



FAIR Pipeline in AI Research

Professor Francisca O. Oladipo

10 November 2021



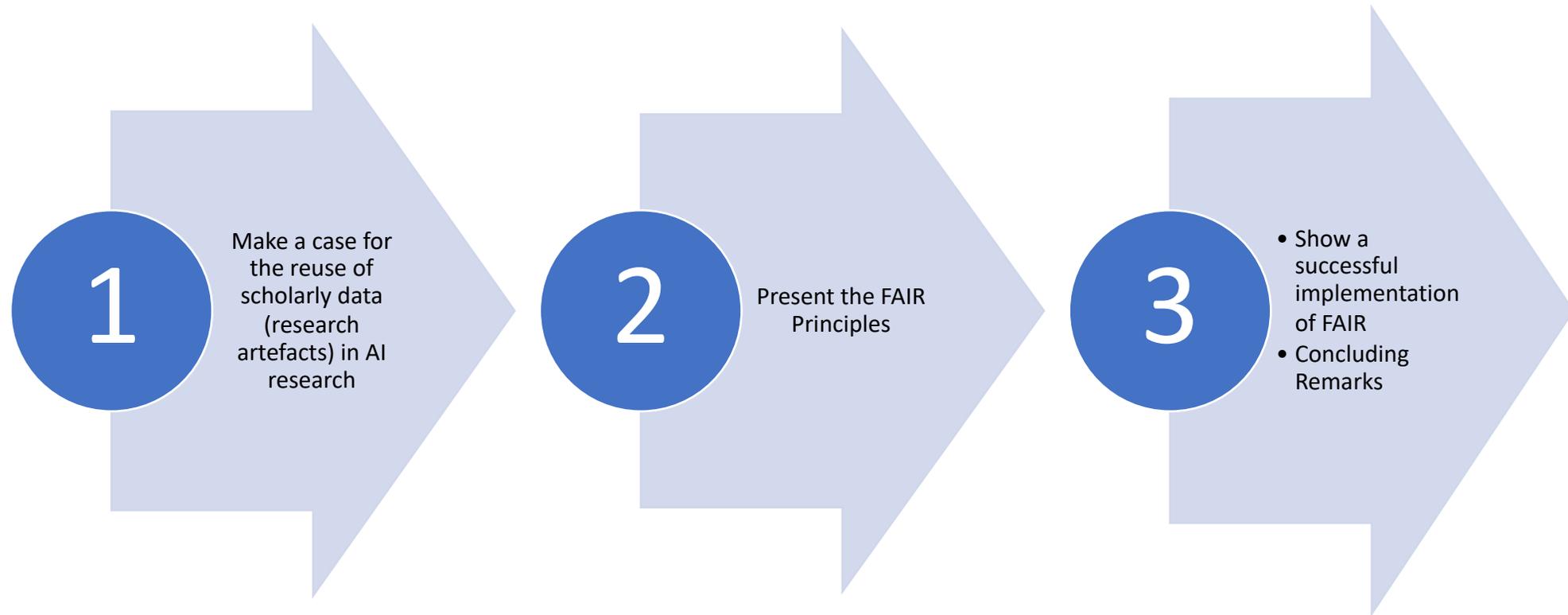
Breaking barriers in science communication and collaboration in African countries using Artificial Intelligence.

#SR2SC2021 will bring together academics, researchers, and programme managers across Africa and diverse disciplines, including social sciences, to analyse the current research landscape and generate a holistic framework bridging north-south science visibility using artificial intelligence.

Supported by: 

 sr2sc@southscieX.io  Nov 10 - 11, 2021

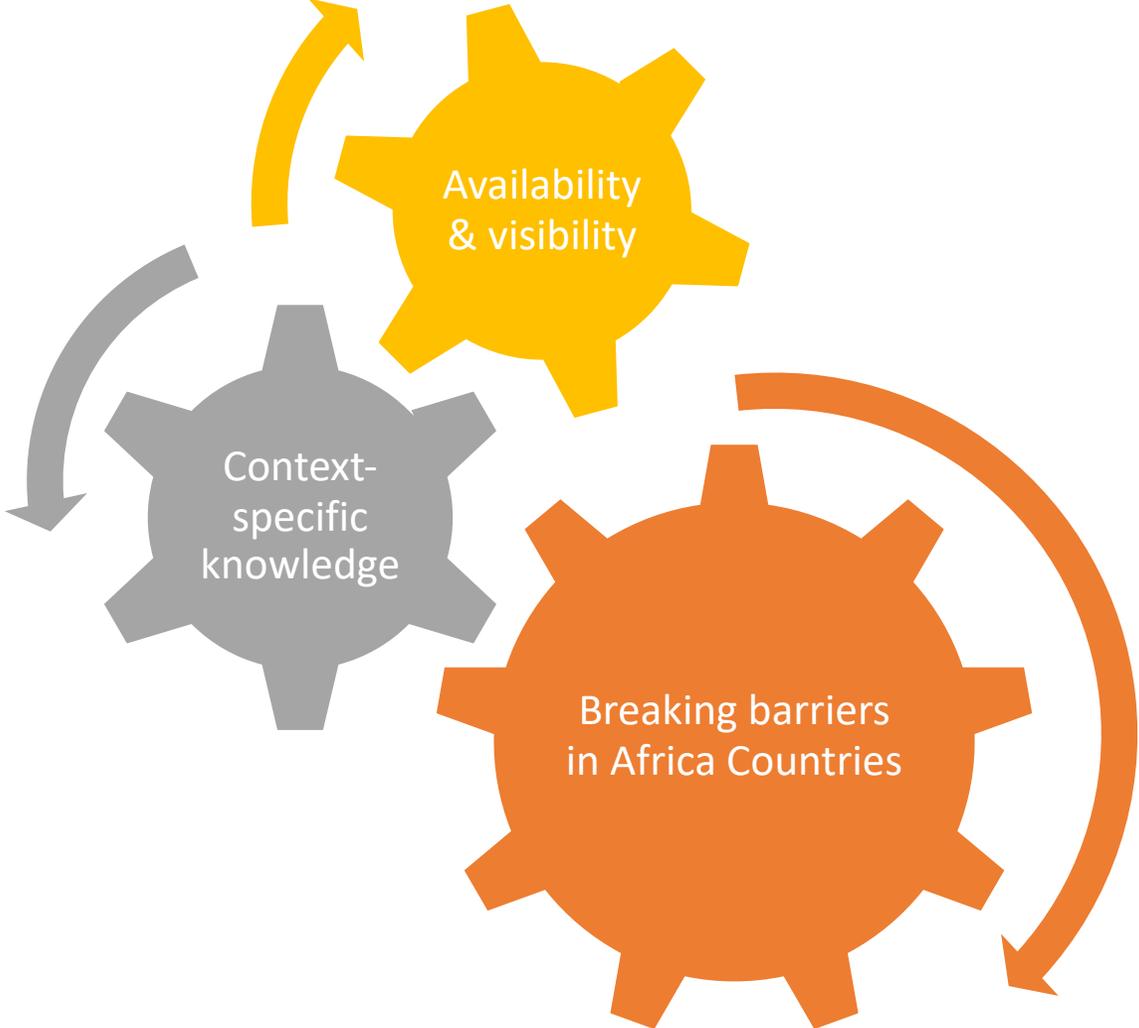
Outline



Housekeeping

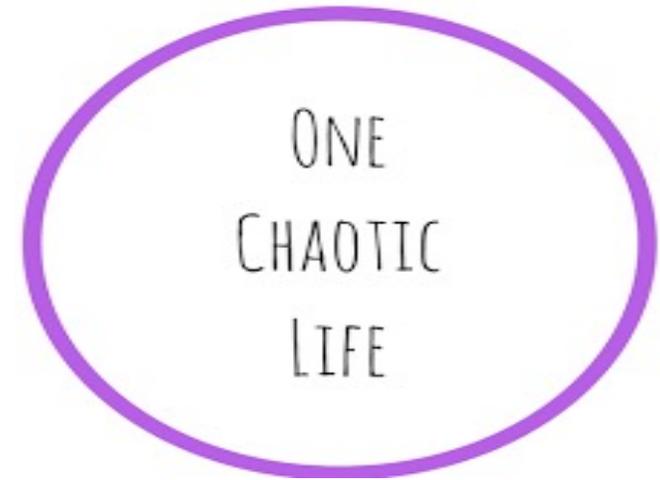


The Context of social change



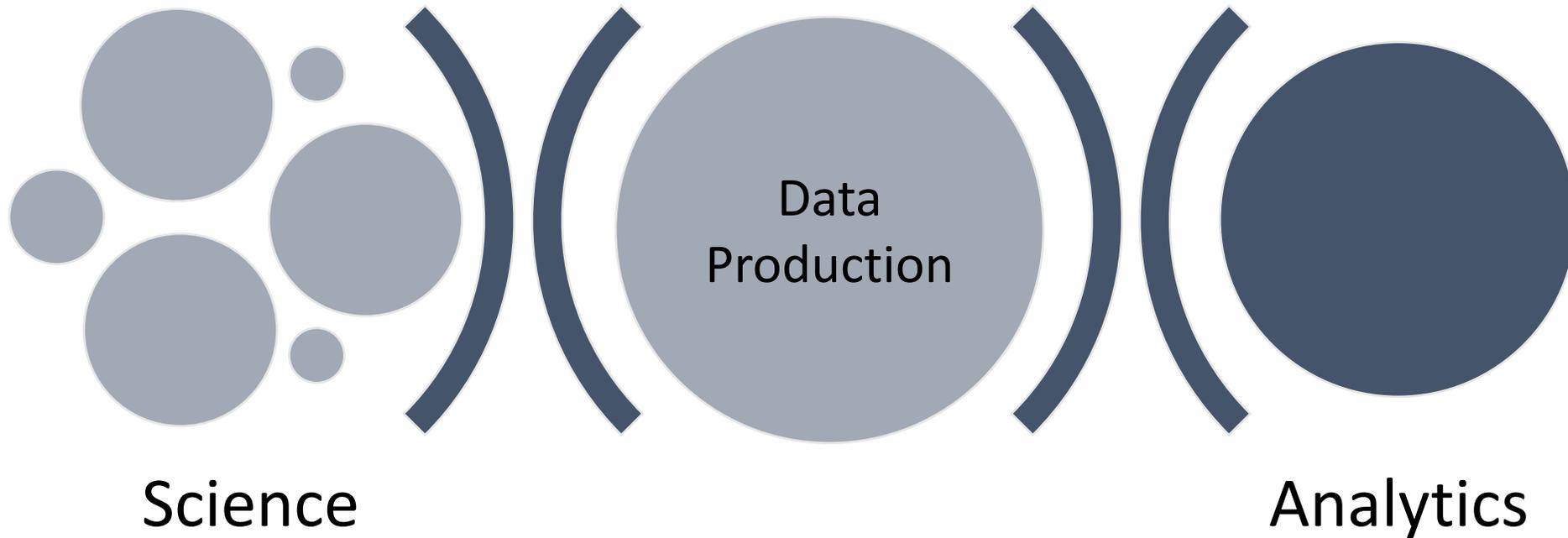


=



Researcher (Science) in Africa Countries

Data is at the center



Science (Research) plays its role based on data that are trustworthy, has provenance, etc

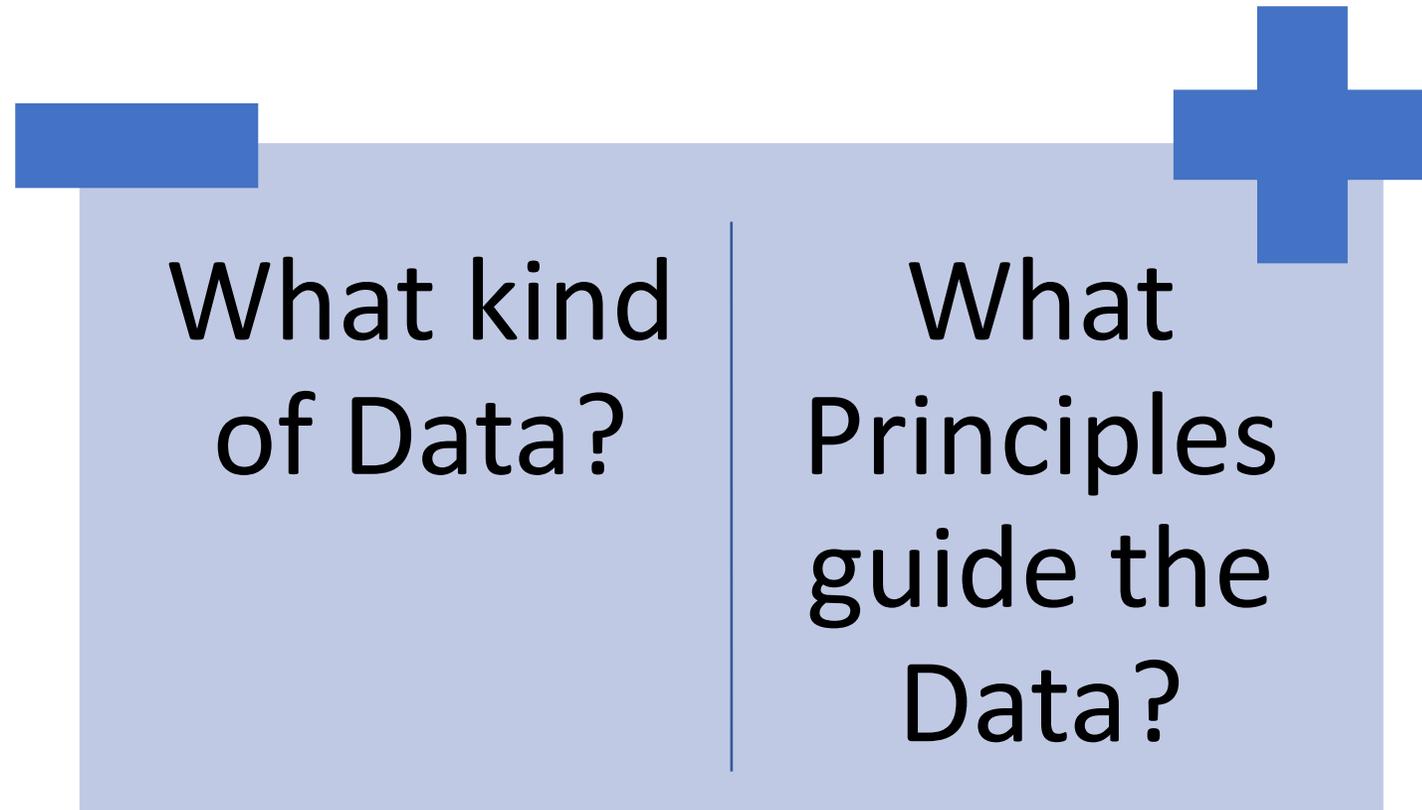
Data Collection/Production Process (Africa)

- The Question
 - Hectic
 - Repetitive/Circle
 - Collected and mostly not published
 - Data Vs Guiding Principles?



The BIG Question

- What is the right question to ask?



There is a need to support the discovery of Southern Knowledge through proper research management

There is the need for

- proper collection and annotation
- proper archiving
- provision of a 'long-term care' of valuable digital assets (including southern knowledge)
- ensure that they are discoverable and and re-used for downstream investigations, alone, or in combination with newly generated knowledge



A Different Kind of AI

—

FAIR Data Science

- The kind of Science that ask the questions of:
 - How do we provide access to critical data across locations
 - How do we ensure that the data is 'official'
 - How do we ensure data governance
 - How do we protect data ownership
 - How do we ensure that all Research data remains:
 - Findable
 - Accessible
 - Interoperable
 - Reusable
- Even after the 'completion' of the research and the researcher may have moved on to something else
 - How do we contribute to the work of others and build a strong research community
- To answer these kind of questions is an emerging and exciting Computing Research Field
 - The Science of FAIR Data

FAIR Principles

- **Data Production Process**
- Describe how research outputs should be organised so they can be more easily accessed, understood, exchanged and reused
- FAIR data is the default for most funding bodies
 - Maximizes the integrity and impact of their research investment.
 - Started as an EU Standard but now being adopted worldwide





What kind of Principles Guide the Production of Data?

Sounds New?

2020 Survey
on the
State of
Open Data

Approx 5000 Researchers

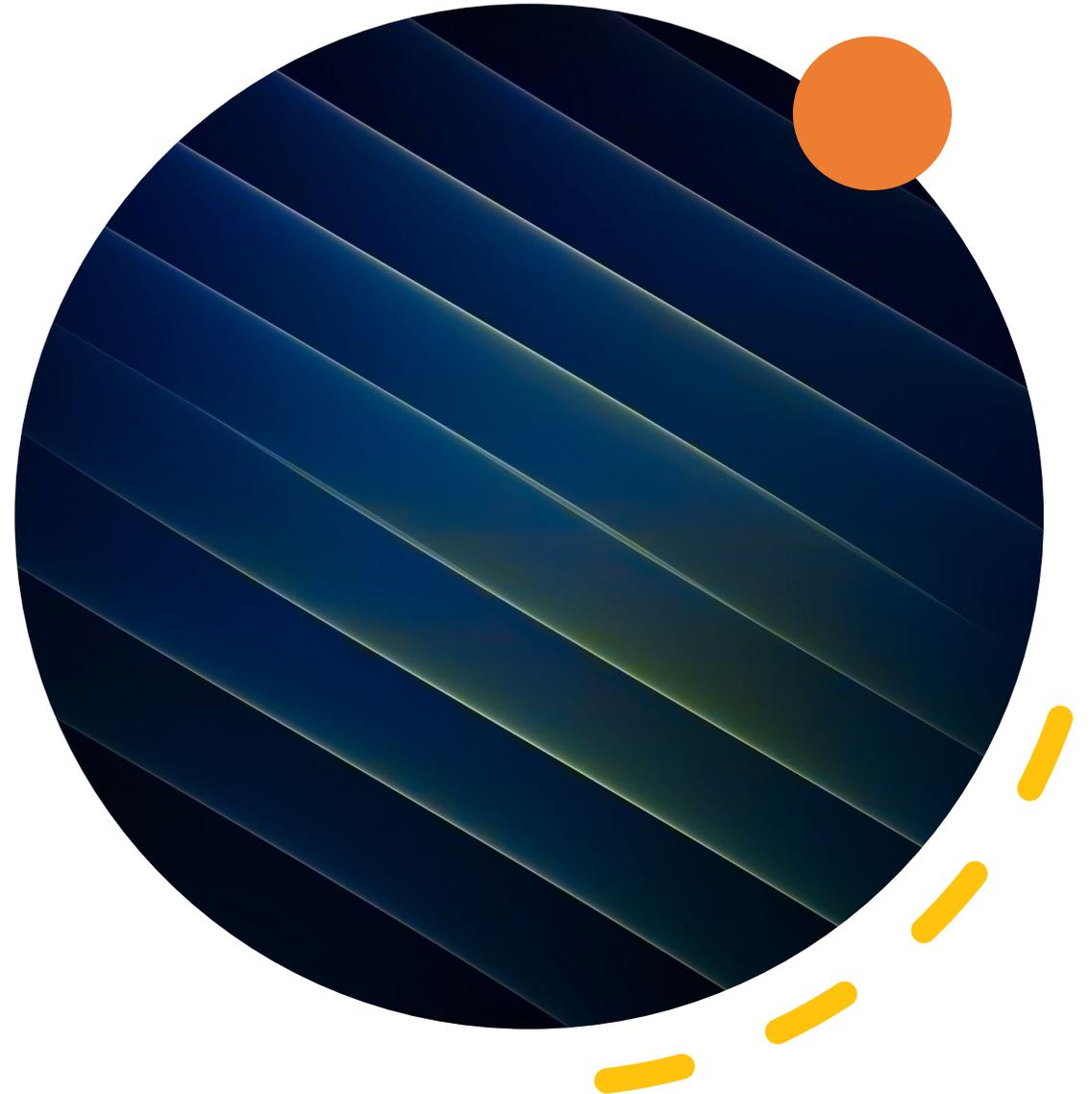
190 Countries

39% never heard of FAIR

More
awareness

=

More
Researchers
are getting
to know





The FAIR Principles

RESEARCH DATA - OPEN BY DEFAULT



14

- A set of principles to make digital resources (in this case data) Findable, Accessible, Interoperable, and Reusable—FAIR
 - FAIR principles provide guidelines
 - Do not require specific technologies
 - Allow communities of stakeholders to define specific FAIR standards
 - Develop metrics to quantify them

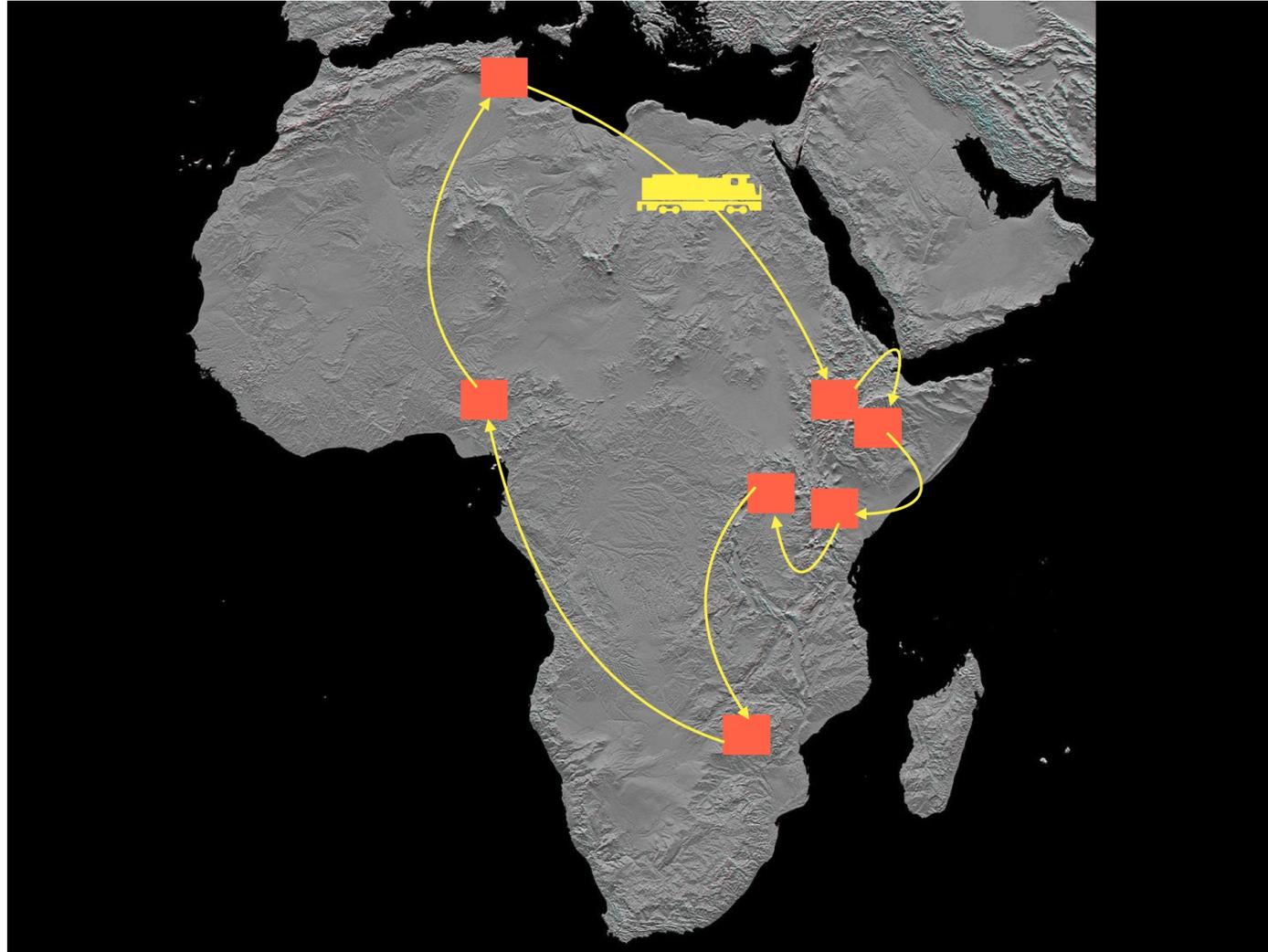
FAIRification

- The A. B. C of getting started on FAIRification
- A. Deposit your data in a repository. Data which are archived to a data repository are more likely to be accessible for the longer term. They are also easier to find, to reuse and to cite.
- B. Include a data availability statement (DAS) in your manuscript. This is a short narrative statement explaining to readers how and where your data can be accessed.
- C. Cite the data you have used, and that you have generated in your study.

<http://datadryad.org/>
<https://data.mendeley.com/>
<http://figshare.com/> <https://vivli.org/>
<http://zenodo.org/> <https://osf.io/>
<https://dataverse.harvard.edu/>
<https://www.agu.org/-/media/Files/Publications/Generalist-Data-Repository-Grid.pdf>



Successful case study: VODAN Africa & Asia





The Crises of COVID-19

- Consequences
 - Fears
 - Disruptions...
- Positives
 - eLearning
 - Collaborations and New Partnerships
 - Experiencing something new
 - Opportunities for knowledge sharing and convergence.
- Realizations
 - Relevance of ML and AI in the scheme of things and why universities should invest in it
 - Need for science to play a role based on data that are trustworthy, etc
 - Preparing For Future Outbreaks Consequences

Successful case study: VODAN Africa & Asia

- Learning from the Ebola Crisis in WA –what happened to the Data
- A network of group of universities, university hospitals and MoHs
- Deploying Artificial Intelligence in the support of the fight against the COVID-19 pandemic by:
 - Making available Machine-actionable data on COVID-19 to frontline health workers
 - Deploying FAIR Guiding Principles for COVID-19 Healthcare Data Management and Stewardship
 - Installing COVID-19 Machine-actionable FAIR Data points in 6 African Countries Uganda(2), Nigeria(2), Ethiopia(2), Kenya (1), Zimbabwe (1), Tunisia (1)

Successful case study: VODAN Africa & Asia

- Provide answers to questions like:
 - How do we provide access to critical data across locations (in Africa and now Asia)
 - How do we ensure that the COVID-19 data is 'official'
 - How do we ensure the COVID-19 data governance
 - How do we protect the ownership
 - How do we ensure that all COVID-19 data remains:
 - Findable
 - Accessible
 - Interoperable
 - Reusable

VODAN Africa & Asia: Team



Public Datasets

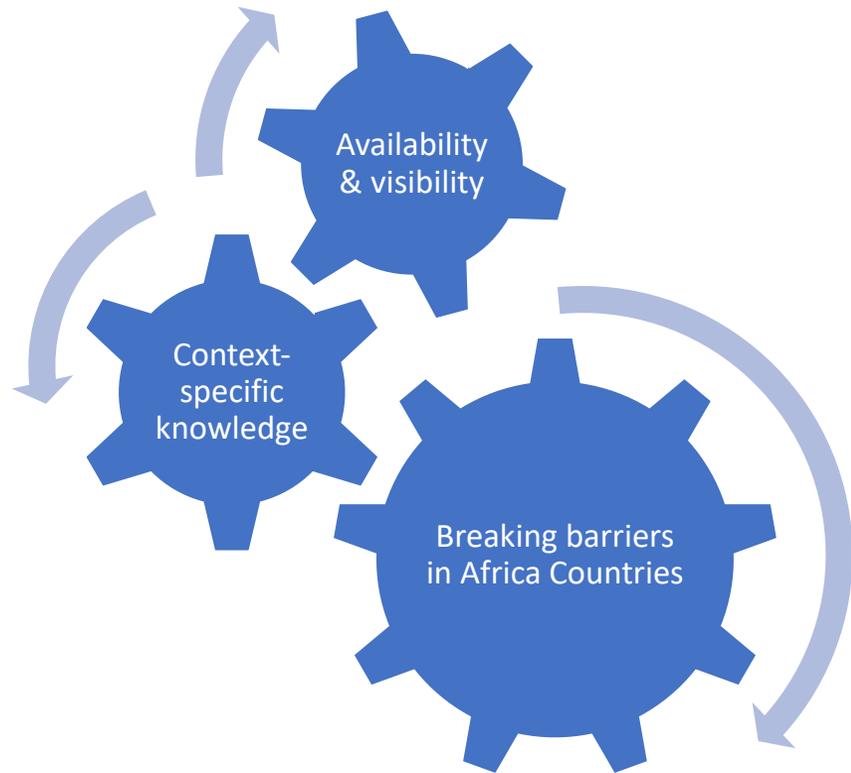
Oh Please!



What is Africa's
contribution to these
Public Datasets?



Intent



Make a case for the urgent need to improve the infrastructure supporting the discovery of Southern Knowledge

- Emphasize on enhancing the ability of machines to automatically find (and use) all digital research objects from the global south (data, models, algorithms, tools, workflows, ...)
- Show a case-study of successful implementation

