Global Partnership for Sustainable Development Data

Better data. Better decisions. Better lives.

Africa Regional Data Cube:

One-Year Post-Launch

ABOUT THE AFRICA REGIONAL DATA CUBE

The Africa Regional Data Cube (ARDC), based on the Open Data Cube infrastructure, is a technological innovation that layers 17 years of satellite imagery and Earth observation data for five African countries: Ghana, Kenya, Senegal, Sierra Leone, and Tanzania. It stacks imagery across a time series and makes the data – which is compressed, geocoded, and analysis-ready – accessible via an online user interface and Python application notebooks.

The data cube was developed by the Committee on Earth Observation Satellites (CEOS) in partnership with the Group on Earth Observations, Amazon Web Services, Strathmore University in Kenya, Office of the Deputy President - Kenya, and the Global Partnership for Sustainable Development Data.

The formal launch of the ARDC, and an in-depth training workshop for government representatives of participating countries took place in Nairobi, Kenya in May 2018. Since the launch, there have been a number of in-country and remote training sessions with users to continue building capacity.

USE CASES

While each country has specific challenges and priorities, the ARDC is being used to help address issues related to land use, water extent and quality, agricultural productivity, and urbanization. The ARDC can help monitor vegetation cover, land degradation, changes in water extent, the growth and spread of informal settlements, and much more.

The ARDC is enabling stakeholders in all five countries to develop and use new methodologies for collecting data to monitor the following SDG Indicators:

2.4.1: Proportion of agricultural area under productive and sustainable agriculture
6.6.1: Change in the extent of water-related ecosystems over time
11.3.1: Ratio of land consumption rate to population growth rate
15.1.1: Forest area as a percentage of total land area
15.3.1: Proportion of land that is degraded over total land area

LESSONS LEARNED

The ARDC was created in response to data needs and gaps identified by partner countries and based on examples of some countries' successful usage of Open Data Cube technology. It is a solution that can help address countries' respective needs and fill data gaps. However, once adopted as a solution, it takes a significant commitment of time and resources to effectively build capacity and increase use. The following learnings were compiled through inputs and insights provided by the GPSDD secretariat, the NASA team, partner country stakeholders and training participants. The learnings are intended to be a distillation of critical reflections on what has worked, what has not and most importantly, what are the lessons from the various experiences that are important to consider in informing continued improvement of the ARDC and related activities.

What does it take to build capacity?

Political Engagement

- **Time investment in country:** It takes 4-6 months of discussions with stakeholders to establish an institutional framework and buy-in from key stakeholders to invest time and staff resources to adopting the ARDC. The best approach was to conduct one-on-one meetings with each institution to understand needs and concerns prior to convening all stakeholders together to discuss opportunities and next steps. Once buy-in is established, it is important to maintain these relationships to improve outcomes and explore future opportunities.

- **Institutional governance framework:** Identifying an institution within which to anchor the ARDC is critical. This ensures that there is a clear focal point with a mandate, responsibility, and incentive to promote adoption and use of satellite data. This includes both political support and logistical management. However, settling on which institution should be the anchor can be a point of contention. In some contexts, it helps to separate political control and technical control between two different institutions. The statistical office is not always the best fit.

- **Embedding within existing workstreams:** It is important to embed the ARDC governance structure within larger national structures and ensure that it is aligned with national priorities. Streamlining the ARDC with existing national collaboratives and programs has been useful in ensuring that the use cases identified are aligned with national priorities and SDG needs.

Technical Engagement

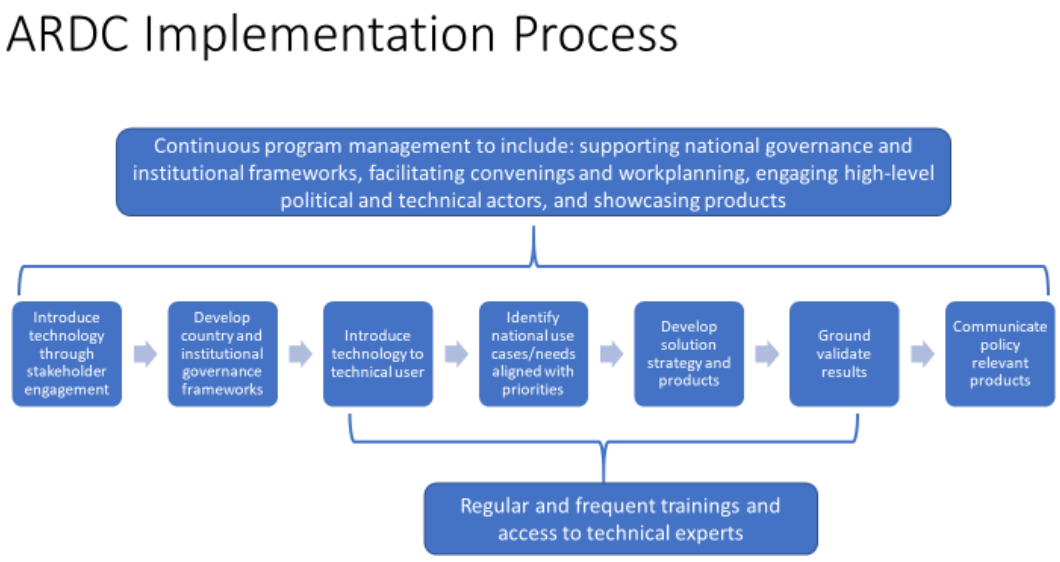
- **Regular and frequent engagement:** Technical engagement and trainings need to be regular and frequent to address the novelty and complexity of the data cube and the diversity of users: it is a new technology, a non-traditional data source, requires specific technical skills and knowledge, and involves a potentially new problem-solving process and approach. In addition, users' baseline GIS/coding understanding and skills fall on a broad spectrum ranging from none to advanced.

- **Time investment in understanding the technology:** The majority of the ARDC's first year post-launch, and much of the initial engagement within each country, was focused on introducing and understanding the technology and its potential benefits. Given that the data cube, and often satellite data, is a new data source, a significant amount of time needs to be invested in getting users comfortable with the technology, so they fully understand what it is, what information it can provide, and how to access and use it. Once users know what the technology can do, they can identify specific relevant applications – problems to tackle using the ARDC – and start answering questions.
- **Engagement between trainings:** There is a moderate learning curve associated with using the ARDC. Because of this, it is unrealistic to expect users to actively use and troubleshoot using the ARDC on their own after the first initial trainings. Users' time investment in understanding the ARDC and improving their skills is often, if not always, additional to their full-time job/duties. Engaging with users on a regular and frequent basis through consistent trainings – both in-person and remote – in addition to follow ups with dedicated experts is critical to effectively improving their skills to a point that ARDC use can be integrated into their day-to-day work.
- **In-person trainings seem most effective:** While it is critical to have multiple modes of knowledge sharing (virtual trainings, training videos, documentation, etc.), users tend to engage by asking more detailed questions, offering new use cases and methods to explore, and voicing concerns and feedback on the user interface or functionality of the ARDC during in-person trainings. Feedback from these trainings has led to many interface and functionality improvements, thus contributing to the improvement of the ARDC. However, multiple trainees noted that the training videos are very helpful and that they plan to share those with colleagues not attending the in-person training.

- **Trainees:**
  - **Relevant:** It is important that the right people are attending the trainings. It was noted that the first several trainings included individuals who did not have anything to do with spatial data/GIS. While there will inevitably be a range of skill levels, it is important to ensure that those attending the trainings will be the technical users. It is also important to consider training logistics and avoid perverse incentives in terms of location and per diem.
  - **Consistent:** Consistently building the capacity of one cohort would be the most effective, as it will enable consistent progress over the engagements and trainings and will develop a cohort that can then train others in a 'training of trainers' model. Every time there is a new trainee, there will need to be some degree of time invested in understanding motivations and the needs he/she wants to address. There have been some missed opportunities in this regard to date. Given limited resources, the approach has been to conduct as many trainings as possible, as often as possible. However, targeting who receives the training will enable more consistent progress for the users.

- **Ground validation:** Once users have developed a solution or product, it is critical that they validate it with ground data. Ground validation is valuable and necessary to ensure that the users' products meet a certain degree of accuracy, particularly prior to being shared as policy-

relevant information. This step is very important and has yet to be addressed with any of the ARDC users.



What does it take to increase use?

- **Effectively communicate value to users:** While timely and potentially new policy relevant insights are the value-add worth communicating to policymakers, it is important to communicate the operational value as well:

    1. The top value add for users is the time saved by using analysis-ready satellite data. Processing data at this scale would be incredibly time-consuming. With the ARDC, users do not need to process the data— it has already been done. In addition, accessing this data through an online user interface or Python notebooks is much easier than traditional methods.

    2. Access to multiple datasets, particularly radar data. Pre-prepared radar data makes it possible to generate new products. ARDC users will be learning how to use radar data along with the rest of the global community, rather than catching up.

- **Identify policy implications:** When developing a product, it is important to identify specific questions it can help answer or policy issues it can help address. For example, the ARDC can help inform a new policy, or it can test the effectiveness of an existing policy. Identifying policy implications in advance of analysis can help users develop more effective ways of communicating results to a policy audience. For example, some ARDC algorithms are directly focused on the SDGs and aligned directly with the indicator methodologies. These SDG algorithms can be an impactful contribution to the 2030 Agenda for Sustainable Development.

- **Managing cloud computing capacity:** Due to cloud credit constraints and costs, simultaneous access to the data and analysis is currently limited. Users have flagged this as a concern and

potential hinderance to continued momentum in using and increasing use of the ARDC. As the ARDC is scaled, it will be important to consider how access to and use of cloud computing will be handled given the number of users. It is possible to optimize this process by knowing user needs and appropriately allocating access per type of task, etc., but this will require a cloud computing expert and consistent user engagement.

- **Develop and manage a knowledge bank:** Users have expressed interest in having an ARDC-specific knowledge repository that houses algorithms, training videos, presentations, literature, and more that can be a resource for existing and new users. This could help users share knowledge on developing solutions per given topic or SDG indicator and consider strategies on developing new or adapted algorithms per different contexts. Short, targeted training videos have already proven useful. One published in January 2019 has already had 200 views.

- **Incentivize through showcasing:** In order to continue building high-level buy-in and support for users to invest time and resources into using the ARDC, it is important to showcase the benefits of accessing and using the data cube. This can be done by documenting and showcasing results at national and international venues and convenings and through technical publications as well as through more general, mainstream platforms and outlets. It is important to increase visibility of utility to encourage increased use.

- **New products**: The NASA team has developed a mobile app for the ARDC based on the fact that mobile phones are prevalent and can potentially offer on-the-spot, quick spatial insights to users. The next step in this process is to apply a use case to test the utility of the app.

This document was published June 2019. For more information visit:
http://www.data4sdgs.org/initiatives/africa-regional-data-cube