

Job advert: Data Quality Specialist at BASEflow, Malawi

2024-09-10

Download the [PDF version](#) of this job advert.

Job Title: Data Quality Specialist

Reporting to: Programmes Manager

Workstation: Blantyre (primarily), but open to deliverable-based remote working arrangement

Application due date: 1st October 2024

Note: Carefully review the submission requirements. Incomplete applications will not be considered.

Project background

The openwashdata community was established in 2023 with the vision for an active global community that applies FAIR principles to data generated in the greater water, sanitation, and hygiene (WASH) sector. The project is managed by the Global Health Engineering group at ETH Zurich and in its first 1.5 years developed a 10-week “data science for openwashdata” training programme and established a clear and efficient data publishing workflow for datasets and code that follows FAIR principles and highest standards for computational reproducibility and version control. This second phase of the project envisions a more hands-on approach by embedding a data steward within two partner organisations in South Africa and Malawi. These data stewards will collaborate with the partner organization to identify existing Open Research Data practices, formulate a preliminary data management strategy, pinpoint datasets ready for open publication and those requiring more careful scrutiny before publication, and deliver data science training to a community of organisations who are interested in improving their open data practice.

The openwashdata team will provide supervision and training to the data stewards for aspects of data management, including data management plans, data privacy, data ethics, and data publication. The training will cover aspects of file organization, file types, and different pathways for collaboration and task management.

The strategic partner for this follow-up project on data stewardship in Malawi is a social enterprise called BASEflow. Operating from Blantyre, this non-profit organization has a committed team of 15 individuals and is dedicated to enhancing water security in the country. They have demonstrated leadership in data collection and sharing within the WASH sector in Malawi, offering support to research institutions and government agencies in both data collection and management.

The overall outcome of this follow-up project is an increase in WASH related data that is published openly. The funded proposal and more information are available at this website: <https://openwashdata.org/pages/gallery/proposal-02/>

Job description

We seek a Data Quality Specialist who will take on the role of a data steward for a 1-year contract within BASEflow and will collaborate directly with the openwashdata team to deliver the data stewardship project in-country. The candidate will be responsible for the following:

Identifying Current Data Management Practices and Developing a Draft Data Management Strategy for BASEflow

- **Conduct an Audit of Current Data Management Practices:** Review how data is currently collected, stored, accessed, and utilized within BASEflow. Identify any gaps, inefficiencies, or potential risks.
- **Stakeholder Consultation:** Engage with key stakeholders (e.g., data users, IT staff, management) to understand their needs, challenges, and suggestions regarding data management.
- **Benchmarking Against Best Practices:** Research and compare BASEflow's data management practices with industry standards and best practices to identify areas for improvement.
- **Draft Data Management Strategy:** Develop a strategy document outlining recommended practices, tools, policies, and procedures for data management, including data governance, storage solutions, data sharing protocols, and data security measures.
- **Review and Feedback:** Present the draft strategy to stakeholders for feedback and make necessary revisions before finalizing the document.

Publishing at Least 10 Datasets of Two Different Types Available to BASEflow

- **Identify and Categorize Available Datasets:** List all datasets available to BASEflow, categorizing them into types (e.g., environmental data, community surveys, operational data).
- **Select Target Datasets for Publication:** Choose at least 10 datasets from two different categories for publication, prioritizing those that are most valuable or frequently requested.
- **Data Cleaning and Standardization:** Prepare the selected datasets by cleaning, organizing, and standardizing the data to ensure consistency and usability.
- **Metadata Documentation:** Create comprehensive metadata for each dataset, including descriptions, data sources, collection methods, and usage guidelines.
- **Publication on Accessible Platform:** Publish the datasets on an appropriate platform (e.g., data repository, cloud storage) that ensures easy access while maintaining data security.
- **Communication and Outreach:** Inform relevant stakeholders about the availability of the published datasets and provide guidance on how to access and utilize them.

Supporting Data Science Trainings with Varying Styles, Focus Topics, and Audiences

- **Needs Assessment and Audience Analysis:** Identify the specific data science skills gaps and training needs within BASEflow and its partners. Analyze the target audience for each training program, considering factors like experience level, job role, and learning preferences.
- **Curriculum Development for Diverse Topics:** Design tailored training curricula covering a range of data science topics (e.g., data visualization, statistical analysis, machine learning) suited to different skill levels and organizational needs.
- **Design Training Materials:** Create engaging training materials, including presentations, hands-on exercises, tutorials, and reference guides, adapted to various teaching styles (e.g., workshops, seminars, online courses).
- **Deliver Training Sessions:** Conduct the training sessions using a mix of delivery methods (e.g., in-person workshops, webinars, self-paced online modules) to cater to different audiences and learning styles.
- **Post-Training Evaluation and Support:** Gather feedback from participants to assess the effectiveness of the training. Provide follow-up support, such as Q&A sessions, refresher courses, or additional resources, to reinforce learning.
- **Continuous Improvement:** Use feedback and outcomes from the training sessions to continuously refine and improve future training programs.

Desired Qualifications and Experience

Education:

- A Master's degree or equivalent, approved qualification in Computer Science, Data Science, Engineering, Media and Communication, Statistics, Development Studies, or related

Experience:

- At least 3-5 years of experience using data science tools (e.g. Git, GitHub, R, Python, RStudio IDE, VS Code, etc.)
- Experience in developing curriculum, facilitating, coordinating, and conducting training sessions to audiences of varying diversities in Malawi.
- An understanding of innovative/creative teaching strategies that ensure inclusivity will have an added advantage.
- Experience managing and conducting interdisciplinary (research) projects in environmental and social sciences, engineering, and international development.

Other Essential Attributes:

- A willingness to challenge and change institutional practices that present barriers to different groups.
- Strong networking and relationship-building skills and an interest in building broad collaborations.
- A sense of humor is mandatory

Recruitment Dates

- 2024-10-01: due date for submission of application
- 2024-10-11: notification about passing first selection round (two selected candidates)
- 2024-10-14 & 2024-10-18: a week reserved for a personal and a technical interview
- 2024-10-23: final notification about selection
- to be decided: Start date of employment

Interview panel

- Muthi Nhlema, Team Leader, BASEflow Limited, Malawi
- Lars Schöbitz, Project Management, Global Health Engineering, ETH Zurich

Submission requirements

1. An online sample/portfolio of previous work showing the programming code used for a data analysis project (if no public portfolio exists, a script with programming code can be submitted with the application package).
2. An updated CV in your current format (no more than 3 pages including 3 references).
3. An essay (maximum 500 words) that argues in favor of the following statement: “The data we generate belongs to us and I don’t see the benefits of sharing it openly.”

Equal Opportunity Employer

BASEflow is an equal opportunity employer and values diversity. All qualified applicants will receive consideration for employment without regard to race, color, religion, sex, sexual orientation, gender identity, national origin, disability, or veteran status.