I thought my job was about data. It turned out to be about people.

openwashdata.

Emmanuel Mhango
BASEflow Malawi, Global Health Engineering, ETH Zurich

2026-08-11

Note: This post was originally published by Emmanuel Mhango on LinkedIn on 11 August 2026 (https://www.linkedin.com/pulse/i-thought-my-job-data-turned-out-people-emmanuel-mhango-rszvf/).

When I joined BASEflow Malawi as a Data Steward, I thought I had a fairly clear picture of what my work would involve. I expected to organize datasets, improve data quality, and publish information that could support research and decision making. It sounded straightforward.

I was wrong.

One of the first questions I asked was surprisingly simple.

What data do we actually have?

I expected someone to point me to a folder or a database where everything was neatly organized. Instead, I found data spread across different systems and storage locations. Some groundwater monitoring data was stored in CSV files. Other information was managed in mWater. Different projects had developed their own ways of storing information over time. Some datasets were well known, while others had almost been forgotten. Even within the same platform, related datasets were not always connected. In some cases, there were duplicate records created through different workflows. Documentation was limited, and many datasets had very little information explaining where they came from, why they had been collected, or who was responsible for them.

That was the moment I realized my job was not simply about managing data. Before I could improve anything, I first had to understand the organization's data landscape.

The first thing I built was not a dashboard or a publication. It was a data inventory.

I wanted to know every dataset that BASEflow owned or managed. For each one, I recorded information such as who collected it, when it was collected, where it came from, its purpose, who funded the activity, and who was responsible for maintaining it. I then worked closely with colleagues to confirm ownership and fill in gaps where information was missing.

Looking back, this inventory became much more than a spreadsheet. It became a map of our organizational knowledge. For the first time, we had a clearer picture of the information we held and where it lived. It gave me direction, and it gave the organization a foundation on which we could build better data governance.

As I continued my work, I discovered that one of the biggest challenges was finding the right balance between openness and responsibility.

Part of my role was to publish datasets so they could be shared with researchers, government partners, and development organizations. However, not every dataset could simply be placed online. Some contained personal information that had to remain confidential. Others had been collected through partnerships, and some was owned by government.

There were moments when I found myself asking difficult questions. Who owns this dataset? Do we have permission to publish it? How do we protect people's privacy while still making valuable information available for public good?

Those experiences taught me that data stewardship is not only about technology. It is also about trust. Every decision requires careful consideration of ethics, ownership, and responsibility.

One lesson has stayed with me throughout this journey.

To me, data without metadata is orphaned data.

A dataset may contain thousands of records, but if nobody knows where it came from, who collected it, why it was created, or how it should be interpreted, much of its value is lost. Metadata gives data its identity. It transforms isolated files into knowledge that others can understand and use with confidence.

One of the moments I will always remember came after we had started publishing datasets.

I received communication that UNICEF Malawi had accessed some of the groundwater datasets we had published. The dataset contained information from borehole forensic assessments, including the condition of boreholes, their functionality, and their locations. By analyzing this information alongside district authorities, they were able to identify communities where water infrastructure required attention and target rehabilitation efforts more effectively.

That moment changed the way I thought about Open Science.

Publishing data is not simply about making information available on the internet. It is about creating opportunities for others to solve problems that none of us can solve alone. Data that sits on a hard drive helps very few people. Data that is responsibly shared can help governments plan better, researchers ask better questions, and development partners deliver more effective interventions.

Looking back over the past year, I am probably most proud of the Data Management Strategy we developed at BASEflow. The strategy has provided a clear direction for how the organization will manage its data over the coming years. It established priorities, defined key performance indicators, and created a roadmap that aligns data management with the organization's broader mission. More importantly, it has helped position data as a strategic asset rather than simply a project output.

Perhaps the biggest lesson I have learned has very little to do with data itself.

Successful data stewardship begins with people.

Technology can help organize information, but it cannot build collaboration. Data governance only works when people share a common understanding of why it matters. Every dataset represents the work of people who collected it, communities who contributed information, and organizations that invested resources to create it. Good stewardship means respecting all of those contributions while ensuring that the knowledge created can continue to benefit others.

As more organizations embrace digital technologies, artificial intelligence, and evidence based decision making, the importance of trustworthy and well managed data will only continue to grow. My hope is that more NGOs will see data stewardship not as an administrative function, but as an investment in learning, collaboration, and long term impact.

If there is one lesson I would leave with any organization beginning this journey, it is this.

Before you can share your data, you first need to know your data.

That simple step can become the foundation for better decisions, stronger partnerships, and greater impact for the communities we all aim to serve.