Project title	**DIGITAfrica** – Towards a Comprehensive Pan-African Research Infrastructure in Digital Sciences
Grant agreement #	101187966 (CSA - HORIZON-INFRA-2024-DEV-01-02)
Project duration	36 months (01/01/2025 – 31/12/2027)
Project URL	www.digitafrica.eu
Due date	30 / 06 / 2025
Submission date	30 / 06 / 2025
Dissemination level	Public
Version	1.0

D6.2 DATA MANAGEMENT PLAN

Responsible author(s): André Cardozo Sarli, Sébastien Ziegler, Renata Radócz (Mandat International)

Reviewers: Adam Belloum (University of Amsterdam), Cheikh Bamba Gueye (Université Cheikh Anta Diop)

Funded by the European Union

DIGITAfrica project has received funding from the EU Horizon Europe research and innovation Programme and Switzerland under Grant Agreement No. 101187966. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Executive Agency (REA). Neither the European Union nor the granting authority can be held responsible for them.

Document version history

Version	Date	Inputs	Responsible
0.1	22/04/2025	Initial draft	André Cardozo Sarli and MI Team
0.2	09/06/2025	Version for review	André Cardozo Sarli (MI)
0.3	17/06/2025	Internal Review	Adam Belloum (UvA)
0.4	20/06/2025	Internal Review	Cheikh Amadou Bamba Gueye (UCAD)
0.5	23/06/2025	Document sent for consortium review	André Cardozo Sarli (MI)
1.0	30/06/2025	Final version	André Cardozo Sarli (MI), Émilie Mespoulhes (SU)

Executive summary

This Data Management Plan of the DIGITAfrica Project provides detailed information about the compliance of the partners with the compliance, legal, and ethical requirements of the funders, and particularly the application of the FAIR principles.

The DIGITAfrica project, at its heart, consists on co-designing a research infrastructure for Digital Sciences in the partner African countries, through the development of a framework and an ecosystem. As such, DIGITAfrica involves both the application of a previously test set of blueprints and technologies and especially the development of a novel one, adapted to the local needs. Concerning the relevant data practices for existing datasets, the DIGITAfrica primarily exploits the analysis, outcomes, recommendations and expertise derived from previous projects and initiatives, rather than directly using specific datasets. However, the forthcoming activities have a great potential to generate new dataset related to digital infra-structure, such as experimental data, reference datasets, ontologies, trained models and software.

Annex I provides the Questionnaire that was distributed within the Consortium in order to further evolve the data management strategy of the partners, identify and address any upcoming needs.

Table of contents

List of tables

List of figures

Abbreviations

Abbreviation	Definition
AI	Artificial Intelligence
CA	Consortium Agreement
DMP	Data Management Plan
DoA	Description of Action
DPO	Data Protection Officer
EU	European Union
FAIR	Findable, Accessible, Interoperable, Reusable
GA	General Agreement
GCC	Global Code of Conduct for Equitable Research Partnerships
GDPR	General Data Protection Regulation
IPR	Intellectual Property Rights
MoU	Memorandum of Understanding
RI	Research Infrastructure
WP	Work Package

1 Introduction

DIGITAfrica is a Horizon Europe initiative funded under the call HORIZON-INFRA-2024-DEV-01, aiming to design a sustainable, long-term pan-African Digital research infrastructure. The project supports the African Union's Science, Technology, and Innovation Strategy for Africa (STISA-2024) by strengthening research infrastructures and technical skills in Information and Communication Technologies (ICT). DIGITAfrica intends to foster transformative Euro-African research and innovation collaboration, building on existing EU-AU partnerships and focusing on strategic fields to advance digital sciences, education, and training across Africa.

The project brings together partners from five African countries to collaborate with seven European institutions managing major digital research infrastructures (SLICES-RI and SoBigData RI) with the objective to exchange expertise and co-develop a strategic roadmap. Structured into seven work packages, DIGITAfrica will conduct needs assessments, initial design, capacity building, sustainability planning, and stakeholder engagement over 36 months starting January 2025. Key outcomes include a design study for the pan-African infrastructure, a network of research nodes with a shared agenda, a capacity-building portfolio prioritizing women and youth, and a BluePrint platform to pilot future collaboration. The Blueprint refers to a practical, openly available, and reproducible environment designed to lay the foundations for a pan-African research infrastructure (RI) in Digital Sciences Ultimately, DIGITAfrica aims to boost Africa's digital transformation, foster inclusive innovation, and strengthen AU-EU research cooperation with mutual benefits.

2 Methodology

The methodology for DIGITAfrica's Data Management Plan (DMP) is grounded in FAIR (Findable, Accessible, Interoperable, and Reusable) data principles, ensuring that data generated, collected, and processed throughout the project lifecycle is handled responsibly, transparently, and in compliance with relevant legal frameworks, including the General Data Protection Regulation (GDPR) and national data protection laws across participating African and international partners. To support this process, all project partners were asked to complete a comprehensive Data Management Plan questionnaire. This tool included dedicated sections on the types of data collected or generated, storage and security measures, adherence to FAIR data principles, ethical considerations, the use of artificial intelligence, and intellectual property rights (IPR). The responses provided a structured and comparable foundation to develop a shared and robust DMP across the consortium.

3 About this deliverable

The European Commission defines[1] Data Management Plans (DMPs) as key elements of good data management. A good DMP should describe the data management life cycle for the data to be collected, processed and/or generated. As part of making research data findable, accessible, interoperable and re-usable (FAIR), a DMP should include information on:

- The handling of research data during and after the end of the project;
- What data will be collected, processed and/or generated;
- Which methodology and standards will be applied;
- Whether data will be shared/made open access;
- How data will be curated and preserved (including after the end of the project).

Deliverable D6.2 serves as the data management plan for research within the DIGITAfrica consortium, outlining ethical and GDPR-compliant data handling practices. Produced under Work Package 6 – "Dissemination, Exploitation and Communication," this deliverable summarizes the types of data generated throughout the project and the associated data processing activities.

This version is the initial and sole release of the plan, intended to be a living document that will be regularly updated based on partners' inputs. It will reflect any changes related to data generation, usage, processing, storage, and ethical considerations throughout the project's duration.

4 Data Summary

4.1 Overview

Within DIGITAfrica, data plays a central role across all phases of the project, supporting the production of key deliverables, informing design decisions, and enabling collaboration between African and European partners. The data collected encompasses a broad range, including research outputs, operational data, models, software, and knowledge gathered through workshops and consultations. These inputs are essential to the project's objectives and ensuring its outputs are relevant, evidence-based, and tailored to the African context.

Most project deliverables are directly dependent on data collection and analysis. WP1's demand analysis (D1.1, D1.2) relies on stakeholder inputs and workshops, while the BluePrint and Design Study deliverables (D2.1, D3.1, D5.2) synthesize project-wide findings

[1] https://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-data-management/data-management_en.htm.

into concrete proposals for the future research infrastructure. Similarly, capacity-building deliverables (D4.2, D4.3) require data on existing competencies and training needs, and the Report on open research data and reproducibility (D3.2) addresses the project's internal data practices regarding open access.

Work Package 1 focuses on mapping needs and gaps in existing Digital Sciences RIs through surveys, stakeholder consultations, and analysis of prior initiatives. These findings guide subsequent work, including the prioritization of services to be developed in WP2. Here, tasks such as requirements gathering (Task 2.1) and Proof of Concept deployment (Task 2.4) involve technical and operational data collection to validate initial RI designs and identify infrastructure and skill needs.

Work Package 4 uses data-driven insights to tailor education and training efforts. By identifying priority competencies and gaps, particularly among female and young researchers, WP4 supports the development of a capacity-building portfolio that includes curricula, credentials, and competence frameworks aligned with local needs. These efforts ensure training activities are relevant, inclusive, and effective.

The BluePrint cooperation platform, developed in WP2 and finalized in WP3, will provide a space to test and demonstrate the project’s vision for a future pan-African RI. It aims to support joint activities among AU and EU researchers by offering ICT services and handling various types of data, including synthetic datasets and operational performance data. The BluePrint acts as both a collaborative tool and a testbed for refining infrastructure concepts.

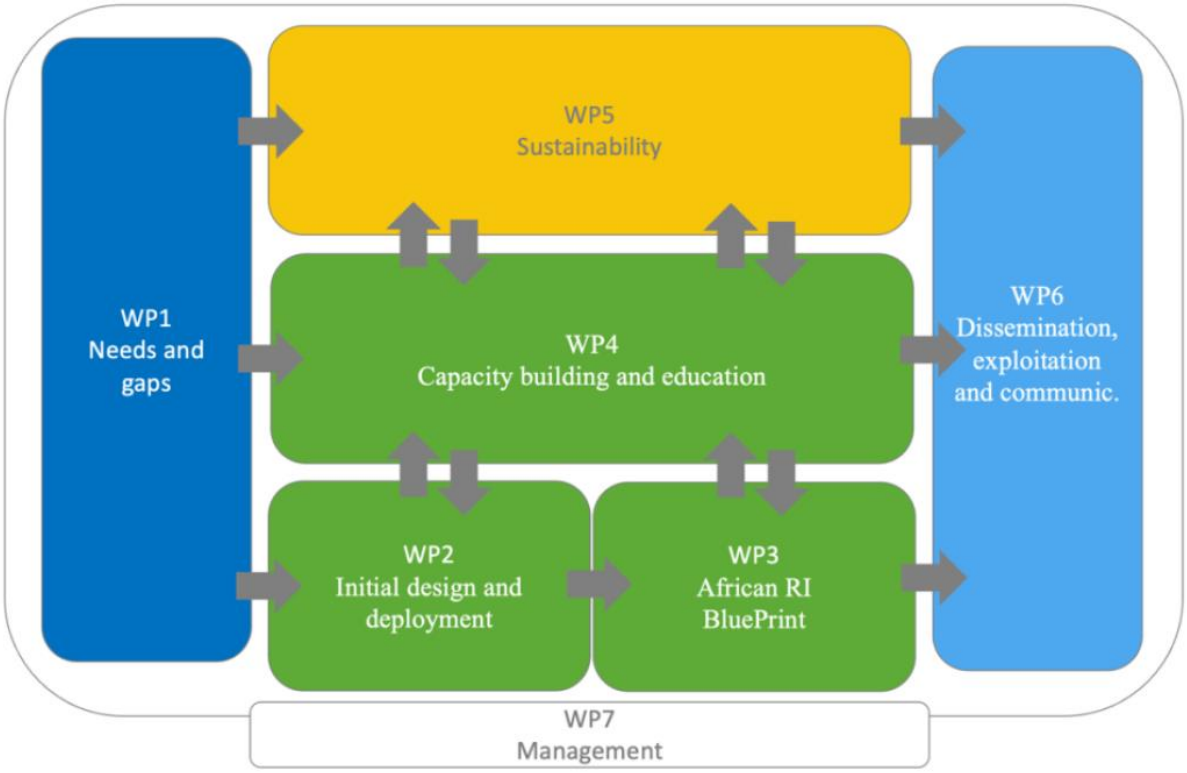

Figure 1 - Overall Workplan of DIGITAfrica

4.2 The datasets

DIGITAfrica engages with datasets in two main ways: through the mobilisation of existing datasets and the co-production of future datasets throughout the project's development. This Data Management Plan, in its initial iteration, is drafted at the stage where the main activities of Work Packages 2, 3, 4 and 5 are still in planning and preparation. Therefore, most of the datasets underlined in existing are auxiliary datasets from other work packages (1, 6, and 7), and the forthcoming activities are only estimated at this stage.

For Work Packages 2, 3, 4 and 5, it is important to explain that while DIGITAfrica draws on the technical infrastructures and methodological resources of the SLICES-RI[2] and SoBigData[3] initiatives, this engagement is primarily focused on making use of expertise and reusable tools rather than directly importing datasets. In this sense, the use of RI-based infrastructures is more about methodological alignment and exchange than about data

[2] SLICES is a flexible platform designed to support large-scale, experimental research focused on networking protocols, radio technologies, services, data collection, parallel and distributed computing and in particular cloud and edge-based computing architectures and services. Website: https://www.slices-ri.eu/

[3] SoBigData is a European open science research infrastructure for social mining according to EU values of fairness and privacy protection. Website: https://www.sobigdata.eu/

transfer. It is worth noting that the European partners in DIGITAfrica are themselves partners of these projects.

4.2.1 Existing datasets

As the project is still in its early stages, most of the working datasets currently serve as auxiliary resources supporting DIGITAfrica's main objectives. Partners are primarily gathering data for dissemination activities and event organization under Work Packages 6 and 7. For these purposes, the nature of the existing data is essentially personal, albeit limited to elements related to capacity and programs assessment. The rationale is that knowing and connecting to stakeholders and early dissemination justify the responsible use of minimized personal data, as seen in Table 1.

To clarify, two work packages have advanced into their tasks with the use of existing datasets and establishment of new ones. In in Work Package 1—particularly through Deliverable 1.1 (Tasks 1.2 and 1.3)—partners focused on (i) providing a comprehensive overview of the existing digital research infrastructure landscape in five African partner countries (Tunisia, Senegal, South Africa, Kenya, and Cameroon), and (ii) identifying the demand and strategic needs of the research and innovation ecosystems through stakeholder consultations. In Tunisia specifically, partner UMA utilized a pilot stakeholder survey dataset to support this effort.

In Work Package 4, the partners are organizing a survey to assess digital skills and training needs across Africa. The objective is to observe possible skill gaps in order to tailor micro-credentials and training pathways. This survey, under tasks 4.1 (Identification of the priority needs) and 4.2 (Articulation with the SLICES/SoBigData Academy and TheNetworkingChannel)

The general summary of existing datasets is represented below:

Dataset	Partner	Description	Contains Personal Data	Purpose	Data Format	Storage Location	Work Package
Stakeholders list	All	List of stakeholders of potential interest to the DIGITAfrica project	Names, email addresses, and organizational affiliations	Dissemination, exploitation and communication of DIGITAfrica activities	.xls	Project Drive	WP6
DIGITAfrica Skills Survey Responses	WP4	Survey responses assessing digital skills and training needs across Africa.	Names, email addresses, and organizational affiliations	To assess skill gaps and training needs for designing micro-credentials and training pathways	.xls, .csv	Project Drive	WP4
2018 Directory - Research Structures	UMA	This document provides data on Research Structures,	It lists the names and email addresses of	The dataset is used to run a survey on the state of IT	PDF	University Drive	All

		Shared Research Service Units, Doctoral Schools, and national and international cooperation programs.	the heads of the research structures	research infrastructure within research structures, and to identify needs and gaps			
Mailing list	UTH, CNR	Mailing lists created through the events participants, newsletters, questionnaires, etc	Yes: name, email, affiliation	Organisation of events, dissemination activities	.xls	Project drive	All

Table 1 - Overview of Existing datasets

4.2.2 Forthcoming datasets:

DIGITAfrica is expected to generate a variety of datasets as the project evolves, particularly through activities related to stakeholder engagement, capacity building, and the deployment of digital research infrastructures. At this stage, the specific formats, collection methodologies, and metadata standards for these datasets are still under development. Not necessarily these tasks will yield datasets, but they also might yield white papers, general reports or other form of publications. These future activities will be defined progressively in collaboration with project partners and communities involved:

Dataset	Partner	Description	Contains Personal Data	Purpose	Data Format	Work Package
DIGITAfrica BluePrint (versions 1 and final), Blueprint assessment and lessons learned	INRIA, UCT, STR, UTH, TUB, UCAD, UN, UMA, CNR, SU, UvA	Includes specifications for baseline services, technical choices, reference architecture for the pan-African Digital Sciences RI, and assessment of the Blueprint's application.	No.	To define a practical framework and a "playground" for the future RI, empowering research communities.	Documents and reports outlining specifications, design, and architecture.	WP2 – Initial design and deployment and WP3 – African RI BluePrint
Report on open research data and reproducibility	UvA, CNR, INRIA, STR, UN, MI	A comprehensive report on DIGITAfrica's approach to open research data and reproducibility, integrating European and	Unlikely directly.	To ensure the future RI adheres to FAIR principles and promotes Open Science.	Document, report.	WP3 – African RI BluePrint

		African experiences.				
Guidelines for female and young researchers, Competence framework and modular curricula, Proposed Qualification and credentials on digital technologies competences and skills	UTH, STR, UCAD, SU, BSC, CNR, UMA, UN, UCT, UvA, TUB	Includes training needs, recommendations for female and young researchers, a competence framework, curricula, and proposed qualifications in digital technologies.	Potentially aggregated demographic data for analysis.	To enhance education and skills in Digital Sciences, addressing identified gaps and empowering female and young researchers.	Documents, reports.	WP4 – Capacity building and education
Preliminary requirements for the design study, DIGITAfrica Design study, DIGITAfrica network, research agenda and MoU	UMA, SU, CNR, BSC, UTH, UCT, UCAD, UN, UvA, STR, TUB	Outlines initial requirements, a comprehensive design study for the future RI (including gender aspects and sustainability), and a Memorandum of Understanding (MoU).	No.	To identify conditions for sustainability, define a design pathway for the RI, and establish governance principles.	Documents, reports.	WP5 – Sustainability

Table 2 - Overview of forthcoming activities

4.2.3 Data storage retention and management

Within DIGITAfrica, all datasets containing personal or confidential information are securely stored and managed by designated project partners, who are responsible for implementing security updates and applying the technical and organizational safeguards required under relevant data protection regulations. To support secure collaboration and data handling from the outset, a dedicated repository for internal project work was established by Sorbonne University in January 2025. Best practices regarding data storage, access control, and information security were communicated to all partners and team members to ensure consistency across the consortium.

Publicly available data related to the DIGITAfrica project is hosted on the project's official website (www.digitafrica.eu). This includes information such as:

- Project objectives, missions, expected impact, and structure;
- A list of partners and their roles in the work packages and governance;
- Announcements of events and public dissemination activities;
- Public deliverables;
- Open-access scientific publications and media content.

Non-public materials—such as internal reports, presentations, datasets containing personal or sensitive information, and confidential deliverables—are stored in a private repository (DropSU and OneLab) managed by the coordinator Sorbonne University. These are only accessible to consortium members and governed by the terms of the Grant Agreement and Consortium Agreement. Any datasets intended for open access that include personal information (e.g. interviews, pilot studies, or focus groups) must undergo anonymisation before release, in accordance with the European Union's GDPR Guidelines.

In addition to the shared repositories, partners may also store data on their own institutional servers or secure facilities. Data retention follows these general principles:

During the project, data is available and accessible on all relevant servers for as long as necessary to support ongoing work.

After the project ends, the public website and project repository will be maintained until project completion. Afterward, wherever feasible, datasets will be anonymised and made accessible in accordance with FAIR data principles. Specific procedures for long-term storage and publication, especially for data of a sensitive nature, are under discussion, and will take into account the Memorandum of Understanding (MoU), especially Task 5.4, and future legacy activities of DIGITAfrica.

Beyond the provisions outlined above, the Table 3 presents the data storage, management and retention policy as declared by project partners. Since these are additional measures, we are only sharing the policies of partners that declared additional retention and storage measures.

Partner	Data Storage Management & Retention Policy
BSC	Data is stored on secure institutional servers and/or GDPR compliant platforms such as Google Workspace or institutional repositories, with restricted access and encryption where applicable. Data will be stored for 2 years post-project completion
INRIA	According to SLICES Metadata Registry System and SLICES Data Management Infrastructure[4]
MI	Data is kept on secure institutional servers, and data will be stored for the project's duration and will be kept for a maximum 5 years after the project's end.
STR	Data will be retained for a minimum of 5 years post-project, in line with institutional policies and funder guidelines. Anonymized datasets intended for reuse in future research or training will be preserved indefinitely in public or institutional repositories.
SU	Data is stored in the Project's dedicated drive (DropSU= and locally (+security copy on Owncloud – hosted by SU - and external hard drives) and will be kept for maximum 5 years after the project's end.

[4] https://doc.slices-ri.eu/BasicServices/MRS/MRS.html

UCT	All personal data is stored on UCT's institutional systems in secure, access-controlled environments, in compliance with the Protection of Personal Information Act (POPIA). Backups are maintained on encrypted drives. Personal data will be retained for 5 years post-project, or earlier if required by ethical considerations or upon request of data subjects.
UN	No additional data storage policy. Data will be retained for 5 years after the project's completion to allow for reporting and evaluation, and then securely deleted.
UvA	According to UvA internal policies and the project lifetime.

Table 3 - Data storage retention and management

5 FAIR Data

In its FAIR Data Management Horizon 2020 Guidelines, the European Commission notes that "Good research data management is not a goal in itself, but rather the key conduit leading to knowledge discovery and innovation, and to subsequent data and knowledge integration and reuse". Therefore, beneficiaries are explicitly encouraged to make their research data findable, accessibly, interoperable and reusable (FAIR).

In alignment with Horizon Europe's open science requirements, this project will ensure open access to all peer-reviewed scientific publications resulting from the research. A machine-readable version of each publication—either the published version or the final peer-reviewed manuscript—will be deposited in a trusted scientific repository at the time of publication. These publications will be made immediately accessible under the latest Creative Commons Attribution license (CC BY), or a license with equivalent rights. For long-text formats like monographs, licenses permitting non-commercial use or prohibiting derivatives (e.g., CC BY-NC, CC BY-ND) may be used. All deposited publications will be accompanied by metadata compliant with the FAIR principles and openly licensed (e.g., CC 0), including persistent identifiers and funding information.

Open access to data will be ensured through licensing under CC BY, CC 0, or equivalent, with consideration for legitimate interests such as intellectual property protection or confidentiality. When access is restricted, the DMP will include justifications. Metadata for the datasets will also be openly available and machine-actionable, ensuring transparency and reusability.

Where additional open science practices are required by the call, such as during public emergencies, the project will comply by depositing outputs in trusted repositories and providing immediate access under CC BY or CC 0. If open access is not feasible, the beneficiaries will grant non-exclusive licenses under fair and reasonable terms. These measures ensure that research outputs remain accessible and verifiable, supporting scientific integrity, transparency, and societal benefit.

5.1 Limitation

It is important to note that at the time of drafting this DMP many activities that may generate datasets and open access resources are still in the planning phase. As a result, partners' efforts toward ensuring findability, accessibility, interoperability and reusability remain limited at this stage.

5.2 Findability

According to this principle, data should be easily discoverable and identifiable through the use of persistent identifiers and descriptive metadata. In the questionnaire, partners were asked to reflect on whether their data will be findable. The Table 4 provides a high-level summary of the responses.

Partner	Findability Approach	Metadata Standards	Persistent Identifiers	Metadata Registration
BSC	Not specified	Dublin Core for general metadata and DataCite for dataset citation metadata. Metadata will include fields such as title, creator, date, subject, methodology, and coverage	DOI through Zenodo. Additional identifiers may include ORCID IDs for contributors and PURLs for supporting documentation	Zenodo, OpenAire, DataCite
CNR	Not applicable as the only processed data is private.	Not applicable	Not applicable	Not applicable
INRIA	Centralized Metadata Registry System (MRS) from SLICES-RI with RESTful APIs and search portal enables easy discovery of all digital objects (datasets, software, tools, services).	Uses a hierarchical metadata profile structured around SFDO (SLICES Fair Digital Objects), aligning with Dublin Core, DataCite 4.3, DCAT 2.0, EOSC, and RDA IG	Persistent IDs are included (e.g., DOI) and information on identifiers is part of the “Primary” and “Management” metadata .	Metadata registry centralizes SFDO info including internal/external identifiers, versioning, license, etc.
MI	Not applicable as data for internal purposes. Website data is public	Not applicable	Where applicable, DOI	Website
STR	All datasets will be assigned persistent identifiers (e.g., DOIs) through deposition in trusted open-access repositories such as Zenodo.	Each dataset will be accompanied by rich metadata, following standard schemas (e.g., Dublin Core, DataCite) to enable easy discovery by both humans and machines.	DOI, Zenodo	Metadata will be indexed in searchable catalogs and cross-linked with related publications and project documentation.

SU	Not applicable: only personal data	Not applicable	Not applicable	Not applicable
TUB	TUB will mostly work with documentation and software rather than personal data. If any data will be available, TUB will follow the guidelines from the Research Data Management Office.	Dublin Core	DOI	Zenodo
UCAD	Not applicable as no dataset or data processing is foreseen at this point	Not applicable as no dataset or data processing is foreseen at this point	Not applicable as no dataset or data processing is foreseen at this point	Not applicable as no dataset or data processing is foreseen at this point
UCT	Aims to make all data open, if possible	DublinCore for documentation, DataCite metadata for DOI	Where publicly deposited, datasets will receive DOIs (via Zenodo or UCT institutional repository)	Metadata will be registered on the UCT Data Repository and/or Zenodo, and will be searchable via OpenAIRE
UMA	Not applicable to the processed dataset, which is kept internal	Not applicable	Not applicable	Not applicable
UN	Central catalog to make data discoverable.	Simple tabular dataset metadata that will include a general description of the data, the meaning of each column, and the data type of each column	DOI through Zenodo	Not specified
UTH	Not applicable to its current datasets as it contains personal data	Not applicable	Not applicable	Not applicable
UVA	Will make available public reports, scientific publications, educational frameworks, policy documents, and other openly available knowledge resources. They will potentially generate synthesized reports and analyses related to digital skills in Africa.	Dublin core metadata	DOIs for repositories such as ZENODO	Metadata will be registered with the Zenodo repository and indexed via search engines and repository indexing services to ensure broad accessibility

Table 4 - Summary of Findability measures at the partner level

5.3 Accessibility

According to this principle, data should be made available in a way that allows authorized users to access it, ideally through standard protocols and clear access conditions. In the questionnaire, partners were asked to indicate how they plan to ensure data accessibility. The Table 5 summarizes their responses.

Partner	Accessibility Approach	Repository / Storage	Access Protocols	Authentication / Authorization	Open / Public Access
BSC	Available unless embargoed	Zenodo	HTTPS.	No, unless data contains sensitive material. All anonymized datasets will be open access. Restricted access data (e.g., sensitive national information) will be subject to institutional or project-based review	Datasets will be made available after project review and anonymization
CNR	Not applicable for their datasets so far	Not applicable	Not applicable	Not applicable	Not applicable
INRIA	Metadata stored in a PostgreSQL Repository backend, accessible via RESTful API secured by SLICES AAI (OIDC/JWT)	Files stored in object storage (via DMI) with pre-signed URLs (HTTP PUT/GET) for upload/download	HTTP PUT/GET, REST API; authentication via OpenID Connect (Keycloak/SLICES AAI) registry system	Web portal (Angular SPA) and API use Bearer JWT authentication; anonymous users have restricted access	To be defined with the consortium
MI	Project website for public documents	Project website	HTTPS	No, unless data contains sensitive material. All anonymized datasets will be open access. Restricted access data (e.g., sensitive national information) will be subject to institutional or project-based review	The data on project website in open.

STR	Where necessary (e.g., for sensitive or personal data), access controls and data use agreements will be implemented while still ensuring that metadata remains publicly accessible.	Institutional Dspace	Data will be made available through open repositories with defined access protocols (e.g., HTTPS, FTP).	Data access policies will be clearly described, including any embargo periods or licensing terms.	To be defined
SU	Not applicable: only personal data	Not applicable	Not applicable	Not applicable	Not Applicable
TUB	If the case, data will be made accessible via open repositories.	GitHUB	HTTPS, API access, using PAT, two-factor-authentication and ssh-keys for automated access etc all these available via Github	To be defined	To be defined
UCAD	Not applicable	Not applicable	Not applicable	Not applicable	Not applicable
UCT	Data will be made accessible via open repositories, with embargoes where necessary for internal project phases	UCT Data Repository (https://ziva.uct.ac.za) and/or Zenodo (https://zenodo.org)	HTTPS and secure login protocols. Public datasets will use open access links with persistent identifiers	Required only for internal datasets and drafts (via UCT credentials or shared tokens). Open datasets will not require login.	After 12–18 months embargo period (aligned with project deliverables and publications)
UMA	Not applicable	Not applicable	Not applicable	Not applicable	Not applicable
UN	All data will be downloadable from the project website. All data will be	Project website	HTTPS	Only access to personal data will require authenticated user accounts	All data will be released under an open access license immediately

	released under an open access license immediately upon generation				upon generation
UTH	Not applicable	Not applicable	Not applicable	Not applicable	Not applicable
UVA	Data will be made publicly available immediately upon curation and internal approval, with no embargo period.	Zenodo	HTTPS	None of open access data	Not required

Table 5 - Summary of Accessibility measures at the partner level

5.4 Interoperability

According to this principle, data must be designed to work with other data and systems, using common standards and vocabularies. Interoperability is one of the main targets of the DIGITAfrica project as the planned milestones and deliverables are thought, planned and optimized to be applied in other African countries, including the potential integration of new partners. In the questionnaire, partners were asked to indicate how they plan to ensure interoperability. The Table 6 summarizes their responses.

Partner	Data Formats	Metadata & Documentation Standards	Shared Vocabularies/Ontologies	Compatibility Measures
BSC	Data will be stored in machine-readable and widely used formats such as CSV and JSON for structured tabular data. Documentation and narratives will be in PDF/DOCX, while metadata descriptions will use README.md and .txt	Dublin Core for general metadata and DataCite Schema for citation and attribution	Align responses with vocabularies such as UNESCO ISCED, OECD FOS, and EUROVOC, where applicable	standardized file formats and metadata schemas and by adhering to FAIR principles2. This ensures easy integration with EU-wide research infrastructures and initiatives like SoBigData++ and SLICES
CNR	Not applicable	Not applicable	Not applicable	Not applicable

INRIA	MRS supports converter tools exporting/importing metadata to EOSC, Dublin Core, DataCite, RDA, ensuring semantic and technical interoperability	Maintains mappings between SFDO and external metadata models	Uses standard metadata schemas and optionally SigMF (domain-specific) .	Metadata exportable via API to EOSC; import functionality supports cross-platform integration.
MI	PDF and .docx for deliverables published in the website	Project's templates	Project's templates	Not applicable (website data)
STR	Data will be stored in widely accepted, machine-readable formats like CSV, JSON, RDF, XML that are compatible with common data analysis tools.	Metadata will adhere to community standards and will be structured for integration with other data sources.	We will adopt standardized vocabularies, ontologies, and classification systems like domain-specific ontologies to ensure semantic interoperability.	By storing data in widely accepted, machine-readable formats (e.g., CSV, JSON, RDF, XML) that are compatible with common data analysis tools
SU	Not applicable (personal data)	Not applicable	Not applicable	Not applicable
TUB	Data will be stored in interoperable formats and machine readable. CSV, JSON, RDF	To be defined	TBD: IEEE Taxonomy, DOI, DCAT	Aligned with the project.
UCAD	Not applicable	Not applicable	Not applicable	Not applicable
UCT	CSV, JSON, XLSX	Data will be documented using community standards, and README files will accompany each dataset	Not applicable	will ensure compatibility by aligning datasets with those from other WP partners
UMA	Not applicable	Not applicable	Not applicable	Not applicable
UN	CSV, XLS	Not specified	Not specified	Not specified
UTH	Not applicable	Not applicable	Not applicable	Not applicable
UVA	Reports will follow standard formats (PDF, TXT, CSV); metadata uses controlled vocabularies	Data will be stored in machine-readable formats such as CSV and JSON. Reports will also be provided in human-readable formats like PDF.	Publicly available ontologies and vocabularies where applicable	Data structure and documentation will be validated against repository and community standards.

Table 6 – Summary of Interoperability measures at the partner level

5.5 Reusability

The ultimate goal of FAIR is to optimise the reuse of data. This can be achieved, when metadata and data are well-described so that they can be replicated and/or combined in different settings. In the questionnaire, partners were asked to indicate how they plan to ensure reusability The Table 7 summarizes their responses.

Partner	Data Reusability Approach
BSC	BSC ensures reusability by accompanying each dataset with detailed documentation (README, data dictionary, methodology, license) and implementing version control for all updates.
CNR	Not applicable
INRIA	Enhances reusability by supporting type/domain-specific metadata and automatic metadata production, improving both machine-actionability and interoperability. Includes type/domain-specific attributes for more precise metadata. Automates metadata production for streamlined data handling and reuse.
MI	Not applicable
STR	Each dataset will include clear licensing information (e.g., Creative Commons Attribution 4.0 International – CC BY 4.0) to facilitate reuse. Comprehensive documentation and data dictionaries will accompany each dataset, describing data sources, processing methods, and quality checks.
SU	Not applicable
TUB	Research data reusability is ensured through version control, standardized metadata, and automated validation.
UCAD	Not applicable
UCT	Will ensure data reusability through comprehensive documentation (README files, methodology briefs, data dictionaries), licensing (CC BY 4.0 or CC BY-NC 4.0), and quality control procedures including version control via Git and internal checks.
UMA	Not applicable
UN	The University of Ngaoundéré will ensure reusability by employing data versioning and provenance tracking and providing a data dictionary
UTH	Not applicable
UVA	All outputs will be released under an open license (e.g., CC BY 4.0) and will be accompanied by sufficient documentation, including README files and references to metadata. Provenance and version history will be maintained

Table 7 - Summary of Reusability measures at the partner level

6 Ethics, Data Protection & Security

DIGITAfrica as a project, its consortium and the actions related to the project have complied with ethical and legal principles, standards and regulations. This includes undertaking

activities in compliance with ethical principles and applicable international, EU and national law. The most important guiding principles are outlined in this section.

6.1 Data Protection

As part of the DIGITAfrica project, all beneficiaries, including those based in non-EU countries, are required to comply with the General Data Protection Regulation (GDPR) and other relevant data protection laws, such as the regulation he processing of personal data by the Union institutions, bodies, offices and agencies (2018/1725).

Mandat International oversees the project through a continuous legal monitoring process to ensure all data-related activities remain in line with current legislation, particularly concerning personal data and the use of AI. African partners will also be responsible for ensuring alignment with their own national data protection laws.

In accordance with the General Agreement and in harmony with the above said regulations, the partners agree and pursuit to ensure that any personal data is:

- processed lawfully, fairly and in a transparent manner in relation to the data subjects
- collected for specified, explicit and legitimate purposes and not further processed in a manner that is incompatible with those purposes
- adequate, relevant and limited to what is necessary in relation to the purposes for which they are processed
- accurate and, where necessary, kept up to date
- kept in a form which permits identification of data subjects for no longer than is necessary for the purposes for which the data is processed and
- processed in a manner that ensures appropriate security of the data.

6.2 Alignment with the Global Code of Conduct for Equitable Research Partnerships

The “Global Code of Conduct for Equitable Research Partnerships” (“GCC”) or the “Trust Code” is developed by the TRUST project to ensure equitable, respectful, and responsible international research partnerships, and especially cater to situation where there is a disparity of available resources and relevant cultural differences. In August 2018, the European Commission adopted the GCC as a mandatory reference document for Horizon 2020. The University of Cape Town adopted it in 2019, being the first University to do so. The GCC is expressly mentioned in the DoA, and herein DIGITAfrica reinforces its commitment to adopt the GCC and details its approach at the data level.

The GCC is structured around four key ethical values, which are fairness, respect, care, and honesty, and includes 23 articles designed to guide all phases of international research collaboration. In the next sections, we will detail how DIGITAfrica aligns with these principles:

6.2.1 Fairness: Equitable Partnerships and Local Ownership (Articles 1–7)

DIGITAfrica adopts a co-construction model with African partners and stakeholders, explicitly fostering a "partnership of equals" as articulated in Article 2 (Inclusion of local communities and participants). The project is grounded in mutual goal-setting, with African institutions centrally involved in identifying local needs, defining the research agenda, and co-developing tools and infrastructures.

The DIGITAfrica Blueprint, a key deliverable, will be hosted on African nodes and shaped through stakeholder consultation. This supports Article 1 (Local relevance of research) and Article 4 (Inclusion of local researchers) by ensuring that research outputs are contextually grounded, usable, and under local control. The project also prioritizes capacity building, in line with Articles 3 (Feedback to local participants) and 7 (Fair compensation of local support systems), through training, joint educational activities, and mobility schemes.

6.2.2 Respect: Cultural Sensitivity and Informed Engagement (Articles 8–11)

DIGITAfrica places a strong emphasis on community engagement, starting with a survey of existing capacities (Task 1.1), followed by consultation workshops and gap analyses. These efforts reflect Article 8 (Respect for cultural sensitivities) and Article 9 (Community assent through local structures), by ensuring that research activities are designed with and not merely for local communities.

Task 1.1 also focuses on early and ongoing stakeholder involvement, including "direct end-users: the R&I communities." This participatory model reinforces the importance of inclusive governance and ensures that research directions are not externally imposed but co-defined with those most affected.

6.2.3 Care: Minimizing Harm and Promoting Local Wellbeing (Articles 12–19)

DIGITAfrica incorporates the value of care through its attention to both technical and social risks, fulfilling Article 13 (Feedback and complaints procedures) and Article 15 (Protecting participants from stigmatization and harm). Potential harms related to AI and digital technologies, such as algorithmic bias, lack of explainability, or data misuse, will be strictly monitored, and project outputs will conform to both EU AI regulations and local oversight mechanisms.

The project also integrates an environmental dimension, working with the GreenDIGIT initiative to reduce the ecological footprint of digital infrastructures. This addresses Article 18 (Environmental protection), promoting long-term benefit and avoiding unintended

negative externalities. Moreover, DIGITAfrica aims to generate locally developed solutions in critical sectors like agriculture and water management, ensuring that the knowledge triangle of research, education, and innovation yields direct societal benefit.

6.2.4 Honesty: Transparency, Data Integrity, and Legal Compliance (Articles 20–23)

DIGITAfrica aligns with the TRUST Code's final cluster of principles around honesty and integrity by committing to open science, robust data governance, and legal oversight. The project ensures open access to publications, data, software, and models through trusted repositories, adhering to FAIR data principles and promoting reproducibility, as advocated in Article 20 (Clear understanding of roles and responsibilities) and Article 21 (Honest, comprehensible communication).

This Data Management Plan establishes compliance with GDPR, Creative Commons licensing, and persistent identifiers, satisfying both EU legal obligations and Article 23 (Upholding strong data protection regardless of local standards). DIGITAfrica also embraces the FACT principles (Fairness, Accountability, Confidentiality, Transparency), embedding them into its governance model and data infrastructure. These principles complement the TRUST Code's guidance on ethical research conduct, reinforcing transparency and trust among all stakeholders.

6.3 Personal data handling

In virtue of its nature, DIGITAfrica main outcomes are non-personal data, in the form of blueprints for research-infrastructure, and capacity building reports, among others. Therefore, activities that handle personal data are only auxiliary and limited to supporting the Work Packages, such as through dissemination, events and contacting stakeholders.

The personal data collected may include names, institutional affiliations, email addresses, professional roles, and country of residence. No sensitive personal data or special categories of data were collected or aim to be collected in this project.

As stated in the DoA and the Grant Agreement, personal data processing activities comply with the GDPR (Regulation (EU) 2016/679), as well as applicable national data protection laws in the countries where data collection occurs, including in African partner countries. In accordance with these regulations:

- Lawfulness and Fairness, focusing on the identification of a valid legal basis, as per Art. 6 (1) and 9 GDPR, and the processing of data in a way that is not unjustifiably detrimental, unlawfully discriminatory, unexpected or misleading to the data subject;

- Informed Consent: Data subjects are fully informed about the nature and purpose of the data processing activities prior to providing consent. Consent is collected in a clear and accessible manner and may be withdrawn at any time without consequence.
- Data Minimization: Only personal data strictly necessary for the intended purposes is collected and stored.
- Storage and Access: Personal data is stored securely on GDPR-compliant servers and is accessible only to authorized members of the project team who require access for legitimate project purposes.
- Data Retention: Personal data will be retained only for as long as necessary to fulfill the purposes for which it was collected, or until consent is withdrawn, after which it will be securely deleted.
- Third-Party Access: No personal data will be shared with third parties outside the project consortium without the explicit consent of the data subject, unless required by law.
- Cross-Border Data Transfers: Personal data collected in the EU is not expected to be transferred to Non-EU countries. Data from African partners shared with European partners is only stored in the Project's Drive which is GDPR Compliant.
- Data Protection Officer: The partners agreed that the representative of Mandat International is appointed as the Data Protection Officer.

In addition to this, the partners declared the following elements concerning their adoption and handling of personal data. The Table 8 summarizes their declarations:

Partner Name	Purpose of Personal Data Collection (Categories)	Key Data Processing Activities	Further Processing?	Information Rights Disclosure	Consent Mechanism	Ethical Approval
BSC	To engage stakeholders in microcredential co-creation and inform AI/digital science training design. Data: names, roles, affiliations, emails.	Distributing surveys, stakeholder analysis, communication, anonymization of data.	No	Via a data privacy notice in the initial survey.	Informed digital consent collected via survey form.	Not required for their activities.
CNR	For dissemination and communicati	States they do not perform processing activities on	No	Not required	Not detailed.	Not required.

	on via mailing lists.	personal data in mailing lists.				
INRIA	Does not collect personal data	Blueprint	No	SLICES-RI policy	Per Slices-RI	Not required.
MI	Dissemination and communication activities. Website and organization of events: names, roles, affiliations, emails.	Communication purposes and limited to outreach and events	No	Yes, Privacy Policy in the website	Yes, Privacy Policy and Consent in the Website	Not required
STR	To support research and training activities	Specifically, data will be used to train AI models for decision-support systems, evaluate the impact of digital technologies, and deliver personalized learning or services within micro-credential programs.	No	Information sheets and consent forms	Consent will be obtained through digital forms embedded in university-managed LMS platforms or physically signed consent documents. All consents will be recorded and stored in secure, access-controlled systems, along with metadata (e.g., date, scope, withdrawal rights	Ethical approval will be obtained from the university's Institutional Research Ethics Committee prior to any data collection involving human subjects. All research activities involving personal data or human interaction complied with ethical research standards and legal requirements.

SU	Facilitate the logistics of the event organisation, daily project management.	Communications and Project Management	No	The project partners can request at any time request to be arise from the mailing list. Regarding the participants to the project events, a contact email is available and they can to be removed from the project database.	Data is directly collected from partners	Not required.
TUB	Does not collect personal data	Not applicable	Not applicable	Not applicable	Not applicable	Not applicable
UCAD	Not applicable	Not applicable	Not applicable	Not applicable	Not applicable	Not applicable
UCT	To facilitate stakeholder engagement, conduct surveys, coordinate pilot participants, and disseminate updates. Data: names, emails, institutional affiliations.	Collecting survey responses, analysing feedback, managing contact lists, documenting pilot testing.	No.	Privacy statements appended to survey tools/forms; UCT's privacy policy applies.	Informed consent via online forms (logged with timestamp).	Ethical approval sought from UCT's Research Ethics Committee for human participant activities, when applicable.
UMA	To conduct a survey on IT research infrastructure and identify needs/gaps.	Primarily statistics.	No	Explanations provided within the survey and accompan	Collected "By email".	Not necessary as they deal with public

	Data: names and email addresses of heads of research structures.			ying email.		email addresses.
UN	To manage and administer project workshops.	Collecting contact information (name, email, phone, organisations) for communicatio n, and recording attendance.	No.	A detailed privacy notice will be provided before data collection.	Explicit, informed consent by checking a box on a form.	Not specified.
UTH	For event organisation and disseminatio n activities. Data: name, email, affiliation.	Organisation of events, participant lists, questionnaire s.	No	"Through privacy notice".	Simple Consent Mechanism via online form.	Not required.
UVA	Does not collect personal data directly	Publicly available data only; minor exposure to partner-collected personal data possible (name, affiliation, email)	No further processing	Not required	Not required	Not required

Table 8 – Summary of data processing activities related to personal data

6.4 Technical and Organizational Measures

As required by the DoA and in line with Article 32(1) of the General Data Protection Regulation (GDPR), the DIGITAfrica consortium is committed to implementing appropriate technical and organisational measures (TOMs) to ensure a level of security appropriate to the risks associated with the processing of personal data. These measures take into account the state of the art, the costs of implementation, and the nature, scope, context, and purposes of the data processing activities carried out within the project.

Specifically, the following TOMs are considered appropriate and are being adopted by the DIGITAfrica consortium:

- Pseudonymisation and encryption of personal data, where applicable, to reduce the risk of identification in case of unauthorized access;
- Measures to ensure the ongoing confidentiality, integrity, availability, and resilience of data processing systems and services used in project activities;
- Procedures to restore the availability and access to personal data in a timely manner in the event of a physical or technical incident;
- Regular testing, assessment, and evaluation of the effectiveness of technical and organisational safeguards to ensure they remain fit for purpose throughout the lifecycle of the project.

These measures are implemented by each partner on their respective servers and applications, and by the consortium through the project's drive, managed by Sorbonne University.

7 Intellectual Property Rights

The DIGITAfrica project explicitly recognises that the management of knowledge and Intellectual Property Rights (IPR) are fundamental for smooth collaboration among consortium members and for the successful exploitation and sustainability of project outcomes both during and after the project. The framework for managing knowledge and IPR issues are integrated within the Consortium Agreement (GA), aligning with the policies and context for EC funded projects under Horizon Europe, and will be further detailed in a MoU in the last stages of project.

A core principle guiding the project is the intention to produce open source components that will be made publicly available to reach a broad audience. While promoting open access to research outputs such as publications, data, software, models, algorithms, and workflows through deposition in trusted repositories, the project intends to support these principles while retaining companies'/organisations' IPR by implementing the principle "as open as possible and as close as necessary".

Furthermore, the management of research data produced by the future RI will comply with the FAIR data principles, but importantly, this will be done while respecting the conditions outlined in the Consortium Agreement regarding IPR, access rights, and confidentiality. Contractual and legal management aspects, including IPR if applicable, are a significant focus within Work Package 7 (Management), specifically covered under Task 7.3, which also notes its relevance concerning the Memorandum of Understanding (MoU) defined in Task 5.4.

The Table 9 presents a high-level overview of the partner-provided inputs to this section.

Partner Name	Foreground IPR Generation Planned?	Ownership & Rights	Background Licences	Foreground Licences	Access & Dissemination of Foreground
BSC	No patents/trademarks currently foreseen, but may be revisited based on WP4 outcomes and digital tools.	Jointly owned by contributing partners, governed by consortium agreement.	Open-source components (MIT, Apache 2.0, GNU GPL) with attribution/obligations.	CC BY-NC 4.0 for content/data sets; MIT or Apache 2.0 for code-based outputs.	Publicly available via Zenodo, GitHub, or project platform, governed by open licences.
CNR	Not specified.	Not applicable	Not applicable	Not applicable	Not applicable
INRIA	Not foreseen	Not applicable	Not applicable	Not applicable	Not applicable
MI	No	Not applicable	Not applicable	Not applicable	Not applicable
STR	Copyrights such as custom-designed micro-credential course, interactive modules, software scripts and code, Know-how, trade secrets, potential trademarks	WP4 Lead, Formal agreements, MoU	The project integrates pre-existing open-source tools and frameworks, including TensorFlow (Apache 2.0), Scikit-learn (BSD 3-Clause), and various Python/R libraries for AI and data analytics. These background assets require proper attribution and inclusion of license texts, but permit commercial and academic use. University LMS platforms and previous eLearning modules may also be reused under Creative Commons licenses	The new code and digital content generated (e.g., AI models, micro-credential content, analytics tools) will be licensed under permissive open-source or Creative Commons licenses. For software/code, MIT or Apache 2.0 licenses will be applied. For course content and documentation, CC BY 4.0 will be used to encourage reuse with attribution.	All non-commercial, educational, and non-sensitive outputs will be made publicly available. Code and tools will be shared via repositories like GitHub or institutional GitLab. Educational content will be disseminated through the university's eLearning portal and public repositories such as Zenodo. Outputs will be released gradually upon project milestones or conclusion, under open licenses allowing reuse with attribution.

SU	No	Not applicable	Not applicable	Not applicable	Not applicable
TUB	Know- how and copyrights (technical specifications, blueprints, scientific papers).	TUB will retain ownership	The project incorporated various open-source components, all governed by various open-source licenses (MIT, Apache 2.0, BSD, GPL2, AGPL3 etc).	To be defined.	To be defined.
UCAD	Not specified.	Not applicable	Not applicable	Not applicable	Not applicable
UCT	Yes: Technical specifications, architectural blueprints, software scripts/tools, research outputs, training materials.	UCT retains ownership of its foreground IPR; joint ownership negotiated for co-developed outputs.	May incorporate open licenses (MIT, Apache 2.0, Creative Commons) with attribution.	CC BY 4.0, CC BY-NC 4.0, or MIT License.	Publicly available via UCT repositories, Zenodo, GitHub/GitLab, and project website after internal review.
UMA	Not specified.	Not applicable	Not applicable	Not applicable	Not applicable
UN	Not specified.	Not applicable	Not applicable	Not applicable	Not applicable
UTH	Not specified.	Jointly owned by contributing partners as per CA.	As described in the CA.	As described in the CA.	As described in the CA.
UVA	No	Not applicable	Not applicable	Not applicable	Not applicable

Table 9 - Foreseen Intellectual Property activities

8 Artificial Intelligence

Overall, DIGITAfrica acknowledges AI as a critical priority sector in ICT and a fundamental driver for Africa's socio-economic development. The project aims to harness AI's power and advance principles of FAIR (Findable, Accessible, Interoperable, Reusable) and FACT (Fairness, Accountability, Confidentiality, and Transparency) across Africa. While AI activities are not extensively developed during the project's direct implementation, they are a key focus for the future pan-African Digital Research Infrastructure (RI) blueprint. Potential risks like algorithmic bias, transparency, explainability, privacy, and data security have been

identified and will be closely monitored. The project will include AI topics in the design of the future RI and strictly monitor compliance with EU legislation, particularly the EU AI Act.

The Table 10 presents a high-level overview of the partner-provided inputs to this section.

Partner Name	AI Purpose / Application in Project	AI Tools / Algorithms Mentioned	Types of Data Processed by AI	Ethical & Compliance Measures	AI Activity During Project (or Future RI)
BSC	Development of microcredentials focused on AI and digital science training.	Not explicitly detailed for their direct use.	Not explicitly stated for AI-specific processing, but general data includes survey responses (names, emails, affiliations). Data will be anonymised.	Adherence to GDPR, local regulations, and EU Ethics Guidelines for Trustworthy AI. Transparent documentation, informed consent, and non-discriminatory outcomes.	AI activities for training design and potential digital tools are foreseen during project implementation.
CNR	Reference to SoBigData.eu's catalogue of AI methods and tools for data mining tasks.	SoBigData.eu offers a set of AI methods and tools.	Detailed description of input and output data formats are publicly available with the tools.	Not specified for direct CNR AI activities within the project, as they state they don't perform processing activities on personal data for AI.	CNR's role is to link to existing AI tools from SoBigData.eu; direct AI processing activities by CNR within the project are not stated.
INRIA	None	Does not apply	Does not apply	Does not apply	Does not apply
MI	No development of AI foreseen	Does not apply	Does not apply	Does not apply	Does not apply
STR	Yes	Will use tools and libraries such as TensorFlow, Keras, and Scikit-learn to implement supervised learning (e.g., decision trees, neural networks), unsupervised learning (e.g.,	Educational platform usage log and feedback, public datasets for training purposes	AI applications will follow ethical principles, with anonymized data, prior ethical approval, and compliance with GDPR and	Will employ AI technologies across multiple project streams, particularly in the areas of data science and analytics, AI for social impact, smart

		clustering), and predictive analytics models.		Kenya's Data Protection Act. Regular audits will address bias and unintended impacts.	city applications, and healthcare data management
SU	No development of AI foreseen	Does not apply	Does not apply	Does not apply	Does not apply
TUB	Not planned	Does not apply	Does not apply	Does not apply	Does not apply
UCAD	Does not apply	Does not apply	Does not apply	Does not apply	Does not apply
UCT	Needs, Gaps, & Environmental Assessment: AI-based modelling for energy efficiency, carbon impact, and sustainability scenarios.2. Requirement Analysis & Clustering: NLP for analysing and categorising qualitative stakeholder input.	Python-based libraries (e.g., Scikit-learn, NLTK), open-source AI frameworks (e.g., TensorFlow/Keras), and structured pipelines for text mining.	Structured survey data, qualitative feedback, and simulation outputs from sustainability models.	Compliance with EU AI Act guidance (where applicable) and UCT's POPIA-aligned ethical standards. Data will be anonymised when applicable. Use restricted to research and non-commercial purposes unless further consent obtained.	AI methodologies will be explored and applied within the project's activities.
UMA	Does not apply	Does not apply	Does not apply	Does not apply	Does not apply
UN	Use of AI for proofreading texts.	Not specified.	No personal data will be processed by AI.	Not specified beyond stating no personal data will be processed.	AI is used for proofreading texts during the project.
UTH	Development of microcredentials focused on AI and digital sciences.	Not explicitly detailed.	Not explicitly stated for AI-specific processing.	All EU directives will be followed.	AI activities for training design are foreseen during project implementation.

UVA	UvA will not develop AI tools in this project. It will use LLM-based tools (e.g., ChatGPT) only to accelerate the search and summarization of public documents. Any use of AI in generating reports will be transparently declared.	Publicly available Generative AI and LLM-based tools	Not specified	Not specified.	Not specified.

Table 10 - Adoption of Artificial Intelligence

9 Conclusions

This first iteration of DIGITAfrica's Data Management Plan showcases the activities in place and future by all project partners towards the definition, integration and implementation of a mature data governance strategy. Guided by living feedback from partners, with the Data Management Plan itself being a "living document" updated throughout the project, and informed by the evolving requirements of the pilots, technical activities and regulatory and ethical guidance, DIGITAfrica prioritized compliance while embracing transparency, scientific integrity and alignment with best practices and international standards.

The consortium adopted the FAIR data principles (Findable, Accessible, Interoperable, and Reusable) as a framework to ease the joint implementation of coordinated data practices with the long-term research and innovation goals in mind. By balancing openness and transparency requirements with a close monitoring of privacy requirements, the project has successfully generated impactful results.

As DIGITAfrica starts its blueprint phase, this deliverable presents readers and consortium partners alike with the foundations for what will be a sustainable data management approach down the road.

10 References

TRUST (2018) The TRUST Code – A Global Code of Conduct for Equitable Research Partnerships, DOI: https://doi.org/10.48508/GCC/2018.05

11 Annex I – Data Questionnaire sent to partners

Instructions

This questionnaire aims at collecting the initial information on data management, ethics and IPR of partners in the present research project.

The current survey needs to be completed **by all partners.**

Please fill in the relevant (contact) information for your organization on the first page of this questionnaire.

Should you not process any personal data in the framework of your involvement in the project, you should at least complete the sections on (FAIR) Data Management, IPR management and AI use.

Please return the completed survey to us as soon as possible.

Partner Organization

Name:
Address:
Country:
Website:
Privacy policy webpage:.

Contact Person

Name:
Email:
Phone Number:

Data Protection Officer

Name:
Email:
Phone Number:

DATA PROCESSING ACTIVITIES

1. **Indicate what categories of data you will collect or process in the context of the project:**

   - ☐ **non personal data** (i.e. environmental data, analytical reports, design specifications, architecture, etc). *Note: All aspects are related to Research Data Management and FAIR compliance*.
   - ☐ **personal data** (any information relating to identified or identifiable individuals, including for instance email or IP addresses)
   - ☐ **special categories of data** (personal data revealing sensitive information such as sexual orientation, racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade union membership, as well as any health, genetic or biometric data related to the data subjects)

2. **Describe the categories of personal data you will be collecting and/or processing: How did you collect these personal data?**

   - ☐ directly from data subjects who belong to your research team
   - ☐ directly from data subjects outside your research team (i.e. early adopters, beta testers, etc.)
   - ☐ indirectly through partners of the project
   - ☐ indirectly through other organizations external to the project
   - ☐ Other
   - ☐ N/A (*you can skip the last section of this document*)

DATA MANAGEMENT

3. **Please list all datasets which your organization will use or has obtained in the context of the DIGITAfrica project** (Feel free to duplicate this table if multiple datasets will or have been used)?

	Please provide your answers in this column:
Dataset(s) name	*What is the name of the used dataset(s)?*

FAIR DATA

Dataset(s) description	*Please provide a short description of the dataset(s).*
Personal Data	*Does the dataset include personal data? If yes, please specify the type of personal data.*
Purpose	*What is the purpose for which you use/ process the dataset(s)?*
Data format	*What format(s) are your dataset(s)?*
Data Storage	*Where will you store the dataset(s)?*
Main Data Source	*What is the main source of the dataset(s)?*
Data Ownership	*Who owns the dataset(s)?*
Country of Origin	*Where does the dataset come from?*
Restrictions on the use	*Are there any restrictions for the use of the datasets?*
Access	*Who has access to the datasets? Please include other work packages which will also access the datasets.*
Retention Period	*How long will you keep the datasets?*
Licence	*Under which licence did you obtain access to the datasets?*
WP and task	*For which work package and which task do you need to use the datasets?*
Additional Comments	*Please add here any additional comments.*

1. **Did you or will you be taking measures in order to comply with the FAIR data principles (making data Findable, Accessible, Interoperable and Reusable)?** *If so, kindly provide additional information on how each of these principles are being met:*

	Please provide your answers in this column:
Findable	*Please explain in detail and with as much technical information as possible how you ensure findability of your research data.*
	What types of data will you collect, generate, or reuse (e.g., datasets, images, software)?
	What metadata standards will you use to ensure discoverability (e.g., Dublin Core, ISO 19115)?
	What persistent identifiers will you use for your datasets (e.g., DOIs, PURLs)?
	How and where will your metadata be registered or indexed?
Accessible	*Please expain in detail and with as much technical information as possible how you ensure accessibility of your research data.*
	When and how will the data be made publicly available (e.g., embargo period, open access)?

FAIR DATA

Dataset(s) description	*Please provide a short description of the dataset(s).*
Personal Data	*Does the dataset include personal data? If yes, please specify the type of personal data.*
Purpose	*What is the purpose for which you use/ process the dataset(s)?*
Data format	*What format(s) are your dataset(s)?*
Data Storage	*Where will you store the dataset(s)?*
Main Data Source	*What is the main source of the dataset(s)?*
Data Ownership	*Who owns the dataset(s)?*
Country of Origin	*Where does the dataset come from?*
Restrictions on the use	*Are there any restrictions for the use of the datasets?*
Access	*Who has access to the datasets? Please include other work packages which will also access the datasets.*
Retention Period	*How long will you keep the datasets?*
Licence	*Under which licence did you obtain access to the datasets?*
WP and task	*For which work package and which task do you need to use the datasets?*
Additional Comments	*Please add here any additional comments.*

1. **Did you or will you be taking measures in order to comply with the FAIR data principles (making data Findable, Accessible, Interoperable and Reusable)?** *If so, kindly provide additional information on how each of these principles are being met:*

	Please provide your answers in this column:
Findable	*Please explain in detail and with as much technical information as possible how you ensure findability of your research data.*
	What types of data will you collect, generate, or reuse (e.g., datasets, images, software)?
	What metadata standards will you use to ensure discoverability (e.g., Dublin Core, ISO 19115)?
	What persistent identifiers will you use for your datasets (e.g., DOIs, PURLs)?
	How and where will your metadata be registered or indexed?
Accessible	*Please expain in detail and with as much technical information as possible how you ensure accessibility of your research data.*
	When and how will the data be made publicly available (e.g., embargo period, open access)?
	Which repository will you use for data sharing and long-term preservation (e.g., Zenodo, institutional repository)?

	Please provide your answers in this column:
Ownership and Rights	*Who owns the overall IPR for the foreground code/solutions, and how will rights be shared or managed among project partners?*
Background Licences	*Please state any pre-existing code/solutions that were incorporated into the project-generated code/solutions. What licences govern these background assets, are there restrictions or obligations? (e.g. attribution, non-commercial use)*
Foreground Licences	*Please state those licences that have been applied to the new code/solutions (foreground) that were generated during the project.*
Licence Compatibility	*Are there conflicts or compatibility issues between the background and foreground licences? How should they be addresses?*
Access and Dissemination	*Will the foreground code/solutions be made publicly available after the project conclusion? If yes, under what conditions? (e.g. repositories, timing, usage restrictions)*

AI USE PRACTICES

	Please provide your answers in this column:
AI use practices	*Did your organization use or plan to use any Artificial Intelligence (AI) technologies or methodologies as part of the project? If so, please describe the specific AI tools or algorithms employed, the purpose and application of AI in your research activities, the types of data processed by these AI systems, and measures taken to ensure ethical use and compliance with relevant data protection regulations.*

ETHICS AND PERSONAL DATA PROTECTION

	Please provide your answers in this column:
Purpose of Personal Data Collection	*For what purpose(s) did you collect the aforementioned personal data?*
Data Processing Activities	*List the main data processing activities your organization has or will perform in the context of the project*
Further Processing Purposes	*Did you or will you process the generated data for any further purposes than the ones it was originally collected for? Please indicate YES or NO*
	If you answered YES, then please describe the purpose of this additional processing.

Information Rights	*How did you inform the individuals (the data subjects) about the purpose of the data processing of their personal data in the project?*
Consent	*How did you plan to collect and document the consent of the data subjects whose personal data will be processed by you?*
Data Storage	*How and where did you or will you store the data?*
Data Retention	*For how long did you or will you keep the data?*
Ethical Approval	*Did you or will you obtain an ethical approval as part of your project-related activities? Please elaborate.*

Funded by the European Union

DIGITAfrica project has received funding from the EU Horizon Europe research and innovation Programme and Switzerland under Grant Agreement No. 101187966. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Executive Agency (REA). Neither the European Union nor the granting authority can be held responsible for them.