Operationalizing FAIR and CARE data stewardship principles in WASH systems

openwashdata.

Emmanuel Mhango
BASEflow Malawi, Global Health Engineering, ETH Zurich

2026-10-27

Note: This post was originally published by Emmanuel Mhango on LinkedIn on 27 October 2026 (https://www.linkedin.com/pulse/operationalizing-fair-care-data-stewardship-wash-systems-mhango-go7jf/).

Moving beyond data collection toward responsible, reusable, and community-centered water, sanitation, and hygiene data.

Introduction

The Water, Sanitation, and Hygiene (WASH) sector generates an enormous amount of data. From borehole inventories and groundwater monitoring records to water quality assessments, sanitation surveys, infrastructure functionality reports, and household access indicators, these datasets are essential for planning, monitoring, resource allocation, and decision-making.

Yet, collecting data is only one part of the equation. The real value of WASH data depends on whether it can be discovered, accessed appropriately, understood, reused, and applied to improve services and outcomes.

Across many WASH programs, datasets remain scattered across institutions, stored in incompatible formats, poorly documented, or difficult to access beyond the projects that generated them. In other cases, data may be technically available but lack the contextual information needed for meaningful interpretation. Community-generated data may also be shared without sufficient consideration of the people and communities represented.

These challenges raise an important question: **How can we make WASH data more interoperable and reusable while ensuring that its management remains ethical, inclusive, and accountable to communities?**

Two complementary frameworks offer a useful starting point: the FAIR principles and the CARE principles.

Understanding FAIR and CARE

The FAIR principles, introduced by Wilkinson and colleagues in 2016, provide guidance for making data Findable, Accessible, Interoperable, and Reusable.

- **Findable:** Data should have persistent identifiers, descriptive metadata, and searchable records so that users can discover relevant datasets.

- **Accessible:** Data should be retrievable through clearly defined protocols, with transparent access conditions. Accessibility does not mean that all data must be publicly available.
- **Interoperable:** Data should use shared vocabularies, consistent formats, and compatible standards that enable integration across systems.
- **Reusable:** Data should be accompanied by sufficient documentation, provenance, licensing, and quality information to support responsible reuse.

The CARE principles, developed by the Global Indigenous Data Alliance, complement this technical orientation by emphasizing Collective Benefit, Authority to Control, Responsibility, and Ethics.

- **Collective Benefit:** Data governance should support tangible benefits for the communities and people represented.
- **Authority to Control:** Indigenous Peoples and other relevant rights holders should have appropriate authority over decisions concerning their data, consistent with their rights and governance arrangements.
- **Responsibility:** Data users and custodians should build relationships, strengthen capacity, and use data in ways that support equitable outcomes.
- **Ethics:** Data practices should respect people, communities, their rights, and the potential consequences of data use.

FAIR helps us ask whether data can be found, accessed, combined, and reused. CARE challenges us to ask who benefits from that reuse, who has the authority to make decisions, and whether the process respects the people represented.

In WASH systems, both questions matter.

A technically excellent dataset can still be used irresponsibly. Equally, a well-governed dataset may have limited value if it cannot be discovered, interpreted, or integrated with other information systems.

The opportunity lies in operationalizing both frameworks together.

1. Making WASH data findable and meaningful

Consider a dataset containing groundwater monitoring observations from several districts. It includes water levels, electrical conductivity, temperature, observation dates, and monitoring well locations.

If the dataset is uploaded without a descriptive title, metadata, geographical coverage, collection methodology, or information about the monitoring equipment, its practical value is limited.

Applying FAIR begins with establishing a structured data inventory and publishing descriptive metadata. Each dataset should have a clear title, abstract, responsible institution, geographical coverage, temporal coverage, collection methods, update frequency, data dictionary, and access conditions.

Where possible, datasets should be assigned persistent identifiers and registered in searchable repositories or catalogues.

However, findability should not be confused with unrestricted exposure. Exact locations of vulnerable water infrastructure, sensitive community information, or records that could expose individuals to harm may require restricted discovery or carefully controlled access.

The objective is to make datasets discoverable by the appropriate users while maintaining proportionate safeguards.

2. Making data accessible without making everything public

One common misunderstanding about FAIR is that accessibility requires all datasets to be openly available. In practice, FAIR supports access through clearly defined and documented conditions, including authentication, authorization, and controlled access where necessary.

For WASH institutions, this means establishing practical data access procedures. These may include public access for approved aggregate indicators, institutional access for operational records, and restricted access for sensitive household-level information.

Data-sharing agreements should specify who may access the data, for what purpose, under which conditions, and with what responsibilities. For example, a district water office might make aggregated borehole functionality statistics publicly available while restricting access to detailed infrastructure security information or identifiable household survey records.

A well-designed data governance framework should document these distinctions rather than leaving access decisions to individual discretion. The guiding principle is straightforward: data should be as open as appropriate, and as restricted as necessary.

3. Improving interoperability across WASH information systems

WASH data is often generated through different platforms, reporting structures, and technical environments. Water infrastructure records may be maintained in asset management systems, groundwater observations in monitoring databases, sanitation indicators in survey platforms, and program performance data in monitoring and evaluation systems.

Even when these datasets describe the same geographical areas or infrastructure, differences in identifiers, administrative boundaries, terminology, units of measurement, and data structures can make integration difficult.

Operationalizing FAIR requires deliberate attention to interoperability. Practical interventions include:

- Establishing common identifiers for water points, monitoring wells, and other infrastructure assets.
- Standardizing geographical references, administrative codes, dates, and measurement units.
- Maintaining shared data dictionaries and controlled vocabulary.
- Using documented APIs and machine-readable formats for data exchange.
- Recording relationships between infrastructure assets, monitoring observations, maintenance activities, and water quality results.
- Adopting relevant national and international standards where appropriate.

For example, groundwater monitoring observations should be linked to a stable identifier for the monitoring well. This makes it possible to combine water-level measurements with infrastructure characteristics, historical maintenance records, and relevant environmental information.

Interoperability is not simply a software integration problem. It is also an institutional coordination problem. Organizations must agree on definitions, ownership, responsibilities, and the procedures through which data is updated and corrected.

4. Embedding CARE into WASH data governance

Although FAIR provides a strong foundation for data management, it does not, by itself, fully address power, representation, or the distribution of benefits arising from data use.

This is where CARE becomes particularly important.

WASH datasets frequently contain information about communities, household water access, sanitation practices, local water sources, and service inequalities. These records can inform important public decisions, but their use should be guided by the interests, rights, and circumstances of the people represented.

Operationalizing CARE requires institutions to go beyond obtaining data and consider the relationships surrounding its collection, management, sharing, and reuse.

Collective Benefit means designing data initiatives around meaningful improvements in water and sanitation services. Data-sharing arrangements should support better infrastructure maintenance, more equitable resource allocation, improved service reliability, and stronger local decision-making.

Authority to Control requires attention to the rights and governance arrangements of the people and communities concerned. Indigenous data sovereignty is particularly relevant where Indigenous Peoples' data is involved. More broadly, WASH institutions should establish appropriate mechanisms for community participation, representation, and decision-making without assuming that a single community representative can speak for everyone.

Responsibility means investing in local data management capacity, providing appropriate access to useful information, acknowledging data contributors, and building partnerships that allow institutions and communities to benefit from the data they help generate.

Ethics requires identifying and mitigating risks such as re-identification, stigmatization, exclusion, misuse, and the disclosure of sensitive information. Data collection and sharing should be proportionate to legitimate purposes and accompanied by appropriate safeguards.

For WASH programs, this may involve consulting communities about the intended uses of household-level data, explaining how information will be shared, establishing complaint and correction mechanisms, and reporting back on how findings inform service delivery.

Community participation should not be treated as a one-time consultation. It should be an ongoing part of data governance.

5. A practical example: from WASH data publication to borehole rehabilitation

The value of data stewardship becomes clearer when we look at what happens after data is collected.

In my work in the WASH sector in Malawi, I have experienced this firsthand. We had collected data on water infrastructure status in Mwanza District and other areas. This dataset covering water infrastructure status and related information was published through an open data platform.

Before publication, this information was stored on computers at the BASEflow office, limiting its visibility and potential for reuse beyond the immediate project team.

Barely 5 days after publication, we received an email from UNICEF indicating that the organization had discovered, appreciated, and downloaded our data. UNICEF subsequently used the information to support borehole rehabilitation activities in the areas covered by the datasets.

Following the intervention, water point functionality was restored in the areas covered by the datasets.

This experience illustrates an important principle: **data creates greater value when the people who need it can find it and act on it.**

From a FAIR perspective, publishing the datasets transformed information that was previously difficult for external organizations to discover into a resource that could be accessed and reused for operational decision-making. The experience also demonstrated how open data can help development partners identify infrastructure needs and direct interventions where they are needed.

The CARE perspective adds another important consideration. The ultimate value of WASH data should not be measured only by downloads, publications, or institutional visibility. It should also be assessed by whether its use contributes to meaningful benefits for the communities that depend on those water points.

This example also highlights an opportunity for stronger data stewardship. By documenting datasets consistently, maintaining reliable water point identifiers, recording data collection dates, and tracking subsequent uses and outcomes, WASH institutions can make it easier to establish how data contributes to service improvements.

Data that remains on an office computer has limited reach. Data that is well documented, appropriately shared, and actively reused can help turn evidence into action.

6. Translating principles into operational practice

The greatest challenge is not understanding FAIR and CARE conceptually. It is translating them into everyday institutional processes.

A practical implementation approach can begin with five steps.

Step 1: Assess existing data assets and governance arrangements.

Develop an inventory of datasets, documenting their custodians, purposes, formats, quality, sensitivity, access conditions, and update schedules. Identify gaps in documentation, interoperability, and accountability.

Step 2: Establish institutional data stewardship responsibilities.

Define the roles of data owners, custodians, stewards, system administrators, and authorized users. Establish procedures for data validation, metadata management, version control, access requests, retention, and incident response.

Step 3: Develop metadata and interoperability standards.

Create common metadata templates, data dictionaries, identifier conventions, and exchange formats. Align these with relevant government standards and sector requirements rather than introducing parallel systems without a clear need.

Step 4: Integrate ethical and community-centered governance.

Classify data according to sensitivity and risk. Establish proportionate access controls, privacy safeguards, consultation procedures, and mechanisms for addressing community concerns. Clarify how collective interests and applicable rights will inform decisions about data use.

Step 5: Monitor implementation and improve continuously.

Measure whether datasets are adequately documented, discoverable, interoperable, and reusable. Also assess whether access decisions are transparent, communities are meaningfully engaged, local capacity is improving, and data use contributes to better services.

This approach makes FAIR and CARE part of routine data management rather than principles that exist only in policy documents.

7. Measuring success: from compliance to impact

Successful implementation should not be measured solely by the number of datasets collected or published.

A more useful approach combines technical, governance, and service-delivery indicators.

Table 1: Dimensions and example indicators for assessing data governance and responsible data use

Dimension	Example indicators
FAIR implementation	Percentage of datasets with complete metadata, documented access conditions, data dictionaries, and provenance
Interoperability	Percentage of relevant datasets using agreed identifiers, vocabularies, and exchange formats
Data quality	Completeness, validity, consistency, timeliness, and documented correction rates
CARE implementation	Evidence of community participation, transparent governance decisions, and mechanisms for raising concerns
Capacity development	Number of institutions and staff supported to manage, interpret, and responsibly reuse data
Service-delivery impact	Improvements in maintenance planning, response times, infrastructure functionality, and the equitable allocation of resources

These indicators should be adapted to institutional mandates, available resources, and the maturity of existing information systems.

Importantly, the presence of a policy or consultation mechanism does not automatically demonstrate that CARE is being implemented effectively. Institutions must examine whether governance arrangements work in practice and whether the resulting decisions produce meaningful benefits.

Similarly, FAIR compliance does not automatically guarantee accurate data. Discoverable and interoperable datasets can still contain errors, biases, or incomplete observations. Data quality assurance remains essential.

8. The role of digital infrastructure and institutional capacity

Digital repositories, dashboards, APIs, geospatial platforms, and automated quality checks can make FAIR and CARE implementation more practical. However, technology alone cannot resolve weak governance or unclear institutional responsibilities.

A data portal may improve discovery, but it cannot determine whether a dataset should be shared without an appropriate governance framework. An API may simplify integration, but it cannot resolve conflicting definitions of infrastructure functionality. An advanced analytics platform may identify patterns, but it cannot independently determine whether a particular use is ethical or beneficial to the communities concerned.

For institutions working in resource-constrained settings, implementation should therefore be incremental and sustainable.

This includes investing in staff skills, documenting existing workflows, adopting maintainable technologies, strengthening institutional collaboration, and aligning data stewardship with established monitoring, evaluation, and reporting processes.

It also means recognizing that not every institution needs an entirely new platform. In many cases, better metadata, clearer access rules, standardized identifiers, and stronger stewardship practices can deliver substantial improvements using existing infrastructure.

Conclusion

The future of WASH data management depends on more than digitization. It requires institutions to treat data as a strategic resource that must be managed carefully, shared responsibly, and used to improve people's lives.

The FAIR principles provide a practical foundation for making WASH data findable, accessible under appropriate conditions, interoperable, and reusable. The CARE principles add an essential governance perspective by emphasizing collective benefit, authority to control, responsibility, and ethics.

Together, these frameworks encourage a more balanced approach to data stewardship: one that recognizes both the technical value of information and the rights, interests, and responsibilities associated with its use.

For WASH institutions, the immediate priority is to move from principle to practice. This means establishing data inventories, strengthening metadata and quality assurance, adopting common standards, defining stewardship responsibilities, and embedding ethical and community-centered decision-making into data-sharing arrangements.

Ultimately, the success of WASH data stewardship should not be judged only by how much data an institution collects or publishes. It should also be judged by whether that data can be trusted, responsibly reused, and translated into more reliable, equitable, and sustainable water and sanitation services.

The goal is not simply to make WASH data available. It is to make WASH data useful, trustworthy, and accountable to the people it is intended to serve.

Join the conversation

To WASH professionals, government institutions, researchers, and development partners:

How is your organization operationalizing the FAIR and CARE principles in its WASH information systems? What tools, governance practices, or lessons have helped you turn data into better services for communities?

I welcome your perspectives and experiences in the comments on the original LinkedIn article.

Selected references for further reading

1. Wilkinson, M. D., et al. (2016). *The FAIR Guiding Principles for scientific data management and stewardship*. Scientific Data, 3, 160018. https://doi.org/10.1038/sdata.2016.18
2. Carroll, S. R., et al. (2020). *The CARE Principles for Indigenous Data Governance*. Data Science Journal, 19, 43. https://doi.org/10.5334/dsj-2020-043
3. Global Indigenous Data Alliance (GIDA). *CARE Principles for Indigenous Data Governance*. https://www.gida-global.org/careprinciples
4. Research Data Alliance (RDA). *FAIR Data Maturity Model*. https://www.rd-alliance.org/

This article is based on my experience as a Data Steward at BASEflow Malawi through the openwashdata community with ETH Zurich.