more workshop sessions would be possible in the future, as part of the grant received from the SA Dept. of Science and Technology.

White paper for an Open Research Data Strategy in Botswana

View presentation: http://www.slideshare.net/AfricanOpenSciencePlatform/white-paper-for-an-open-research-data-strategy-in-botswana

Dr Patrick Molotsi (*Country Managing Director and Principal Researcher, Joint Minds Consult, Botswana*) thanked the forum for the opportunity given him to share Open Science and Open Data issues in Botswana and said that he hoped that the draft White Paper could be taken back to the Botswana forum after receiving input during the workshop. Data is also receiving attention on governmental level, and the role of data was recognised in a speech on national television. He also advised that: while a lot of data purportedly shows progress made in dealing with poverty, progress is not elimination; statements about increasing job opportunities, e.g. a statement made that 18 000 jobs were created, without context and providing a link to the base of how many people are looking for jobs and how many jobs are created, for example, is also questionable.

Dr Molotsi thanked the previous speakers for setting the tone and providing the background to the subject, given that a rich background helps to progress. He indicated his presentation was a case study and that five other representatives from JMC and universities in Botswana were also at the workshop. The organisation had prepared a Version 0 Draft White Paper (position paper) that would ultimately be a joint product of the universities, even though engagement had already commenced with the Ministry of Science and Treasury. The plans would be shared during the presentation.

Joint Minds Consult is an independent research institute and there are not many of these in Botswana. It is a regional institute that interacts with SADEC, SA, Zambia, etc. JMC is a member of RDA and one director is a member of the Steering Committee. Because of the links with RDA, some of the leading members of RDA and CODATA were invited to seminars with key universities and colleges in October 2016 to discuss the White Paper on Open Science. The initiative was well received. JMC focuses on research, knowledge development and management, and student support.

Open data is the reality of the world today, i.e. we are interconnected and the world is producing data at unprecedented rates - therefore the term Big Data. We have to respond to the global movement of reinventing science, and data is the foundation of that process.

For the past forty years, Botswana has been regarded as one of the most successful states in Africa, based on key indicators of development. It has one of the highest GDP rates and a stable micro economic system, good policies and it has done well i.t.o. the human development index. The country

celebrated its 50th year of independence. However, it has reached a cross-road and there will be no more success if it continues to depend on a mineral economy which is stubborn in terms of diversification, which was the basis of its success in the past. To transform Botswana into a knowledge economy and to measure the performance of every policy it is necessary to advance Open Science and Open Data. The broader environment is now conducive to more research and critical data.

Botswana needs to produce data for practical solutions to deal with problems in society, including poverty, shortage of land, inequality, etc. Existing data needs to be recognised, even though there are limited resources and research infrastructure. Data have been accumulated but it is not used optimally, as it is not accessible to the majority of researchers. Therefore, it remains isolated and fragmented and there are no standards, which means that data is different. This could lead to buying data from each other when we should be sharing data and infrastructure. These arguments need to be used to show government that we need more collaboration on this matter to move the country forward.

The first draft of the White Paper has been finalised, and further consultation is necessary with government and institutions in order to help develop institutions and create benefits for Botswana. The vision of the White Paper is to develop a coordinated national strategy for research data which builds on a continuous dialogue between key research and academic institutions and government through MTERST for the purpose of creating consensus on policies, legal frameworks, governance institutions and strategies required to enable Botswana to take full advantage of the global trends in data-driven Open Science.

He emphasized the need for systematic and sustained research - to produce data for analysis to arrive at data driven solutions; performance improvement anchored in national development premised on data informed policies and efficient decision making processes. Currently, fragmentation and lack of intra- and inter-institutional interaction that require data and information sharing practice. Collaboration and partnerships on data and information sharing for developing in Botswana hardly happen.

The rational for Open Science for Botswana is an environment where fully utilized data could make a significant contribution to economic and innovation goals; the value of data is in its use and fragmented data should be readily available and fully utilised by all parties in Botswana. Scientific partnerships would be enhanced with other countries if they sign up in support of Open Data.

Many datasets are available from different sources. Universities (in particular, the University of Botswana) have collected interesting data on geological resources, indigenous knowledge and different aspects of the economy and this data is at the universities. The new University of Technology is also generating data that is inaccessible to other players. Some initiatives are beginning to develop between the universities, but this is driven by individuals and based on their own goodwill – it is not an organised and systematic way of opening up data, as the process ends when they leave. Therefore, we need to establish policies, systems, standards, procedures and institutions to support OSOD, rather than relying on individuals.

CODATA is directing an African Open Science Platform (AOSP) initiative and it will only be sustainable at country level with the following in place:

- Physical infrastructure with guidelines on acquiring, curating and disseminating data and information.
- Protocols, policies and procedures in the science system.
- Training opportunities for skills development on the use of data.

This will provide the structure and support to ensure that science objectives are achieved.

The landscape in Botswana is a challenge. In 2013/4, the Government of Botswana invited researchers from the World Bank to do a readiness assessment for OSOD; the assessment focused on leadership, policy/legal framework, institutional preparedness, data within government, demand for data, open data ecosystem, financing, technology and skills infrastructure and key datasets. The conclusion was not encouraging. The conclusion of the research was that Botswana was not ready for OSOD i.t.o. government and certain other critical areas. Like in other African countries, leading research institutions in Botswana – universities and colleges and even industry and non-governmental sectors - are keeping fragmented, underutilized and inaccessible datasets of varying quality and national and international relevance. Research collaboration and data sharing initiatives are almost alien to institutional operations.

The rating used was green, grey, yellow and red (worst). Botswana was red in critical areas, e.g. policy and legal framework; this means the legal environment for open data is far from ready. There were also concerns raised about private personal data and security, as government was concerned that - by opening data - they would be giving away the security of individuals and of the country. These are some of the challenges that must be acknowledged.

But even within institutions there is a red landscape. The country is rich for such a small country, with only 2.1m people and it has lots of institutions: public, private, research, academic. Therefore, there is a good opportunity for creating an open data institute. However, in the last two decades, R&D was still low, as was GDP (at 0.348%) by international comparison. But the last decade has seen rapid movement, including: the establishment of the University of Science & Technology, which is an example of government fast-tracking science and technology research; a revised policy being released in 2011. The landscape shows positive signals, but there are still challenges with government readiness – hence our advocacy to be strong drivers of this.

Even within institutions, cooperation and collaboration across disciplines is low. The research community is weak and fragmented by disciplines. The silo culture is strong! No clear policies and advocacy strategy to promote open data currently exist in the country. Hence the White Paper by research leaders in Botswana to get the country together to formulate a strategy on Open Science/ Open Data.

Dr Molotsi highlighted some areas of excellence and opportunities in Open Data: water and hydrological studies in the last 15-20 years in the Okavango Swamps and there are mega computers operating at a high level at the University of Botswana; energy; computing and security etc. They are fully aware that in Botswana, there is a great deal of data in private businesses and they need to engage with these organisations, e.g. the diamond mining company, *Debswana*, has huge quantities of data on mining. When talking about OSOD, we need to talk from a multi-disciplinary perspective, not just about technology and the natural sciences. Data from the social sciences, the humanities, etc. must receive equal attention, as there is social illiteracy and many good research outcomes could have been better shared if other professionals had understood the basis of the data. Breaking down the walls of science and our traditional disciplines is important.

A number of advocacy and awareness raising seminars on the 1st draft White Paper were held, and agreement for a coordinated course of action was secured by the following stakeholders: UB, BIUST, MTERST and private universities. The first national research forum is planned to be held between March and May 2017 and a small team will be formed to help drive the paper before the forum. Stakeholders will continued to be invited to be part of the process. Internal and external funding are needed, as the stakeholders don't have the finances to host the national research forum. However, it is important that those they have already talked to are receptive to the idea. Dr Molotsi indicated that

he hopes to pursue this with the Ministry and have started speaking to them. The 1st draft White Paper must become an instrument for dialogue, towards reaching consensus on making research data openly accessible. Taking the dialogue to governmental level will ensure an open research data strategy for Botswana.

Benefits of Open Data and Policy Developments: perspectives from research institutions and universities

View presentation: http://www.slideshare.net/AfricanOpenSciencePlatform/benefits-of-open-data-and-policy-developments-perspectives-from-research-institutions-and-universities

Prof Joseph Wafula (*Jomo Kenyatta University of Agriculture and Technology, Kenya*) presented on the benefits of OSOD, an OSOD policy, and shared some thoughts on the journey that attendees have been invited to participate in.

According to Prof Wafula, the digital preservation of data is key to an OSOD policy. Publishing the data is not the only important thing – all information describing the data, as well as the data itself, must be preserved. Researchers collect data to find solutions to problems, resulting in new innovations and discoveries. Intellectual property (IP) issues need to be addressed through a policy, and IP needs to be clarified at the time of data creation. Often IP stands in the way of data being shared, resulting in datasets lying on shelves for 15-20 years after which it is retired, while it could have been used for the benefit of all, if it had been made available.

Technology changes, and for data to remain accessible, data needs to be curated and migrated to accessible formats, to be shared and re-used in the unforeseeable future. Researchers often make the mistake to think that they have access to all data available, to respond to research problems. Only making use of available data often limits the researchers' understanding of research problems. Having access and exposure to the wealth of all datasets, can help the researcher to better understand the problem, providing additional insight.

We live in an inter-connected world and so, for example, creating something in the manufacturing field impacts on climate or agriculture. If research is collected in the same timeframe, the dots can be connected and problems solved, with synergy adding value to the processes.

Cultural challenges need to be dealt with in order to make data public by default, in contrast to the current approach of it being private unless the researcher can be convinced otherwise.

The key objectives of an Open Data policy can be captured as follows:

- Promote data publication, preservation and reuse.
- Promote multi-disciplined research capabilities and activities that are ICT enabled.
- Accelerate ICT innovation through equipping innovators with requisite skills and readible and quality data.
- Change culture to keep data private.

Often we think of Big Data in terms of volume (size) as the preferred data to be preserved, while smaller data sets (created on desktop computers/laptops) may contain valuable information needed to tweak something and achieve a breakthrough in innovation. Therefore – the size of the data should not matter. All data supporting the work of postgraduate students should be preserved, so that others can build on their work. An Open Data policy should address this differentiation.

Clear Open Data policies and strategies are key pillars to successful data governance. There has been great progress in terms of policies – especially from European countries – but many countries (including in Africa) are struggling. The AOSP initiative will provide a framework for all African countries to move forward and learn from each other. If the continent does not collaborate, there is a risk of initiatives collapsing. We therefore need to work together, and ensure that there are clear policy and legal guidelines.

A summary of best practices – of which some borrowed from Botswana – were shared. Please refer to the slides.

Technical documentation and metadata describing the data are most important; however, some scientists provide data, but fail to understand and explain some components of their own data (context), while some data notes are lost completely, as information is not kept after a paper is published. Therefore, data about data must be made available and reasonable descriptions must be provided so that it is easy to understand/interpret; and this must be provided in useable formats, without too much delay. The costs associated with data collection needs to be recovered, since someone somewhere had to pay for it – often the tax payer. Business models and sustainability therefore also important. We should remain focussed, and AOSP should not only be about collecting all available data, but also about addressing research needs.

Open Data policy development needs to be based on the following three pillars: context, content and impact.

With reference to the **context pillar**, key factors include:

- Level of government organization: is there support from government, which institutions need to be involved, who will champion Open Data.
- Key motivations, policy objectives: should be clear, and all stakeholders need to buy in into the process.
- Open data platform launch: must communicate a clear message and provide direction.
- Resource allocation & economic context: how sustainable are the resources.
- Legislation
- Social, cultural and political context
- Drivers for open data
- Forces against opening data

The private sector is pushing for open data and many businesses are now running with open data.

With reference to the **content pillar**, key factors include:

- Licensing: which license will apply, to determine the conditions for use of the data. Also indicate how the data must be cited.
- Access fee: will a fee be charged? Even if data is open, it doesn't necessarily mean it is completely free.
- Data restriction
- Data presentation
- Contact with user: a feedback mechanism is required, so that we know how the data is used, the impact thereof, and more.
- Amount published
- Processing before publishing
- Cost of opening
- Types of data
- Data formats and standards

- Data quality
- Provision of metadata

Many questions need to be asked, including about data quality, and there may have to be some processing step involved before data is published, e.g. a certification body to provide credibility, to guarantee that the data has been collected and made available according to policy. Trust needs to be established.

With reference to the **impact pillar**, key factors include:

- Re-use of published data
- Possible predicted risks
- Benefits aligned with motivation
- Public value
- Transparency and accountability
- Economic growth
- Entrepreneurial open data use/innovation
- Efficiency
- Environmental sustainability
- Inclusion of marginalised

Some motivators to publish data include that researchers and scientists are doing this to inform economic growth and entrepreneurship.

For open data initiatives to succeed and to be sustainable, key strategic pillars need to be in place:

- A support open data infrastructure based on open data policies, standards and supportive legal and licensing frameworks needs to be available.
- User friendly data publishing tools should be accessible. Remove barriers, and make the process of data sharing easier for scientists.
- Feedback channels need to be created for data users
- High in demand datasets that users want, need to be prioritised
- Quality issues regarding datasets need to be addressed
- Privacy rights need to be protected, which depends on country policies.
- Clear, consistent and useful metadata need to be collected and preserved together with the datasets. Data must be intelligible, so that people don't call to question the researcher about the data.

JKUAT in consultation with CODATA developed and implemented an open research data policy (JORD). Already researchers are benefitting from this policy, with about 2 000 post-graduate students per year producing new datasets.

Continued expected benefits of the policy:

- RDI
- Encouragement of international studies and experience
- Promotion of new areas of work not envisioned by the initial investigators
- Strengthen the credibility of scholarly publications
- Development of new products and services
- Support JKUAT open data platform

Botswana is an example that African governments can drive Open Data policies if they want to. In Kenya, a national Open Data policy still needs to be produced, although the Freedom of Information

Act has been passed. JKUAT will be working with IBM to build a cloud-based open data platform that will be launched in 2017. Proprietary technology is involved, but there is an option to marry it with other extensible technologies for growth and to ensure that all partners on the continent are reached. All stakeholders are most welcome to use this.

Open Data as a resource can increase trust in governments, boost economic growth, create jobs, and improve essential products and services. To harness Open Data, proper policies and strategies need to be in place. Dialogue within the open data ecosystem needs to be promoted so as to create open data programs with a balanced supply and demand. We need to schedule round-table events similar to the workshop, to discuss the fears and how to overcome them.

Prof Wafula also acknowledged barriers towards a national Open Data policy.

Prof Wafula concluded with an appeal that the current process would ensure the development of a template, model or vision that all could provide input to and which can then be used to convince governments and institutions to support the process. He also indicated that he was available to provide technical support from CODATA.

Questions: None

Data and communication of research: incentives and disincentives

View presentation: http://www.slideshare.net/AfricanOpenSciencePlatform/data-and-communication-of-research-incentives-and-disincentives

Dr Louise Bezuidenhout (*Researcher, University of Exeter, UK*) - a social scientist that has been collecting data in SA and Kenya over the last two years - indicated that she would focus on what drives scientists to share or not share data and try to flesh out the issues.

Best data practices are not always embedded in communities or acted upon on the ground by individual researchers. This indicates a need to do more than pay lip-service and to get buy-in through: incentives that will stimulate responsible practices; disincentives that will prevent the unwanted practices.

Research has been done on incentives and disincentives for sharing data, and why researchers share or not share data, and a wide range of reasons have been uncovered. In all cases, the ways in which the scientific community views sharing are fluid and changing. The research demonstrated that reasons for data sharing reflected social, cultural and the physical context of research conducted. Also important to note is that the percentage of scientists that share varies considerably between countries and that the motivations differ significantly.

Many of the reasons to share are motivated by community norms and are aspirational and community focused (on the end product), i.e. how the community can benefit. This is seen as a positive aspect. Sharing is further often influenced by requirements from funders and institutional codes of conduct.

The reasons for not sharing are pragmatic, e.g. lack of money or other resources (time, expertise), because of issues relating to individual credit (scooping, misuse, misapplication), or because of confusion regarding requirements that might be conflicting. In most of these cases the researchers are individually focussed and not motivated by the end product. Also, policy may be absent or confusing.

An examination of incentives and disincentives in an African context was conducted, as the situation is different to that in Western institutions. The questions asked:

1. Are the incentives of sharing adequately communicated in a manner that resonates with African scientists?

2. Are the personal incentives of sharing also effectively communicated and realizable?

3. Are the challenges of producing data in low-resourced research environments adequately considered as disincentives to sharing?

4. Do policies offer any protection against these disincentives?

Fieldwork done across 4 laboratories in sub-Saharan Africa included 56 semi-structured interviews with postgraduate and research scientists, laboratory observations, policy reviews and governance interviews. The objective was to understand the daily pressures to share data or not share data and questions were asked that fell into four categories: data generation; data processing; data sharing; data re-use.

In SA, the general thinking reflected by 1/3 of the researchers regarding data sharing was: it's a nice idea in theory. Further outcomes indicated:

1. Most respondents supported the ideal of Open Data in principle.

2. Many respondents discussed Open Data as something "out there" and not in their research context.

3. Very few respondents had awareness of the personal benefits of sharing or evidence for it in their research facility.

4. Daily challenges of producing data in low-resourced research environments were frequently cited as disincentives for sharing.

5. Few respondents felt policies offered any protection against these disincentives and current research infrastructures compounded these problems.

While most respondents supported open data, the issue was discussed in very abstract terms. Open data was seen as a Western concept, which demonstrated incomplete/limited buy-in and lack of ownership among Africans.

There was no evidence of the personal benefits of sharing data and the absence of mentors made it difficult to comprehend the benefits. No-one had approached them and they were not willing to engage.

There was also confusion around predatory publishing and Open Access, showing a lack of willingness to participate in open access.

The examples provided of issues that detracted from them sharing were many and wide-ranging and these can be classified into: physical (ICT, Internet, power outages, funding, and equipment), personal (data skills, location, and work load), social (policies, IP, high student turnover, procuring resources).

Feedback was collected on how perceptions were shaped regarding data use:

- The perception exists that the data is out there, but the researcher just doesn't get to it.
- Fees shape how and whether data can be accessed.
- The researchers would like to work faster, but promises are made (regarding technology), and not delivered on.

Feedback was collected on how perceptions were shaped regarding data sharing:

- Lack of resources was indicated as a problem, with researchers arguing, for example, that if funds have to be spent on buying chemicals, they cannot spend time on publishing. Also, they may have to pay for publishing out of their own pockets, but it is expensive to publish.
- Some people in Africa feel disadvantaged for not being able to have sufficient resources, compared to their international colleagues, who can do the same research must faster.
- Controlling or managing data was indicated as a challenge. With sufficient resources, it would have been possible to mine the data, also where incomplete or not organised.

Policy was mentioned many times. People felt they did not know what was happening on the ground, policies cannot be followed, there is a lack of protection of data, and people don't trust the way data is protected.

Understanding of data sharing incentives and disincentives is crucial and a complicated aspect that is more complicated in Africa than in other countries i.t.o. protecting data.

- The strong and complex cultural component must be recognised when thinking about the carrot and the stick.
- The full scope of data engagement activities that the programme would like data scientists to contribute to must be addressed.

Incentivization must include:

- adding long-term value and advance individual and community aspirations

- pragmatic short-term rewards

- are in line with accepted community values and have community support

Addressing disincentives must consider:

- daily research challenges of individual scientists
- achievability of data requirements *in situ*
- full scope of data engagement activities

Questions and comments

a)   Comment from **Prof Geoffrey Boulton**:

It's a nice theory, but in one domain it is absolutely required in practice. When evidence of science data is published, you can't discuss it. The progress of science is like a dessert trail strewn with litter … and you can't do that if you don't have the evidence. You need the data when publishing science or you enhance the risk of questionable results. Without data being contested, the risk of using it is high. So, contestation must be high and to do otherwise is malpractice.

A comment was provided by **Dr Bezuidenhout**: More engagement is needed for open science and it should be a fundamental part of being a scientist – people shouldn't be able to choose.

b)   Question from **Dr Magdalena Eriksson**:

If you publish data and someone else uses it, what is the acknowledged practise? It should be as rewarding as quoting.

Response from **Dr Bezuidenhout**: More and more people are starting to recognise the re-use of data as a benefit. But for many of the scientists who publish, there is no benefit in releasing their data until they had exhausted the publishing route.

A comment was provided by **Dr Eriksson**: Some publishing units require that the data is deposited.

Response from **Dr Bezuidenhout**: Many of the publications used by these scientists did not require this to be done.

The Chair concluded by stating that - when there is no data, it is merely an opinion. When the data is however available as proof and evidence, it can be used for predicting and drawing trends.

What are the data skills requirements and how initiatives scale to meet the need?

View presentation: http://www.slideshare.net/AfricanOpenSciencePlatform/data-education-and-skills-initiatives

Dr Magdalena Eriksson (*Director, AIMS-Next Einstein Initiative, Rwanda*) explained the difference between data, information and knowledge. An understanding of the fundamentals is important. *Datum (pl. data)* refers to a record of a raw, untreated fact. *Information* refers to a set of data, which is at least partially related by subject or nature of the data. It also implies some level of real or conceptual homogeneity. *Knowledge* refers to the body of information that has been evaluated with reference to a particular circumstance, e.g. level of education in a country.

Information is primarily used for research and data is collected and structured so that knowledge can be deduced. Knowledge is required for effective decision making (evaluation), training (sharing), learning (acquisition) and a common understanding (culture).

Earlier in the day it was shown that our capacity to generate, store and process data is enormous, but we would like to ensure that there are well-trained people with the necessary scientific and technical skills, who are versatile in using/applying the data. If someone studies one subject, you should ideally be able to transfer knowledge to something else. This is the educational perspective.

Dr Eriksson argued that data training should include level of responsibility. Researchers have the responsibility to use data correctly, use it for valid purposes, and to report conclusions responsibly. If there is no research funding screening body, there is sometimes less control regarding what is asked, and this needs to be useful.

AIMS is a non-profit network in the mathematical sciences space. The driving force is not learning per se, but using and applying knowledge in a productive way for the African continent. Its vision is to lead the transformation of Africa through innovative, scientific training, technical advances and breakthrough discoveries for the benefit of society. AIMS support data education and skills through the training of Master's and PhD's, conducting research, engaging in communities, training teachers, and the Next Einstein Forum (NEF), with local and international partnerships which strengthen the work done by AIMS. An AIMS training institute was started in SA in 2003, after which research and other functions were added. With students from 42 African countries, most of the continent is well represented. On average, 300 students conduct structured research on Master's level annually, with at least one-third representing either gender. Faculty members are from all over the world, to maintain a high quality of lecturers and tutors. Students have access to a computer, there is a library and students live at the facility, which offers an active, buzzing environment. There are three phases to the course, which runs for a year. After the first block (computing & problem solving), which all students do, they then choose the course to be followed. In the third phase, they prepare a written mini thesis and defend this in the expert community.

Students learn best when actively engaged and when doing. The library plays an important role, and is frequently used. Students do not participate in traditional exams, and assessment is based on

performance in terms of the mini thesis by each student. All programmes offered are accredited and recognised certificates are issued.

The emphasis is on computing and there is an open source policy to use open source operating systems to do research and assignments, e.g. Sage, Python. The idea is that students should be able to proceed without making a huge investment and that they should share what they do, learn and produce. A pilot programme on big data was presented, with internships in the programme and research being done where the students work. Big data is highly relevant (more than computer security) and students learn the techniques of working with datasets. Seminars prepare the students for joining industry and all learn as they go, preparing them for lifelong learning.

It is expected that more AIMS centres will be opened during 2017. Each AIMS centre has a research facility with junior and senior chairs that are well-funded and these are growing. Researchers visit these to enrich the environment. More than 100 papers are published per annum, and 25-30 PhD and Master's students are supervised to add to the pool of African scholars. All students are trained in an atmosphere of sharing and doing work that is good for the continent in general. The focus is now on new areas, e.g. quantum science and climate change, and will generate and use large quantities of data with a special track in the Master's training programme.

AIMS provides an example of a training paradigm. It trains students to be versatile and to use open access materials. The research will advance science, for example with how you structure datasets. Graduates are ready to participate in open science at various levels and AIMS can contribute to open science in Africa. AIMS responds to some degree to skills requirements technically and with some versatility. There is responsibility as they try to ensure an atmosphere of discussions and validating what they do, including with the students.

The AOSP workshop is not enough, but much happens quickly, e.g. in 2003, AIMS had 25 students and there are now over 300. AIMS will also benefit from AOSP in order to facilitate the process for interaction.

Questions

Held over for later in the session.

Data infrastructure and research strategy priorities

View presentation: http://www.slideshare.net/AfricanOpenSciencePlatform/data-infrastructure-research-strategy-priorities-in-africa

Ms Nodumo Dhlamini (*Association of African Universities, Ghana*) spoke briefly about the Association, discussed the data infrastructure picture from the point of view of the Association of African Universities (AAU), and provided some suggestions for Africa. Her talk included the following commentary:

- The AAU was established in 1967 and celebrates its 50 years of existence in 2017. All participants are invited to the golden jubilee celebration that will be held in Ghana.
- Membership is growing and includes members from north, east, south and west Africa.
- With data infrastructure building blocks, it is important to understand that this is composed of policy at national, regional and continental level. We include e-infrastructure in this, e.g. software, hardware and protocols. With data infrastructure, it is important to empower researchers with online access to facilities, resources and online tools.

- I.t.o. research quality in Africa, the picture is bleak and not bright, as what we do is not visible. A lot is being done, but we don't communicate well and initiatives are not coordinated well. She agreed with the earlier analogy provided of the iceberg.
- Some players i.t.o. data and statistics in Africa are large, e.g. World Bank, UNESCO, and national statistics offices are failing to match data demands at continental level i.t.o. synthesis.
- There is a record of development partners previously believing that higher education would not provide the return that lower levels would and so investment was not made at the higher levels. This is an example of how wrong decisions are made, as we don't have our own local data to help make decisions.
- AAU faces demands regarding data and is currently trying to build capacity to respond to demands for data. They are often asked about research funding in Africa and there are many queries that they cannot respond to, but they are aware that data is resident at national statistics offices or at universities or elsewhere.
- I.t.o. the data initiatives that the AAU is involved in, they respond to content and policies, e.g. Agenda 2023, the science and technology strategy and policies championed by the AU, and the Strategy for Harmonisation of Statistics in Africa.
- Initiatives the AAU has been involved in include improving teaching and the development of the Ubuntu Alliance. They collaborate with Arab states through research and education networks. This is important, as big data and data infrastructure initiatives will succeed once the internet infrastructure is reliable and scientists can collaborate and share. This is not for the sake of IT, but in support of teaching, researching and collaborating.
- The AAU is also working with or have worked with: the Africa Connect project which supports connections to pan-European networks to connect African communities through infrastructure; the Academy of Science of SA on a database of theses; and the Association of Universities, collecting data about universities in Africa. They have also worked on IPR issues, with data being collected, but the data is owned by IAU and the publisher, so there are many issues with ownership of data.
- They have learnt lessons about universities completing data forms and systems that are flexible enough to allow for the unique demands in Africa.
- Going forward, systems to respond to different needs will be required.
- AAU is involved with: the World Bank, which supported Africa centres of excellence, one of which is a centre of excellence for data in Rwanda; the pan African university and partnership for skills in applied sciences and engineering technology. African governments have put money in a basket and this is to be used to train 10000 PhDs. There are specific targets for Africa that look at how we teach, mentor, etc. The AAU also does commissioned research that is relevant to Africa's problems. One priority is to align to the integration Agenda 2063, which speaks to industrialising Africa, improving quality of life in Africa, lighting up Africa, etc.
- The AAU is also keen on monitoring results, specifically i.t.o. public accountability, good governance and helping to measure the impact of higher education on research.
- The AAU feels there is a need to build a management system for data analysis sharing and visualisations, and norms and standards should be developed. Universities should set up systems to speak to MIS to ensure that data is shared and filtered. AIMS building data skills is an important role player.
- Data policy formulation and the implementation of protocols should be addressed.
- National research and education networks are fundamental to the data initiative in Africa.
- AAU is keen to continue collaborating, so as to ensure that we don't duplicate work, but learn from what has already been done as we progress.

In closing, the Chairperson commented that all of this was possible because of the internet being started in the early 1990s. She said that we should promote the web to remain open: in the words of Tim Berners-Lee, create a web we want for the world we want. Through this vision, we can accelerate research and find solutions faster.

Questions

Dr Simon Hodson indicated that the next step in the process would be the identification of existing activities and opportunities for collaboration on existing initiatives (North-South and South-South cooperation). This would be done after an opportunity was provided for the question time that had been held over following the presentations in the previous session.

Questions and comments were then taken on all six presentations and responses provided, as follows:

Question from Prof Julius R Atlhopheng (*University of Botswana*): We have seen a demonstration of the impact of open data; how do we convince people to buy it? Joseph has dealt with this and perhaps he can respond and perhaps also Geoffrey.

Response from Prof Joseph Wafula (*JKUAT*): Impact is the ultimate goal. Last month he won a hackathon, which is a positive thing, i.e. bringing together the people who are creating the data, those with the technical skills to manipulate the data and entrepreneurs who can discern a business model in the data under discussion. So, when data is converted and packaged into a format, we can demonstrate usefulness. For example, they had a researcher working on different vegetables, but he didn't know the chemical and nutritional compositions and the data was manipulated to show iron content, etc. Others have built a model to support lifestyle and another has to do with lifestyle change, e.g. where oxygen is not sufficient, something can be done about it. These are some of the quick wins and some benefits can be built around such things. There is also collaboration with Sweden, which has a project in Tanzania, but which cannot advance, as the scientists cannot bring things together, e.g. one is working on soil, another on nutrition, etc.; so, we bring them all together to collaborate. The other aspect is transport, e.g. we have Uber and instead of sitting in a traffic jam for two hours, we can use Uber to get anywhere. The more the people open data up, the more solutions will be found. These are just some examples and people need to show these going forward. The speaker also indicated that he was available to assist with this via the forums and that arguments can be built using real case examples.

Response from Prof Geoffrey Boulton: Some arguments in the African context were provided earlier and these will be circulated. There are four elements: two in the research community; two in the economy. There are benefits from data informatics from openly sharing data: they benefit individually, applications benefit, then society benefits. Difficulty arises when asking about promoting science into policy and science into practical applications. It is difficult to push science into practical value, as many of the conclusions of science are highly questionable and many are producer-driven, not consumer-driven. Results on research in rice cultivation, led to extraordinary progress in management and fertility and the benefits were enormous. In the commercial domain, the US weather service opening its data led to commercial exploitation. Another is that we give away our personal data every day to Amazon, Google, the telephone company, the travel company, etc., and they register it and this is valued in billions. The commercial community is miles ahead of the research community with the value of this. But integrating data from different sources is difficult; although integrating natural science with social science may provide some benefit. The problem is how to get the various perspectives on science pulled together and integrated.

Question from Dr Edson T Selaolo (*Botswana International University of Science and Technology (BIUST)*): Regarding migrated data, data is collected over millennia and it resides elsewhere, e.g. China, UK, but is applicable to all. The initiative we heard about today could be used to politely obtain that data, as some could be extremely useful. For example, when he was doing his PhD, he got to know about the first geologist in Botswana in 1904, who was German; he wrote a book and Botswana had to ask for a copy from Munich. Such a reference should not be lost. Could this initiative be used to obtain such data; but without opening up old wounds or tramping on confidential information?

Response from Prof Geoffrey Boulton: It's about clawing back some information from the past, i.e. closing time back. It's also about clawing it back to the place where it was created, but which it has left. The expectation is that the extension of open practice will enable more data to find its way back home. There is some interest in UN circles to create a decade of data and finding ways of integrating data better. Another way of approaching this is to say that global patrimony of data is important and that important elements will depend on the use of global data. One useful thing is for principles to be developed by a task team that considers the possibility to do this – unless someone has already done it.

Question from Ms Jacqueline Nnam (*Regional Universities Forum for Capacity Building in Agriculture (RUFORUM)*): When talking about data, what is the scope? The Bill and Melinda Gates Foundation has an open data policy, but it is concerned with peer reviewed data. So, in the African context, is it academic publications, or where do we draw the line? Are we talking about certifying data, or what?

Response from Dr Simon Hudson: We are talking about the data that underpins those publications – open data that underpins the articles, but also the open data for access to evidence to verify claims made in the articles. But that data resides in larger datasets, which may be part of an ongoing project at a university. What is the potential use of that data and how can it be re-used? Other types of data are created for research and other uses, e.g. monitoring (tidal, meteorological) and often for application by government services, but they have a profound research utilisation and national statistics agencies gather these. We would include these. Regarding the limits of openness: the European Commission says it is open by default and as closed as is necessary and earlier we heard about dual use and privacy issues.

Response from Prof Geoffrey Boulton: Don't fall into the trap of thinking that peer reviewed publications means that data is peer reviewed: very few publications can verify data and there are more instances of data being created to serve outcomes. But we are now increasingly able to establish if there are artefacts in the data. So, peer reviewed does not mean peer reviewed data.

Question from Dr Joseph Mwelwa (*Joint Minds Consult*): He indicated that his PhD was in bilingual education. As he listen to issues on this subject, his concern is for all speakers. Are we thinking about providing access in bilingual formats, as Africa is a diverse linguistic society and we seem to be talking from a mono-lingual perspective directed at a multi-lingual society? At the Centre for African Studies they have started to map languages for use in SA and up to the Congo. He is disturbed that providing for a multi-lingual mix in Africa is not on the table. This is a critical issue bedevilling education in Africa, as children cannot access information coded in English.

Response provided: In order to re-use data, it has to be intelligible and understandable. If that captures your concern, so be it. That is the essence of re-use – until you understand, you cannot re-use.

Question from CODATA representative: If one has carried out a study in an African context African scientists have a serious problem of policing data. Not that they are not using it. There is some deep understanding why this is available. Do they have a theory in mind why they keep this in mind? In my country, farmers don't use it. But when you try to load it into your car they will stop you; but they don't care when the waves take it or the water takes it. So, what is the conception in their mind? Can we work on this problem? Implementation will be done at country level and that is where the data lies. We look at macro elements; let's also look at micro elements.

Response from Dr Louise Bezuidenhout (*University of Exeter*): We are drilling down into why they do or don't share data and why it is important for scientists. We are not aware of many studies done in African countries. Much has to do with a lack of awareness and problems faced by people in their daily

environment as to when and where they can release data. There is also a question: with people who own data, the following view was common: 'I look for money to do the research, so it belongs to me'. Even with a PhD, no soft copy is provided by the author and it is said to be self-funded. So, who owns research and what constitutes university research? Therefore, ownership is one area: I am doing the research and it's mine.

Response from Makerere University representative: The issue about ownership of data is about IP and laws governing IP. At Makerere we have a number of policies, including IP, which states that if something is done with university funding for research or using university workers, then the work belongs to the university. If it is commercialised, then there is a formula for sharing the benefits. Many countries should be encouraged to formulate IP policies. Uganda doesn't have one. It is said to start at university level and with the research institutes and to feed this into national policy. Most IPs say that if research is commissioned by an employer it belongs to that employer. The issue is a delicate one, as students say they have funded it, so a policy is needed, as the student was supervised by university staff.

Dr Hodson concluded the session by thanking all participants for the questions and responses, especially on the subject of IP and monetarisation of data afterwards.

Unpacking of themes, short framing presentations and group work

Dr Simon Hodson explained the procedure for the last session, with five groups being formed, each with a Chairperson and a Rapporteur. After the working session, the groups would provide feedback on their discussions and recommendations to the plenary session.

The themes to be discussed were:

- how to build national forums and how to determine a national data strategy;
- training and capacity building;
- incentives and disincentives;
- policy development and issues to be tackled;
- infrastructure roadmaps and the role of institutions.

Each group was required to deal with the following questions i.e.:

- Understand the landscape: What are key data initiatives that we need to be aware of - actual or incipient? What are the key lessons from these initiatives and the key institutions that we need to be involved with? How should the platform engage and collaborate with these initiatives?
- What are the recommendations?
- What are key issues and challenges?
- What should we achieve under these headings (in the pilot programme) and how?
- What are the most effective and realistic steps for the platform to take? Please provide at least five concrete recommendations for the pilot programme.

National Data Forums

Rapporteur: Dr Joseph Mwelwa (*Joint Minds Consult*)
Chair: Dr Patrick Molutsi *(Joint Minds Consult)*

Members: Dr Julius Chisenga (FAO); Mrs Ina Smith (ASSAf); Prof Jianhui Li (CODATA); Mr Zipho Tyoda (DST)

The group looked at existing and ongoing initiatives and four countries were identified. Forums have been conducted successfully in three countries. GEO is also working in the field of OSOD.

Existing activities:

- Selected African governments have started Open Data initiatives (Kenya, Ghana, Namibia, Rwanda)
- RECs such as SADC are also talking about Open Data policies.
- FAO is working on OSOD in agriculture (awareness raising; forums in Kenya, Namibia, Ghana)
- AfDB is working with national statistics offices (Open Data for Africa) (e.g. Namibia)
- Also the GEO lead initiative focusing on making Earth observation data available. Member states also include Zimbabwe, SA, Egypt.
- Awareness of Open Data and Open Science has been created in some countries.
- SCI-Gaia promotes OSOD infrastructure: frameworks & training in sub-Saharan Africa.
- DataFirst creating and curating data.

Recommendations on how AOSP can be approached:

- Inform itself about existing initiatives.
- Market itself and seek partnerships where appropriate.
- Conduct a baseline study of the African continent to avoid being blind-sided. AOSP is expected to develop robust marketing initiatives, so OSOD initiatives can become visible.

A few things need to happen on **national level**, but we first looked at what a forum should comprise of. At national level, there are different stakeholders with different mandates and funders. In many African countries, there are lots of entities with different, and sometimes competing, mandates. The problem is to identify who should lead the drive at national level. There is then also a challenge with acceptance of the driver, who must be neutral. Delegates can identify with this. It must also be sustainable beyond the first forum. But most important is the mind set challenge and perceptions were discussed earlier today regarding OSOD.

At **international level**, the forum must be inclusive and participatory, e.g. there are language issues. We also need an OSOD framework to be developed at national level to inform the next steps at international level. Case studies should be provided as examples and for best practice from the African continent, India and from other continents. Steps to be taken:

- Bilateral and multilateral meetings
- Stakeholder mapping
- Agree on coordinating/lead agency
- Case studies
- Sharing lessons

The group made the following **recommendations:**

- Each country should have an AOSP representative, who can advise and connect. Avoid initiatives being viewed as top down. Initiatives often fail in Africa, as they are seen as being imposed from outside. Nationally agreed platforms, agencies and people should be used.
- AOSP Advisory Board should organise government level consultation regarding AOSP. Use influence.

- AOSP should develop a high level implementation log frame. The project needs a project plan with a start date and end date. Impact needs to be monitored.
- AOSP should have a mechanism for multilingual access. An African platform covers other linguistic blocks that are not here today, e.g. French, Portuguese. So, we need awareness of linguistics in the platform at formation.
- AOSP should present impact case studies to convince various stakeholders in African governments.

Coordinated Policy Development

Prof Joseph Wafula (*JKUAT*) chaired the group, which made the following recommendations:

- We need to do a baseline study to navigate the platform going forward. This will be a useful and worthwhile exercise to see what will come out of it. It will help us to see things at institutional level and to determine the understanding and readiness status of some countries, etc. This can come as an abrupt shock to some institutions and the level of readiness will inform the process going forward with the 54 African states.
- Sensitisation is necessary to ensure awareness of what this is, to ensure people are not confused about access and open data. We need to understand what new details need attention, as there will be some disruption in some areas, as some companies (e.g. mining) spend a lot of money mining data and if you say it will be open, they will not entertain you, as you are touching on something that will dispute their source of income. So sensitisation is important.
- Mainstreaming of open data is required to achieve inclusivity. This must be done so that even if it is in the education sector, all actors and stakeholders have mainstreamed open data policies.
- Harmonisation of policies is required. The issue of regulatory frameworks needs attention, as this could become a barrier to the process; this will ensure the necessary engagement and understanding regarding who is driving it and how we will navigate around it.
- There must be alignment and coordination at national and regional level.
- A participatory process with key stakeholders is required and key stakeholders need to be seen to be participating after the survey has been done.
- Capacity building is important and training specialists are needed.
- A champion is needed and must be found at various levels, e.g. national, institutional, etc. There is benefit to be gained from the process and the champion must be well-informed and equipped i.t.o. knowledge and the political will to oil the process and ensure fast achievement.
- This must be a sustainable model. With the right political will and champion, it will be easier to obtain the required funding and people will more easily listen to options and the models offered.

Incentives and Disincentives

Feedback was provided by **Ms Sarah Callaghan** (*CODATA*):

- This group started with a discussion on initiatives outside of Africa that could be of use.
- One thing that came up when talking about incentives and disincentives is that we need to find a balance so that researchers feel supported and not bullied, and so that they are able to comply with policies without unnecessary compliance pressure. Certain themes kept coming up, e.g. it is not enough to mandate a policy and beat people with it; rather provide support to enable them to comply with the policy and make it easy to make data open.

- Motivation kept coming up: in order for researchers to make data open, we need to show the benefits at personal level, for institutions and for wider groups and projects. In order to demonstrate the benefits, we need stories and case studies where it has worked elsewhere, as humans are not impressed by statistics.
- We talked about the potential role of research funders: we can create policies and can require data management plans, and use a stick to force compliance with policy, e.g. no payment without having made data available. But carrots are also required and although the group struggled with this, they thought that centralised structures and funding elements would be good, but the available pool of money is constant and if you take money out of it for training, then there is less money to fund science. The group also realised that once there is a culture change and doing it is the norm, people will just do it. An analogy is using seatbelts: years ago, no-one used them and the stick that was used was point fines on the driver's licence. Now everyone uses a seatbelt. Mind sets need to change in a similar way.

Ms Callaghan advised that the key points that came up were as follows:

- Evidence of the value of sharing, both within projects and outside. Real evidence of impact is required, as people need convincing. Stories from outside of scientific research can be used.
- Policies should be created; but the tools and support to enable those affected by policy to engage with it easily and well are also needed. If the policy says, xyz and no-one supports this, then no-one will comply and that is not what is needed.
- With IP licencing and the fear of getting scooped, education is needed about the issues. For example, I was scooped, but it didn't affect my career – it was just annoying. But how many people get scooped? So, we need to talk to people and say this is a good thing and it is possible. But we can't talk in the abstract; we must use real examples that are relayed by people who have experienced this. We could focus on early scientists, as they may get the job of dealing with it and may be interested in using it in social media and other ways that older scientists might not.
- Stakeholder engagement and policy is important, as enough people need to buy into it, otherwise it won't work.
- Technology: there are plenty of systems out there and publications, etc.; but there is no one-size-fits-all and customisation will be required. This is fine, as off-the-shelf will often capture 80% of what needs to be captured and this is good, as it is 80% more than in the past.

The recommendations of the group were as follows:

- Promote open science to users and institutions by using real examples.
- Develop policies in conjunction with researchers who will engage with them. Buy-in is needed, as the more buy-in there is, the easier it is to change culture.
- Develop tools and services to make it easier to follow policies.
- In forums such as this workshop, share experiences of developing and using policies. Ensure an open and safe space for people to discuss experiences. All need to share lessons learned and their experiences so that we can all learn from each other, as no-one has the complete solution. Don't just do this i.t.o. Africa – go international, as there is lots to learn from each other and there are lots of problems elsewhere and people could learn from you.
- Clarify the distinctions between IP ownership and use and re-use of data. These are things people get worried about. Issue guidelines and guidance to remove one barrier to OSOD.

Education, Capacity Building and Training

The group reported that they had engaged in a lengthy discussion on the environment and that their recommendations were:

- Establish a framework to guide implementation of open data. There are challenges with the University of Nairobi, as government realised there is no framework guiding what is being done and who owns it.
- Develop guidelines for implementing open data, e.g. guidelines for developing countries for implementing open data in a big data world. Guidelines for developing countries are needed, as other guidelines are being developed elsewhere and we need what works for us.
- Establish benchmarks for good practice in open data implementation, as you have to have standards that guide activities.
- Secure buy-in from institutional top management, otherwise all other efforts will be in vain. This requires awareness of activities, e.g. meetings.
- Organise training sessions to build capacity / re-skill staff at institutions. The University of Nairobi experience was discussed. They had champions at six colleges and there was a champion for open data or someone from the IT Department was assigned this role. But a champion is required to drive the process and these people provided vibrancy and followed up on submitting theses in time and ensuring upload and online access, for example.
- Support knowledge transfer programmes through learning visits, mentorship, documenting lessons learned, community of practice, etc. Not everyone is here to listen to the sharing of experiences where things have worked. So, we need institutions to document their experiences in order to help others, as we can't all go and visit universities to learn from them when they implement open data.
- Develop curricula for courses in data science based on a needs assessment. Some universities have courses in data sciences, but there is a need to re-visit the courses to see if they address our needs, otherwise people will be trained in vain.
- Set up an information hub for sharing information, as people need to visit a one-stop centre to gain access to resources.
- If we don't assess impact, we won't know if we go off course, so we need assessment results to inform capacity-building activities.

Infrastructure Roadmap

Rapporteur: Andiswa Mlisa (*GEO*)
Chair: Ms Nodumo Dhlamini (*AAU*)
Members: Boniface Akuku, Audrey Masizana, Armeline Dembo Mafuta, Paul Loughton, Mostafa Shoeib, Felicitas Nosiku Moyo, Paul Samwel Muneja, Juliana Adisa

The group provided a definition of data infrastructure: Digital structures (systems, tools and policies) that support data collection, access, share, use and visualisation.

A number of existing infrastructure initiatives were identified:

- National Research and Education Networks (NRENs), including regional networks and AfricaConnect2
- Kenyan Agricultural Information Network
- DIRISA
- SAEON

- Africa Data Initiative Research Council
- Egypt National Research Centre
- Centres for High Performance Computing
- Tanzania Open Data Initiative
- Kenya Open Data Initiative
- Ifakara Health Institute Data Portal
- Global Earth Observation System of Systems
- Center for Radio Access and Rural Technology
- Ghana Open Data Initiative
- African Development Bank Open Data Initiative
- ICT Center of Excellence and Open Data (ICEOD)
- SADC Cyber Infrastructure
- Private Connectivity Initiatives e.g. Egypt Russian University

Challenges identified in terms of instrastructure:

- Connectivity (high speed networks within countries and regionally), which has two sub-items: policy on connecting and sharing; physical connection which require resources
- You need computing and archiving capabilities if you want to use and share the data.
- You need capacity to maintain infrastructure and use it, so you need to build capacity i.t.o. infrastructure. Training of network engineers, librarians, data scientists.

Recommendations in terms of infrastructure:

- Collaborate with relevant data infrastructure institutions and initiatives identified.
- Advocate for the development, maintenance and appropriate use of data infrastructure. Often institutions in one country are not aware of one another.
- Provide training – using the train-the-trainer approach, and webinars (online or e-training) which can also be recorded – to build capacity of individuals to maintain and use the data infrastructure.
- Promote inclusion of open source and open standards in the Open Science policies.

Questions

An opportunity was provided for questions and comments to be posed and two comments were made, i.e.:

Comment from Prof Joseph Wafula (*JKUAT*): Perhaps AOSP needs representatives from the Ministry of Transport and Communications as part of the technical advisory committee, as there is a need for technical input to ensure that the notion of OSOD doesn't fall apart.

Comment from Dr Justin Chisenga: How do we encourage the forum to do a quick capacity assessment to avoid defining capacity training needs, when what is needed deep down is different? So, I see a need for a quick assessment of the capacity of individuals and institutional needs.

Identification of Next Steps and Closure

Prof Boulton thanked all participants for the extraordinarily rich discussions that he said were a result of deep deliberations. He offered the following comments as part of the closing session:

- It is clear that the sort of issues focused on go in two directions, i.e. high level issues that are appropriate for the advisory council, and some that are very technical and appropriate for the technical body.
- Most comments affirm issues that had already been identified as being important; but there were also some novel ideas that are worth seizing upon, e.g. the language issue was one that hadn't been identified previously.
- What is important and what is required has been identified and the next step is to decide what is required when a board and council operate.
- We need to focus on the next steps and the information provided is very important. We now need to create a contact network and will use all of you, unless you tell us that you don't want to be pestered. The network could be very active and it ought to be, e.g. a common thread is an initial rigorous landscape study being required. The comments about components of study are important, e.g. the legal framework, who is doing programmes, where is the capacity, etc. All of this must be established.
- Some issues will be forwarded to the advisory council, others need to go to the technical body. They will determine the priorities for the first three years. But the ideas you identified indicate what they should focus on - the question is then what they should focus on first and what are the long-term items. The first thing to do is to set up a site and include information, e.g. papers from this forum, education issues, what is going on, plus the landscape study that will go to all of you for comment.
- The way forward then is to create an advisory board and a technical council, as well as a network and a contact site and then to ensure that you are continued to be informed. We will then be able to draw on your expertise, as this will be crucial.

In closing, Prof Boulton thanked the participants again, adding that it had been a very creative day and that he takes heart from the sense of enthusiasm felt, which was the most important aspect of the initiative. He also stated that he believed the initiative had enormous potential and that it was exciting to see that Africa is able to 'get its act together' in this domain and show the world how to get a community such as this one going on key issues and national aspirations.

He thanked all participants one last time and said that the initiative was looking forward to continued contact with participants.

Dr Hodson thanked Prof Boulton for providing the closing remarks and confirmed that the enthusiasm shown had been noted. He advised that all participants would receive further communication, but were able to sign off of the distribution list, if they wished to. He invited all participants to assist with the landscape study and provided a link to a quick survey on key data initiatives, which is part of the input to the initiative, i.e.: http://bit.ly/African_Data_Initiatives_Survey

Dr Hodson confirmed that:

- The scribed document from the workshop plus the presentations would be uploaded to the website.
- This would lead to a framework and the process would be participatory.
- The workshops and the work of the council would feed into the process.
- The participants will be kept informed of the landscape study as it progresses.

Ms Veldsman thanked Dr Hodson for facilitating the workshop and all participants for their participation. She expressed the hope that the last half-hour of participation by the new group of people who had joined the workshop had added another dimension to the discussions and stated that new groups joining the workshop from other groups demonstrates that, through academics in Africa, this was the type of input and direction that could be obtained. She then thanked those who had joined the workshop from other groups for taking the time to join in on conclusion of their own groups' workshops. She confirmed that the website mentioned by Prof Boulton was already up and running in the background and that all presentations would be made available in due course.

Web URL for AOSP: http://africanopenscience.org.za/

Acronyms and Abbreviations

AF	Africa Forum
AAU	Association of African Universities
AODP	Africa Open Data Platform
AOSP	Africa Open Science Platform
AU	African Union
CHPC	Centre for High Performance Computing
DST	Department of Science & Technology
HPC	High Performance Computing
IT	Information Technology
IP	Intellectual Property
i.t.o.	in terms of
JMC	Joint Minds Consult
MIS	management information systems
NEF	Next Einstein Foundation
NRF	National Research Foundation
OSOD	Open Science and Open Data
RDA	Research Data Alliance
SA	South Africa
SADC	Southern African Development Community
UK	United Kingdom
UN	United Nations
US	United States