

Making ODK Aggregate A Greater Global Good

Submitted by Yaw Anokwa on January 19, 2018 - 12:25pm

Last revised by Web Producer on June 21, 2018 - 3:09pm.

Proposal Status: [Postponing for Future Calls for Proposals](#)

Executive summary

Open Data Kit (ODK) is an open-source and community-owned set of tools that replaces paper surveys with smartphones. The tools are primarily used by health organizations to collect data quickly, accurately, offline, and at scale. This effort focuses on the ODK 1 suite of tools (Collect, Aggregate, XLSForm, Build, JavaRosa) which are widely-deployed global goods that have been used to collect billions of data points. Example projects include:

- For governments working to end polio, access to accurate and timely information makes a world of difference. ODK is used in Jordan, Afghanistan, Pakistan, Somalia, and South Sudan as a key tool in mass polio vaccination campaign quality control. <https://www.youtube.com/embed/zROyvrvt-zk>
- PMA2020 uses ODK to collect a nationally representative sample of data from households and service delivery points in selected sentinel sites. The data is used to estimate health indicators on an annual basis in 11 pledging FP2020 countries. <https://pma2020.org/what-we-do>
- The Neglected Tropical Diseases Support Center has collected more than 50 million data points across 200 projects in 40 countries using ODK. The data has gone to inform The Global Trachoma Mapping Project, which is the world's largest single disease mapping initiative. <https://www.taskforce.org/case-study/global-trachoma-mapping-project-determining-prevalence-help-eliminate-trachoma-2020>
- ODK enabled 400 data collectors to submit more than 80,000 forms in a week in an effort to build a national health map. Combined with satellite imagery, the effort provided the most accurate health sector administrative boundaries to date for Cameroon. <https://forum.opendatakit.org/t/9273>
- Doctors without Borders collected vital mapping and village information to improve ebola response in the 2015 outbreak in Sierra Leone by mobilizing a local community with self-owned smartphones using ODK. <https://forum.opendatakit.org/t/11328>

ODK has been designed for novice users in challenging environments and its robustness in these environments has driven the platform's adoption and evolution. Additionally, the choice to build an active open-source community around ODK has allowed it to benefit from users, implementers, and developers.

Over the last year, the ODK project has transitioned from a single "owner" to community ownership and the project has experienced extraordinary growth during that time. We wish to build on that growth and address long-standing issues with ODK Aggregate, the widely-deployed ODK server. Aggregate is a key component of ODK that if improved would enable a more contributable codebase and greater integrations with the broader health ecosystem.

Consortium team

Nafundi is a software company started by the founders of ODK. Nafundi's leadership, Dr. Yaw Anokwa and Ms. H  l  ne Martin, facilitate much of the software development and community activities on the core ODK tools. Dr. Anokwa, with guidance from ODK's project management committee, will serve as the organizational point of contact. Ms. Martin, with guidance from ODK's technical steering committee, will serve as the technical point of contact.

Over the last year, the team at Nafundi has stepped forward to help the community improve ODK tools. In addition to shipping features requested by the community, we've invested deeply in community processes (launching a community forum, implementing code review, rigorous testing, launching a docs site, actively recruiting and mentoring contributors, sourcing funding for community members) that help ensure the long-term sustainability of the project.

These investments have dramatically grown usage (50% increase in installs of ODK Collect) of the tools and as well as the community (80%

increase in new members of the forum). Community members now participate in every part of ODK, from technical support to software development.

We believe that now is the time to build on this momentum and add functionality to ODK that has been asked for by the community but has not been addressed due to a lack of targeted and consistent funding.

We will continue to use the same open and collaborative process we've used over the past year. And to that end, our colleagues on the ODK 1 technical steering committee and moderators of our community will guide this effort and we continue to seek other partners.

For each activity, we will publish a tentative plan and a call for contributors (developers, testers, writers, designers, managers) on our community forum and social media. Nafundi will facilitate those contributors and serve as a backstop to ensure ongoing and consistent process.

Project description

With Digital Square's support, the consortium will focus our efforts on ODK Aggregate, the ODK server. We focusing on Aggregate because its needs are well-known. Further, improvements to Aggregate are vital to the long-term health of the ODK ecosystem.

Aggregate remains the only OpenRosa-compliant server that has hundreds of thousands of downloads, runs on widely available infrastructure, has been proven at scale in challenging environments, and is governed by an active, transparent, and multi-stakeholder community.

Aggregate is a complex code base that has been maintained by a single individual. We want to use this effort to reduce this risk by reducing the effort required to get started contributing to Aggregate. Reducing the effort will broaden the contributor base so a diverse coalition of individuals, organizations, ideas, and energy can help ensure Aggregate remains a useful tool.

To ensure the influx of contributors can contribute meaningfully, we'll focus on adding the important functionality that the ODK community and the broader health ecosystem have already expressed interest in via feature requests and bug reports. For this effort, we will take on the following activities:

- Add continuous integration to make Aggregate easier to contribute to
- Add basic usage analytics to better inform maintainers about usage patterns
- Add improved logging and error reporting to identify unseen areas for improvement
- Add update notifications to ensure users are using the most recent versions
- Add data consistency checks for large and complex collection campaigns
- Add form-grouping support to enables lightweight multi-tenancy on a single machine
- Add lightweight form updates and versioning to simplify iterative data collection
- Add more robust bulk data exports via web interface and API
- Add new streaming formats that enable easier integrations

The consortium we will use the same open, collaborative, and consensus-based process we've used over the last year to prioritize, design, and build the above functionality. We are confident that with the support of the ODK community and Digital Square, we can harness the renewed energy around ODK to make ODK Aggregate a greater global good.