DATA ZETU

Connecting Flows and Places

Making Data Useful to Hyperlocal Communities in Tanzania

A DATA ZETU RESEARCH REPORT

Connecting Flows and Places

Making Data Useful to Hyperlocal Communities in Tanzania

Prepared by
François van Schalkwyk, SBC4D &
Rosemary Silaa, Health Consultant
for Data Zetu

FUNDED BY

DATA COLLABORATIVES FOR LOCAL IMPACT

PARTNERS

FEEDBACK

We value any feedback, comments or corrections to this report.

Please email your feedback to habari@datazetu.org or complete the form on the Data Zetu website: https://datazetu.or.tz/connect/

ABOUT DATA ZETU

Data Zetu helps communities to make better, more evidence-based decisions to improve their lives. Through partnerships and collaborations with local communities, Data Zetu works with stakeholders to build skills and develop digital and offline tools that make information accessible to everyone.

FUNDING ACKNOWLEDGEMENT

The United States President's Emergency Plan for AIDS Relief (PEPFAR) and the Millennium Challenge Corporation (MCC) have partnered to support innovative and country-led approaches that promote evidence-based decision-making for programs and policies that address HIV/AIDS, global health, gender equality, and economic growth in sub-Saharan Africa. The $21.8 million Data Collaborative for Local Impact (Data Collaboratives) program was launched in April 2015 with funding from PEPFAR and oversight by MCC.

Data Collaborative projects are strengthening the availability and use of data to improve lives and empower citizens to hold governments and donors more accountable for results. The program aligns with broader U.S. government efforts to maximize the effectiveness of U.S. foreign assistance and with the Global Data Partnership's efforts to promote data collaboration to achieve the Sustainable Development Goals (SDGs).

Published in December 2018
under a Creative Commons Attribution-ShareAlike 4.0 International (CC BY-SA 4.0) license by Data Zetu
Located at the Tanzania Data Lab (dLab)
College of ICT, University of Dar Es Salaam
Ali Hassan Mwinyi Road
P.O. Box 31543
Dar es Salaam, Tanzania
habari@datazetu.org
+255 782 374 894
www.datazetu.co.tz | dlab.or.tz

Cover image: United States Africa Command / Stephen Linch

Contents

Abbreviations and acronyms v

Key findings vi

Recommendations vii

Introduction 1

A brief review of the literature on mapping data flows and use in Africa 2

A conceptual framework for data flows 4

Research questions 7

Methodology 8

Scope 8

Mapping datasets and flows 9

Identifying relevant and useful data for hyperlocal communities 9

The health sector in Tanzania 11

Governance 11

Managing the collection of health data 13

Types of health data 14

Collection of HIV/Aids data from facility to national level 16

Custodians and data sources in the Tanzanian health system 17

Flows of health data 17

Vertical data flows from the hyperlocal level to the national level 17

Vertical flows between hyperlocal communities and the district level 21

Return, vertical data flows: National level to hyperlocal level 22

Horizontal data flows: National level 23

Horizontal data flows: Hyperlocal level 24

Feedback loops, data quality and terminations 24

Open data 26

Conclusion 28

The flow of problem-relevant data to the hyperlocal level 29

Relevance of health data to community pain points 30

Usefulness of relevant data to community pain points 33

Intermediation 38

Conclusion 39

Concluding comments 40

References 42

Key informants interviewed 45

Appendices

Appendix 1: Key stakeholder and use network for the Tanzania health system 46

Appendix 2: Example of health data flows mapping method 47

Appendix 3: Custodians and data sources in the Tanzania health system 48

Appendix 4: Descriptions of data relevant to water quality and cholera Kyela District, Tanzania 49

Appendix 5: Facility visited in Kyela District by ward and type 59

About the authors 60

Abbreviations and acronyms

ART	antiretroviral therapy
CSO	civil society organisation
DHIS2	District Health Information System (version 2)
DMO	District Medical Officer
e-IDSR	Electronic Integrated Disease Surveillance and Reporting
eLMIS	Electronic Logistic Management System
EGA	eGovernment Agency
FFARS	Facility Financial Accounting Reporting System
GOT-HOMIS	Government Hospital Management Information System
HDC	Health Data Collaborative
HFGC	Health Facility Governing Committee
HFR	Health Facility Registry
HMIS	Health Management Information System
HRHIS	Human Resources for Health Information System
Ministry of Health	Ministry of Health, Community Development, Gender, Elderly & Children
MoH	Ministry of Health
MSD	Medical Stores Department
NGO	non-governmental organisation
NHIF	National Health Insurance Fund
PLANREP	Planning and Reporting System
PO-RALG	President's Office: Regional Administration and Local Government
RITA	Registration, Insolvency and Trusteeship Agency
U5BRI	Under-five Birth Registration Initiative
WDC	Ward Development Council
WEO	Ward Executive Officer

Key findings

1. Data flows are more definitively outwards and upwards from local communities and their health facilities than they are inwards or downwards.
2. Those structures in place to facilitate the flow of health data at the local level between the district council, health facilities, communities and CSOs are ineffective and do not enhance data flows.
3. There are several outward and inward data flows that terminate, nor are there sufficient feedback loops in the flow of health data. Terminations stagnate the flow of data in the system by precluding the creation of feedback loops.
4. The absence of functional feedback loops limits opportunities for improving the quality of health data, which was found to be of concern. A lack of feedback loops in the system also limits the ability of system to improve on the relevance of the data collected to hyperlocal communities.
5. A problem-focused approach to leveraging data for improved local decision-making makes it possible to identify local datasets which may not otherwise present themselves as useful. Identifying problems also brings into focus the potential of specific data within large supranational datasets.
6. To make health data useful to local communities requires concerted efforts by multiple stakeholders to compensate for a lack of data skills at the local level and a health data system with restricted flows to and between stakeholders at the hyperlocal level.

Recommendations

1. Ongoing efforts to integrate multiple, vertical sources of health data into an interoperable health management information system should be supported by donor funders in place of investments in new, standalone vertical health data systems.
2. New incentives for the collection of accurate and timely health data by health service providers accompanied by a clear definition of roles for those who collect and process data at the health facility level should be institutionalised by government at all levels of health service provision.
3. Ongoing efforts towards the standardisation and digitisation of data, accompanied by the training of dedicated staff responsible for data capturing at the health facility level, should be supported to further encourage accurate reporting of health data.
4. The reporting and correction of erroneous health data should also be encouraged and should be made possible by introducing sufficient time to check and verify data beofre it used for official purposes.
5. Improvements in upward flows of health data made possible by measures outlined in points 2, 3 and 4 above should be matched by stimulating downward data flows in order to hange of the status quo where most health facilities see themselves only as data collectors and not users. This could be achieved by linking the allocation of resources to (i) planning and reporting that is clearly data-driven at both the district and facility levels, and/or (ii) planning and reporting at the facilitity level that takes into consideration other health facilities in the district. To emphasize the need to promote data use at the sub-national level,
6. Horizontal flows of health data at the hyperlocal level could also be improved by broadening access to centralised data sources such as DHIS2. Health facilities, CSOs and research institutes should be granted increased levels of access to data in the health management information system, while remaining mindful of patient confidentiality and other privacy issues.
7. Health governance structures at the district level should be streamlined to improve (i) community-driven problem-formulation, (ii) the identification of problem-relevant datasets and (iii) the exchange of health data between villages, wards, health facilities and the district council.

Introduction

African countries have pledged to achieve sustainable development and inclusive growth by adopting the 2030 Agenda for Sustainable Development and the Agenda 2063. According to the Africa Data Revolution Report 2016, "[t]he need for timely and quality data to inform the two agendas exerts pressure on the national statistical systems more than ever before [...] At the national level, this can be seen in long-term national development plans and numerous legal, legislative and policy reforms aimed at improving the quality, timeliness, relevance, availability and accessibility of data". Referring specifically to the health sector, the WHO states that "information is the pillar of any system, to enable it to operate efficiently and effectively. Availability and accessibility of health information is crucial for generating a knowledgeable and healthy community. However, for information to be valuable it should be reliable, complete, timely and accurate" (WHO n.d.). But data alone is not enough – equally important are the actors and technologies that catalyse the flow of data, in so doing connecting the available data to the policy- and decision-makers instrumental in the allocation and distribution of resources in pursuit of equitable human development.

Research on the socio-technical dynamics of data flows can provide data initiatives wishing to stimulate the evolution of a data ecosystem with evidence-based insights on interventions that are more likely to connect data with decision-making at all levels. Mapping the flow of data between actors can reveal evidence of blockages, breaks, switchers and connectors – critical information to inform and evaluate interventions designed to unlock the potentials of data, and to shape policy to ensure that data's contribution is steered towards equitable development. Such mapping can also inform efforts to support data collaborations working towards the achievement of the Sustainable Development Goals (SDGs).

This research report presents findings from research on the flow of data in the United Republic of Tanzania, and specifically on the flow of data to and from the hyperlocal level in the health system[1] of that country. The report begins with a review of the

1 The reasons for the focus on Tanzania and its health sector, and a definition of what is meant by the hyperlocal level, are outlined in the methodology section of this report.

literature on data flows in developing countries. The conceptual framework which borrows from the work of sociologist Manuel Castells on the 'space of flows' is then set out before the research questions and methods are presented. A brief overview of the Tanzanian health sector is then provided, with a focus on the governance of data and existing data systems in the sector. The overview also presents findings related to sources of health data at different levels of governance. The overview is followed by further findings which are set out in two sections: the first describes findings on the flow of health data between the national and local levels in Tanzania, while the second section describes findings related to relevance and usefulness of data at the hyperlocal level. The report ends with concluding comments that seek to integrate the findings and to link back to the notion of connecting the space of flows with the space of places for the purposes of equitable and sustainable human development.

A brief review of the literature on mapping data flows and use in Africa

A recent attempt to map data ecosystems in Africa was published in the first edition of the Africa Data Revolution Report (ECA 2017). The Report presents findings of an analysis of the data ecosystems of ten African countries and finds that despite evidence of political will and commitment from non-government actors, combined with some evidence of experimentation, the legal, technological, human and financial resource architecture and infrastructure to develop data ecosystems on the continent are deficient. Changes within national data ecosystems are not occurring fast enough or at scale. The report itself offers at most a high-level overview of actors in national data ecosystems and provides little by way of a detailed mapping of actors and flows. Moreover, research on open data in developing countries shows that supply and demand often remain disconnected, that there are gaps to overcome before supply can lead to effective use, that there is more to data than policies and portals, and that the publication of government data is not only the responsibility of central or national governments.

Development Gateway (DG) research in 2016 investigated how what it terms "results data" are collected, shared and used across the health and agriculture sectors in three countries, including Tanzania. In the case of Tanzania, the DG report concludes with three recommendations: (i) reduce data collection burdens, (ii) improve the relevance of results information, and (iii) create incentives to promote data use, especially among district managers and service delivery staff. The research covered 17 districts in 6 regions in Tanzania and was based on interviews with 140 stakeholders. The DG report is an invaluable reference in terms of providing first-hand accounts of the actors and their experiences related to the collection of health and agricultural data in Tanzania. Importantly, the report includes a local-level perspective, but remains at a high level in terms of its analysis with findings aggregated across all districts; nor does the report present a detailed illustration of the data and actors in the data ecosystem to capture their relative positions and connections. The collection, flow and use of survey and other data (e.g. crowdsourced data) is not discussed with the report focusing on the DHIS2 health data management system. The report does not go into detail about

how data is used by communities – either directly, or indirectly via CSOs – to resolve the challenges they face. These 'shortcomings' notwithstanding, the report provides extremely valuable data and insights.

In 2008, Tanzania's Ministry of Health with support from the World Health Organisation mapped the procurement and distribution of medicines and medical supplies. The mapping identified partners, government agencies, donors, faith-based organisations, non-governmental organisations, private companies, and health facilities involved in the medicines supply chain in Tanzania (Ministry of Health and Social Welfare 2008). The study found that (i) most actors target specific diseases while multilateral donors target major endemic and epidemic diseases, and (ii) that donor organisations largely make use of public channels to procure and distribute commodities, while other actors make use of non-public channels. Common throughout the distribution system was a lack of information sharing among actors. More recent mapping of the medicines supply chain was done by Yale University researchers as part of Project Last Mile which sought to establish a novel partnership model that explored the transfer of Coca Cola's core business expertise in supply chain management to improve the distribution of pharmaceuticals and medical supplies in Tanzania (Linnander et al. 2017; Yale University 2012).

While research has drawn attention to the importance of intermediaries in catalysing the flow of data in data ecosystems, their function is dependent both on the availability of data (supply) and on an interest in the use of the data (demand). Data may be supplied by federal, national or sub-national governments, or by non-governmental, donor or research organisations. Or it may be generated by or crowd-sourced from communities. Demand may come from tech-savvy entrepreneurs, from government agencies, from journalists, from researchers, or from community-based monitoring initiatives, if it exists at all. The flow of data between actors of different types, its repackaging and redistribution as it flows between actors, and the incentives of actors to supply and use data create a complex, non-linear arrangement.[2]

Recent research that examined the impact of open data in developing countries, finds, among other things, that problem identification and the articulation of the usefulness of data in relation to real-world problems, greatly enhances data's potential impact: "The most successful open data projects are those that are designed and implemented with keen attention to the nuances of local conditions, have a clear sense of the problem to be solved, and understand the needs of the users and intended beneficiaries. Projects with an overly broad, ill-defined, or 'fuzzy' problem focus, or those that have not examined the likely users, are less likely to generate the meaningful real-world impacts, regardless of funds available" (Verhulst & Young 2017: 264).

Problem-identification is at its most specific at the community or "hyperlocal" level. The hyperlocal level implies any level of governance that is more granular than the

2 See, for example, the mapping of the restaurant health inspection data ecosystem in New York, USA, which describes the expansion of information flows, actors and governance relationships with advent of new technologies, particularly social media (Helbig 2012), and the mapping of the South African public higher education data ecosystem which highlights how even the simplest data ecosystem may feature multiple actors playing seemingly conflicting intermediary roles (Van Schalkwyk 2016b).

typically delineated "subnational" levels of governance such as the regional level (for example, provinces, states or counties) or the metropolitan level (for example, towns or cities). The exact coverage of the hyperlocal will vary across different forms of governance but are likely to include those described as precincts or wards (in the context of cities), or local authorities or villages (in rural areas). It is at the hyperlocal level that data is collected to be fed into subnational and/or national management information systems, and it is at the hyperlocal level that the same data can be most usefully applied to solve the problems faced by communities, and where the problems themselves are most salient. But this is also the level that is the furthest removed from ongoing national and supranational efforts to govern the collection and use of data for the purposes of development.

A recent project in Mexico on data use for improved fiscal governance, mindful of both the problem–value relationship and the local needs–global governance dichotomy, used a unique approach to overcome these challenges. "Data Treasure Hunts" are designed to follow a process of community-led problem definition and assessment followed by actionable strategies that ensure the use of data to solve prioritised challenges. Early findings from the project show that the Treasure Hunt method has helped communities to better analyse incentives and to explore how to leverage the interests of different local actors (Global Integrity n.d.).

To conclude, there remains a lacuna when it comes to an atomic-level understanding of the flows of data in data ecosystems, particularly as it relates to the collection and use of data at the hyperlocal level to solve problems faced by local communities.

A conceptual framework for data flows

Data flows are best understood with reference to the movement and exchange of raw and processed data between humans and machines in complex socio-technical systems.

The arrangement of actors, the process of exchange and the dynamic and changing properties of these systems have led many scholars to refer to these data systems as 'ecosystems' (Harrison et al. 2012; Heimstädt et al. 2014; Van Schalkwyk et al. 2016). Ecosystems consist of mutually interacting organisms; are complex in their arrangement; characterised by the interdependency of and between organisms and resources; are dynamic rather than static (seeking equilibrium through motion rather than stasis); populated by keystone species that play a critical role in facilitating exchange in the ecosystem thereby ensuring dynamism and constant movement. Movement in ecosystems tends to be cyclical and reinforcing making the system resilient (adaptable and restorative) but ecosystems are also vulnerable to exogenous forces which may disrupt or even destroy the ecosystem (Van Schalkwyk et al. 2016).

The presence of humans as key actors in data ecosystems introduces a critical feature of data ecosystems that differentiates them from biological ecosystems: the existence

of complex communication between system actors. Understanding data (and other human-centred) ecosystems therefore requires an understanding of the exchanges (or flows) of some of the key constructs of communication – data, information and knowledge – between actors.

Communication as flows of data, information and knowledge as critical features of socio-technical ecosystems are, in the age of digitisation and the internet, best understood in terms of networks (Castells 2009; Stalder 2006). According to Castells (2009, 2010), the internet and the communication technologies that have been developed to exploit real-time connectivity on a global scale has had profound effects on society. This is most evident in how society is being reorganised according to the programs of global networks, and the effects of this restructuring on the development of society (Castells & Himanen 2014). There are legitimate concerns that rather than making possible a more equitable distribution of resources, the network society is one in which exclusion is structurally manifest and the gap between rich and poor, the powerful and the marginalised, the metropolitan and the rural, is certain to widen (Castells & Himanen 2014; Ravallion 2016).

At the risk of oversimplification, Castells (2010) explains this structural binary divide as being attributable to a particular condition in the network society: the cleavage between two spaces – the space of flows and the space of places. The space of flows describes that placeless space where information is exchanged in real-time across global networks. The space of places describes the local, physical spaces in which social actors live and breathe; where they seek meaning and define who they are.

The disconnect between these two spaces in the network society provides a useful framework for understanding the flow and usefulness of information (and data) in advancing the development of local communities.

Data is routinely collected at the local level by city and municipal governments across the globe. These data are fed into state, national or federal management information systems for planning purposes which will ultimately provide the basis for determining the distribution of resources (political agendas and corrupt practices notwithstanding). At the same time, the data feeds indicators of a nation's financial health, including, for example, its economic growth and the investment climate which, in turn, rely on data describing the social conditions and well-being of defined populations. This is vital data that, when circulating in global financial networks, determines the ability of the state to access (and provide) financial instruments in the globalised financial system. This is not only about access to global financial networks, but about nation states becoming inexorably integrated into the global financial system. According to Castells (2017), three major processes are changing the coordinates of the global political economy, one of which is that global financial markets are increasingly the core of national and international economies.

However, the primary data and those who inhabit the hyper-local spaces from which primary data originate, typically remain disconnected from global communication networks. Stakeholder network analysis of the Tanzanian health data system shows

just how disconnected the local level is from the national (PATH 2017; see Appendix 1). Research has also confirmed a reliance on local experience and discretionary judgement over data for decision-making at the local level (Mori et al. 2014; Wickremasinghe et al. 2016). Custer et al. (2018) argue that improvements in sub-national data collection will bring data closer to place-based decision-making. In sum, aggregated national-level data may hold little value at the local level because that data's purpose is not defined in local but in global terms.

Open data could play a role in ameliorating the exclusionary effects of networks by democratising access to and use of a highly prized commodity in global networks – data (Van Schalkwyk & Canares 2019). The same is true for improvements in the sharing of data within government and between government and external stakeholders. More equitable access, in turn, makes possible, in theory at least, the reprogramming of networks by special actors in networks, referred to by Castells (2009, 2010) as switchers and programmers. In plain terms, increased access to data can mobilise social movements to effect change in networks programmed in terms that will not deliver equal and sustainable benefits. Governments themselves, as programmers, may take the lead to leverage open data for more equitable outcomes. In the health sector, for example, governments in various parts of the world are seeking to unlock the potential of open data to improve the quality of care, lower healthcare costs and facilitate patient choice (Scrollini 2017; Van Schalkwyk et al. 2017; Verhulst et al. 2014).

However, there is an assumption made by data advocates that increases in access to data will automatically result in a two-way flow of data between the local and the national levels. *Open* data advocates miscalculate: while they are attuned to the fact that flows of data in closed networks preclude return flows outside of those networks, they presume that more open flows will unlock blockages in return flows. This presumption ignores the disconnect between the space of places and the space of flows. There is evidence of a growing acknowledgement of the flaws in this assumption. Recognising the poor flow of data from the national to the local, researchers have increasingly turned their attention to the sub-national level (Canares 2016; Piovesan 2017) as have some governance initiatives (see, for example, the Open Government Partnership's Local Program,[3] the South African government's municipal budgeting and expenditure data portal,[4] and the What Works Cities project[5]). The attention paid to feedback loops in data flows further attests to an awareness of and concerns about low levels of two-way flows (Helbig 2012 et al.; Piovesan 2017).

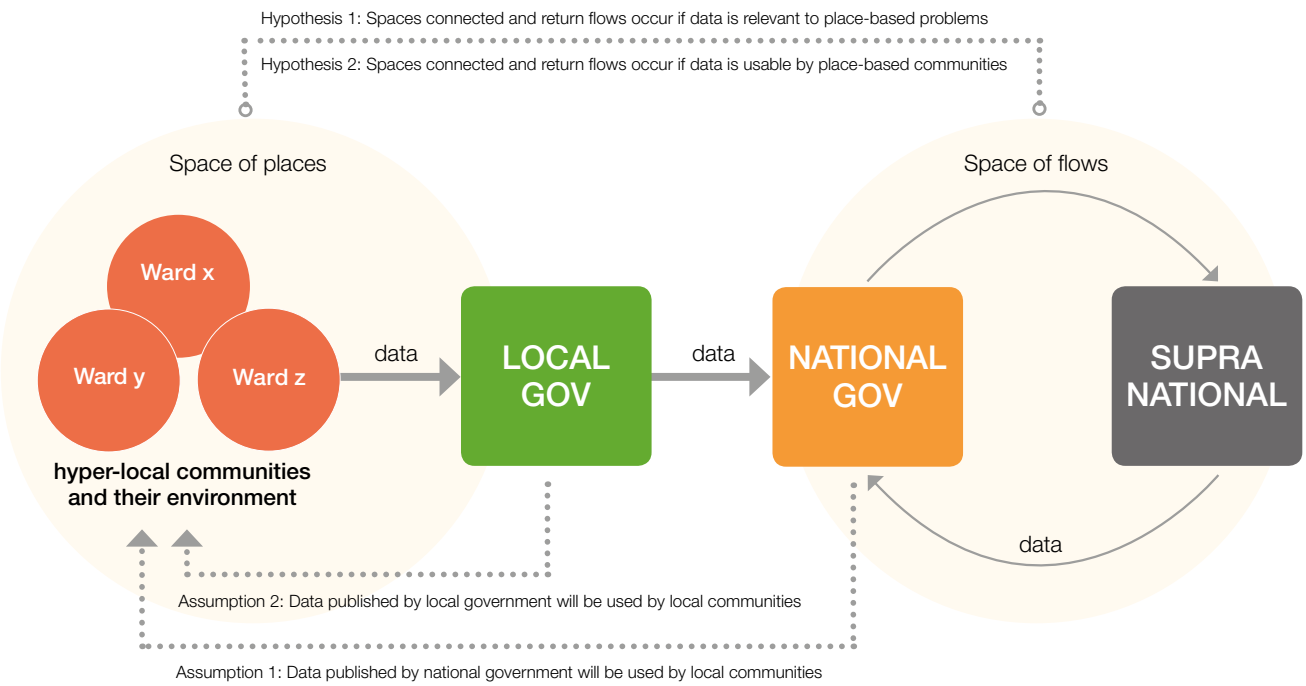
Focusing attention on flows in data published by local authorities and the impact that these may have on local communities is a logical shift. If the data at national level holds little value for local communities, and if local leaders are more connected to communities and therefore more attuned to their needs, then it is more likely that effective bi-directional flows will be observed at the local level. This exposes a second

3 https://www.opengovpartnership.org/local
4 https://municipalmoney.gov.za/
5 https://whatworkscities.bloomberg.org/about/

assumption: that local-level data will be usable by local communities. Research on the assumed usability of data by local communities has shown the important roles of intermediaries in connecting the needs of communities with available local-level data (Berdou & Shutt 2017; Verhulst & Young 2017; Van Schalkwyk et al. 2016a). While intermediaries may also make national data usable by local communities (Van Schalkwyk et al. 2016b), its translation is likely to be more meaningful if done at the local level (Canares 216) through clear problem definition and the forging of partnerships (Verhulst & Young 2017).

The conceptual approach outlined above is presented in Figure 1. The figure locates the two assumptions made about the flow of data from the space of flows (national and supranational levels) to the space of places (hyperlocal level). And the figure indicates the hypotheses that data relevance and usefulness can activate return flows to connect the space of flows with the space of places. To test these hypotheses, this research project will build on existing knowledge to (1) provide both a visual and descriptive account of data flows in a sector-specific data ecosystem, and (2) explore the relevance and usefulness of the data in the ecosystem to the problems faced by communities at the hyperlocal level.

Figure 1: Conceptual framework



Research questions

Based on the above, the broad question that this research poses is 'How does data flow in the Tanzanian health data ecosystem?' To answer this question, a subset of questions is posed:

1. What are the available datasets in the ecosystem?

2. How are the datasets arranged in relation to one another in the ecosystem?
3. How does data flow between datasets?
4. Where are the breaks and blockages in the flow of data?
5. Are the needs of the community connected to the available data?
    a. Is the available data relevant?
    b. Is the available data useful?

Methodology

Scope

This research formed part of the Data Zetu[6] project. The project sought to amplify citizens' voices through data. The Data Zetu project was located in Tanzania, and was funded and implemented in partnership by PEPFAR and the Millennium Challenge Corporation (MCC) as part of the Data Collaboratives for Local Impact program (DCLI).[7]

The Data Zetu project focused on three districts in Tanzania: Kyela, Mbeya and Temeke. The scope of the research was therefore confined to the activities of the project, both in terms of sectors and geography. Within the limits imposed by the scope of the Data Zetu project, the research component limited its focus to Kyela District in the Mbeya Region located in the south-western part of Tanzania. Kyela was selected from the three districts for the following reasons:

8. Temeke District, as a district within the city of Dar es Salaam, and Mbeya as a regional city, are both urban districts and therefore do not offer as distinct a separation between urban and rural as does Kyela.
9. Over two-thirds of Tanzanians reside in rural areas and rely on local health facilities managed by Local Government Authorities (LGAs) to provide them with basic health services (Boex et al. 2015).
10. The Data Zetu project enjoyed the interest and support of the district government in Kyela District.

The health sector was selected as the focus area for the research. The following reasons support the selection of the health sector:

1. The Tanzanian health sector is comprised of a diversity of actors, with interests and investments in the sector from the international donor community, national government agencies, local CSOs and others. Such diversity may not necessarily be the case in other sectors.
2. The Data Zetu project and its partners are familiar with health data ecosystem both at a national and supranational level. The research team therefore benefited from ready access to knowledge on the actors and data sources in the Tanzanian health system.

6 https://datazetu.or.tz/
7 https://www.mcc.gov/initiatives/initiative/mcc-pepfar-partnership

3. Based on the preliminary findings from the Data Zetu project's other activities, health data emerged as both relevant and meaningful to a cross-section of district-level communities.

Mapping datasets and flows

Information on the available datasets in the ecosystem was collected in the first instance by means of desk research and in consultation with the Tanzania Data Lab (dLab) as it works closely with the National Statistical Office and with CSOs in the health sector in Tanzania. By conducting interviews, searching the media and consulting other data mapping efforts such as the *Results Data Initiative: Findings from Tanzania*, 37 health-related data sources were identified and coded for: name [of the dataset]; collection type [e.g. database; dashboard; document repository]; types of data [e.g. health; demographics; education]; keywords [describing the data, e.g. mortality, malaria, TB]; collection [method of collecting the data, e.g. routine, survey, third-party); data owners [primary = owner of the source data; secondary = owner or host of the platform or medium for accessing the data]; data source [from whom is the data collected, e.g. district office, general population, health facility]; level [e.g. supranational, national, regional]; formats [e.g. excel, PDF, csv]; most recent data available [year]; available online [yes/no]; accessible [yes/no]; open data [yes/no]; and URL [to dataset or to page where data can be found].[8] This list is not meant to be exhaustive but rather represents a sample of the types of Tanzanian health data available, as well the levels at which the data are made available for use.

Once the identification of health datasets was complete, the flow of data was traced from the local to the national and/or supranational levels, and vice versa, using the list of data sources. Further unstructured key informant interviews verified mapped flows and identified new data flows (see a full list of key informant interviews at the end of this report). Informants were presented with the data flow map and asked to identify missing data sources, correct inaccurate flows and to describe any new flows suggested (see Appendix 2). This was an iterative process consisting of two batches of interviews in December 2017 and March 2018, and presentations to DCLI and partners in Tanzania in December 2017 and June 2018.

To record flows of data at the hyperlocal level, fieldwork was conducted in August 2018 by a team of three researchers. A structured questionnaire was administered to government health facility supervisors at 31 public health facilities in the Kyela District, to the District Executive Director (DED) and to the District Medical Officer (DMO). The data was captured in a quick-tap survey application using hand-held devices, and exported to MS Excel for cleaning and analysis.

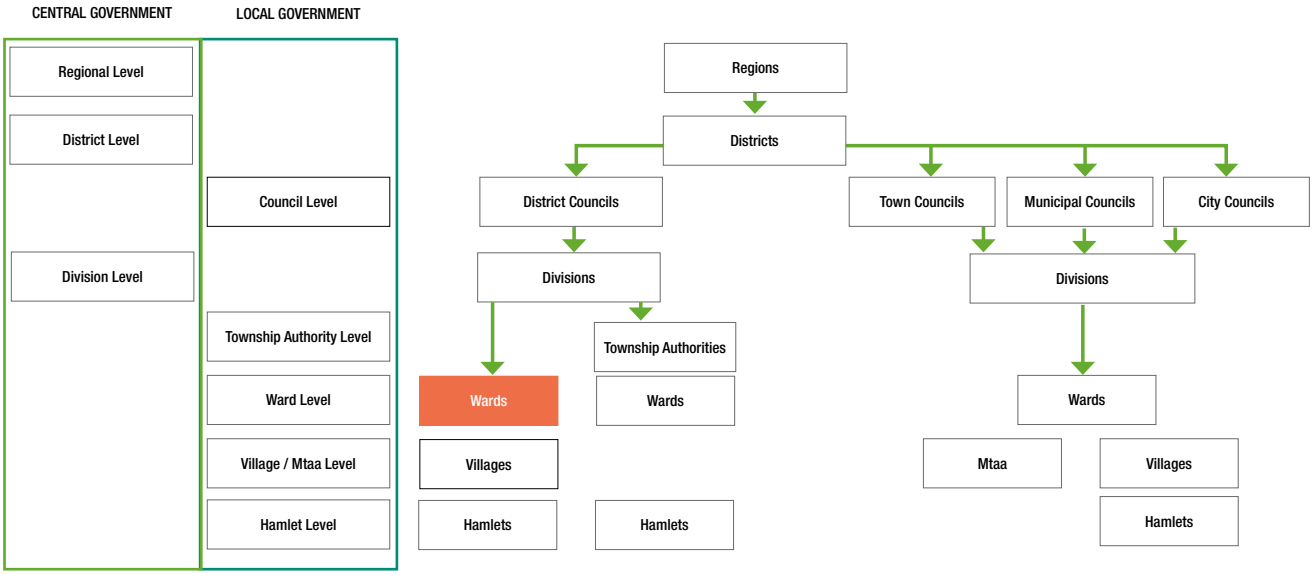
Identifying relevant and useful data for hyperlocal communities

The data needs and priorities at the community level in Kyela were conceptualised as community "pain points" and determined by the findings of the Data Zetu listening

8 The dataset is available at www.bit.ly/dzresearchdata

campaigns.[9] The listening campaigns focused on 4 of the 33 wards[10] in the Kyela District: Kyela, Matema, Mwaya and Ngana. Figure 2 shows how wards fit into the governance structure in Tanzania. Wards in Kyela constitute the hyperlocal level in this context adding a more fine-grained level of analysis to those typically used (i.e. supranational, national and sub-national). The hyperlocal level includes ward communities and their representatives; distinctive communities and administrative units within wards such as sub-wards and villages/*shinas*; and, from a governance perspective, the street representatives.

Figure 2: Levels of governance in Tanzania



Source: United Republic of Tanzania, President's Office, Regional Administration and Local Government (2018), p. 15

As is to be expected, the listening campaigns identified several community-level pain points. The Data Zetu team selected from these two key pain points: (1) water quality and outbreaks of cholera and (2) poor knowledge of rights within in the community. Given the scope of this research project on the flow of data in the Tanzania health system, only the first pain-point directly related to health was used to determine the availability and flows of *relevant* data in the data ecosystem. Each of the data sources identified in the mapping exercise was re-examined to determine whether the dataset contained any data relevant to water quality or cholera in the Kyela District. Additional data sources were identified during the interviews. Data relevance was determined by making reference to common interventions related to cholera outbreaks, including:

> mobilizing communities to promote personal hygiene, safe water and good sanitation; distributing chlorine for household water treatment and safe storage; conducting routine water sampling and laboratory analysis for contamination; and strengthening of case management in treatment

9 Further information on the Data Zetu listening campaign methodology is available online (Data Zetu 2017a), as is a full description of all the pain points identified in the four wards in Kyela (Data Zetu 2017b).

10 http://kyeladc.go.tz/statistics

centers and risk management during burial of deceased cholera patients (WHO 2016).

A final set of 16 relevant datasets were identified, of which 7 were new datasets not included in the original health dataset mapping.[11] These 16 datasets were then 'tested' for usefulness by assessing hypothetically their application and contribution to a selection of possible data-driven solutions to the outbreaks of cholera in Kyela District.

The health sector in Tanzania

Governance

The governance of the health system in Tanzania is described by government as a 'pyramid structure' (MOHCDGEC 2016: 4), with the community at the lowest level and the national government at the highest level. Cascading down from the national level, the institutional governance framework for health consists of zonal; regional; district, town, municipal and city; ward; and village levels (MOHCDGEC 2017). Each level is responsible for the monitoring and supervision of health facilities in their area of jurisdiction, and there is an increasing shift towards decentralisation (MOHCDGEC 2017). The sub-national health governance structures comprise Regional Health Management Teams (RHMT), Council Health Management Teams (CHTM), Health Facility Governing Committees (HFGC), and Village Executive Officers.

The coordination and management functions of the health system are the responsibility of the Ministry of Health, Community Development, Gender, Elderly and Children (MOHCDGEC, or MoH for short), in collaboration with the President's Office: Regional Administration and Local Government (PO-RALG). The main responsibilities of the MoH is 'the formulation of policies and technical guidelines, overseeing service delivery, managing and supervising National and Consultant Hospitals; whereas PO-RALG oversees regional and district hospitals, health centres, dispensaries and provision of various services at the community level through outreach clinics as well as community health workers' (MOHCDGEC 2016: 4).[12]

The health sector in general aims to deliver health services as close as possible to the citizens of Tanzania. Most common illnesses are assumed to be treated at outpatient facilities such as the local dispensaries or at health centres, especially in rural areas where transportation costs typically prevent communities from accessing either the district hospital or private health services (Boex et al. 2015). National Treatment Guidelines specify which health services should be offered at each facility type, and in which cases treatment should be referred to a higher facility level.

11 The dataset is available at www.bit.ly/dzresearchdata

12 Types of health facilities are defined as follows: (1) Hospital: Provides all medical services including surgical treatment and specialised care; (2) Clinic: Where outpatients are given medical treatment or advice, especially of a specialist nature; (3) Dispensary: A primary health facility which offers outpatient services including reproductive and child health services, and diagnostic services; and (4) Health centre: A primary health facility which offers outpatient and in-patient services, maternity care, laboratory, and dispensing and mortuary services.

Decentralisation of public health planning is meant to facilitate public participation in health issues. Health sector reform in the country places emphasis on the participation of lower level health facilities and communities in the health planning process. Preliminary results from an intervention study conducted in two districts in the Lindi Region enabling community representatives to take part in the district's annual health planning process, suggest positive outcomes from recent health policy developments in the country (Sato 2018). However, according to Kilewo and Frumence (2015), despite availability of policies, guidelines, and community representative structures, actual implementation of decentralisation strategies is poorly achieved.

At the district level, the Council Health Management Team (CHMT) has the remit for planning and budgeting activities needed to manage, control, coordinate and support all health services in the district on an annual basis (MoH 2011). District health priorities are set out in an integrated Comprehensive Council Health Plan (CCHP) which must make the best use of the available resources to meet the needs of local communities. Other responsibilities of the CHMT include: ensuring implementation of health activities by hospitals, health centres, dispensaries, and communities; and monitoring and evaluating the implementation of health activities in the district (Maluka et al. 2010). The CHMT consists of: District Medical Officer (chairperson), District Nursing Officer, District Laboratory Technician, District Health Officer, District Pharmacist, District Dental Officer, and District Health Secretary (secretary to the team). Other co-opted members of the CHMT may include the Reproductive and Child Health Coordinator, the Tuberculosis and Leprosy Coordinator, the Malaria Focal Person, the HIV/AIDS Coordinator, all of whom may be invited to CHMT meetings as and when the need arises. Representatives from the non-governmental and private sectors as well as from faith-based and other health service providers are also represented on the CHMT (MoH 2011).

The CCHP must be approved by the Council Health Service Board (CHSB), and the final plan is approved at the Full District Council Meeting. The Full District Council is the highest political body in the district and has overall authority for all district health services. After approval by the Full District Council, the CCHP is forwarded to the Ministry of Health and the President's Office Regional Administration and Local Government (PO-RALG) for final approval (or adjustment) before funds can be disbursed. The CHMT is linked to the communities through user committees and boards which are established at each health facility and in wards and villages (Maluka et al. 2010).

Public–private partnerships are also shaping the governance landscape. One example is the Mobile Health (mHealth) Tanzania Public–Private Partnership (PPP). Piloted in 2009, the partnership has developed several systems to collect and disseminate data on key health issues. The active monitoring of public health budgeting and expenditure combined with ongoing advocacy by the NGO Sikika is a further example of how non-government stakeholders influence the health governance landscape in Tanzania.

Managing the collection of health data

The health sector is in a phase of transition from manual to electronic data management. Development and implementation of ICT strategies in the health system has resulted in the introduction of new tools to manage health data. The change poses many challenges, including the high financial demand for system development, skilled human resources, capacity-building requirements, the high cost of infrastructure at various levels, and management issues. Guidelines, regulation, policies and strategies have been developed to support the performance of the health information systems.

The Tanzania Health Data Collaborative, an initiative set up to achieve a more coordinated and collaborative effort between all stakeholders to unlock the full potential of the government's health information system, acknowledges that Tanzania has made progress in improving its health data and information systems. This includes the introduction of the District Health Information System (dhis2) software for the Health Management Information System (HMIS), the drafting of the Tanzania Digital Health Investment Roadmap, the introduction of an electronic Government Hospital Management Information System (GOT-HOMIS), improved civil registration and collection of vital statistics as part of the Under-5 Birth Registration Initiative (U5BRI) and the Birth Registration System for 4th Generation (BRS4G), and innovations in community data collection through the electronic population register. However, fragmentation and a lack of integration between different data collection tools have impaired the quality of information and the sharing of health data between different stakeholders (United Republic of Tanzania 2016).

In Tanzania, health data is generated through two major channels:

1. Data at the population level, which includes:
    a. population-based surveys such as the Demographic and Health Survey and HIV/AIDS and Malaria Indicator Survey;
    b. the Population and Housing Census;
    c. the Demographic Surveillance System (DSS), which monitors vital statistics at sentinel sites located at various regions in country;
    d. sample vital registration with verbal autopsy, which operates under the sentinel panel of districts; and
    e. national vital registration systems and other specific health research.
2. Routine data at the facility level, mainly coordinated by the MoH via the health management information system (HMIS), also known by its Kiswahili acronym 'MTUHA' (*Mfumo wa Takwimu wa Uendeshaji wa Huduma za Afya*). Hospitals keep departmental registers to capture laboratory, physiotherapy and mortuary data. All health facilities make us of the 15 HMIS books/registers to collect and analyse health data:
    a. Book 1: HMIS guidelines
    b. Book 2: Facility and hospital summary book
    c. Book 3: Community book
    d. Book 4: Ledger book

e. Book 5: Out-patient register
f. Book 6: Antenatal care register
g. Book 7: Child register
h. Book 8: Family planning register
i. Book 9: Diarrhoea treatment book
j. Book 10: Monthly report book
k. Book 11: Dental register
l. Book 12: Labour and delivery book
m. Book 13: Postnatal register
n. Book 14: In-patient register
o. Book 15: Human resource register.

In addition to the two systems, there are specific 'vertical' programme systems (tuberculosis and leprosy, immunisation, HIV/sexually transmitted disease). To illustrate, the MoH through the National Tuberculosis and Leprosy Programme (NTLP) is implementing the fifth TB and Leprosy Strategic plan to strengthen the programme's monitoring and evaluation system (Ministry of Health and Social Welfare 2015, October). To achieve this, a more detailed electronic recording and reporting system has been developed in order to track programme's performance (NTLP 2017). This system has been developed by the Department of Computer Science and Engineering (DoCSE) at the University of Dar es Salaam and has successfully been piloted in 20 Councils in 5 regions of Mbeya, Kilimanjaro, Pwani DSM and Dodoma. The pilot was conducted in the period of January to June 2017. The programme has ambitions of rolling out the use of the system in all regions (NTLP 2017). See also the box below for a description of the collection of data for the HIV/Aids vertical health programme.

Types of health data

Routine data for the planning and oversight of the health system in Tanzania include data on morbidity; mortality; types of services delivered; commodities provided; availability and quality of services; financial sources and expenditures; and all management and supervision aspects. Data is collected in a multitude of different systems. It is claimed that there are in excess of 120 health or health-related data systems operated by the Tanzanian government and its partners (United Republic of Tanzania 2016). According to the MoH, these include:

- Health Management Information System (HMIS)
- Electronic Logistic Management Information System (eLMIS)
- Human Resources for Health Information System (HRIS)
- Training Records Tracking System (Train Tracker and Train Smart records)
- Integrated Disease Surveillance and Response Strategy (IDSR) reports
- Health Facility Sentinel Sites Surveillance Systems
- Supportive Supervision reports
- Routine Demographic Sentinel Surveillance (DSS) system

- National Facility Surveys: Tanzania Service Provision Assessments (TzSPA) surveys and Facility Mapping Surveys
- National Community Surveys: Tanzania Demographic and Health Surveys (TDHS) and Tanzania HIV and Malaria Indicator Survey (THMIS)
- National Identification Authority (NIDA)
- Civil Registration and Vital Statistics System (CRVS)
- PlanRep database
- Epicor database
- Geographic Administration Registry
- Special quantitative and qualitative studies conducted by the MoH and its partners
- Population and Housing Census
  (MOHCDGEC 2016: 52; Ministry of Health and Social Welfare 2015, August: 62)

Efforts are ongoing to make different health data systems within the National Health Management Information System (HMIS) interoperable (Ministry of Health and Social Welfare 2015, August). Currently, the central point of digital integration is the District Health Information system (dhis2), an open source software platform for reporting, analysis and dissemination of all health data, developed by the Health Information Systems Programme.[13]

Data from HMIS are also shared publicly via an online portal that has been developed by the MoH to disseminate cleaned HMIS data to health sector stakeholders and implementing partners. The data is disseminated via the HMIS in the form of tables, figures, graphs and geographic information system maps, and is disaggregated to the district level.

The source of health data in the HMIS is the health facility. Table 1 illustrates the various data capturing systems in operation at a typical district-level health facility (Silaa & Van Schalkwyk 2018). The table reveals the number and complexity of health reporting systems in place at the district health facility level. For example, Table 1 shows that there are at least six different reporting components that comprise the reporting of health data by health facilities in Tanzania. There are several reporting tools – at least 11 if one discounts that each may consist of sub-components – for which data is either captured on paper or digitally. Most reporting tools capture data for all diseases but the table also shows that there are separate reporting requirements for prioritised diseases such as HIV/Aids. Data are consolidated for reporting purposes on a weekly, monthly or quarterly basis depending on the data being reported. In general, it is the MoH that assumes responsibility and ownership of all health and health-related data collected in the public health system.

13 http://www.hisp.org/services/dhis-2/

Table 1: Data collection at district-level health facilities in Tanzania

Component	Reporting tools	Format	Disease	Reporting frequency
Health management	MTUHA registers	Paper	All	Monthly
	Patient Monitoring System (PMS)	Paper	HIV/AIDs	Quarterly
	DHIS2	Digital	All	Monthly
	CTC2 database	Digital	HIV	Quarterly
Hospital management	GoT-HOMIS	Digital	All	Monthly
Inventory management	eLMIS (Report & Request)	Digital	All	Quarterly
	CTC Pharmacy Module	Digital	HIV/Aids	Quarterly
	Ledgers & dispensing registers	Paper	All	Monthly & Quarterly
Financial management	FFARS	Digital	N/A	Weekly
		Paper	N/A	Monthly
Health insurance	National Health Insurance Fund	Digital	All	Monthly
Disease surveillance & reporting	eIDSR	Digital	Communicable diseases	Weekly
		Paper		Monthly

* See box below for a fuller account of HIV/Aids data collection. Source: Silaa & Van Schalkwyk (2018)

Collection of HIV/Aids data from facility to national level

HIV care and treatment services are monitored using the Patient Monitoring System (PMS) which consists of eight tools – five recording and two reporting tools. The recording tools include care and treatment cards (CTC) coded as CTC1 and CTC2, Pre-ART, ART and Cohort analysis registers. Care and treatment cards (CTC1 and CTC2) capture the patient's particulars which are fed into ART and Pre-ART registers for patients on ART and who are not yet on ART respectively. The ART register is used to track patients on ART for the period up to 72 months and provide longitudinal information to be fed into the cohort analysis registers and reports. The two reporting tools are the cross-sectional reports (quarterly) and the cohort analysis (bi-annual). The cross-sectional reports provide quarterly information on "cumulative numbers in care" and "on ART" and on "current in care" and "current on ARV", disaggregated by age and sex. Each facility that provides HIV care and treatment services summarises patient's information from the registers to produce quarterly facility reports, which are shared at the facility level for managing and improving service provision. These facility reports are sent to their respective district medical offices, where they are aggregated to form a district summary report. Eventually, the district summary reports are sent to regional level where aggregation is done to form a regional summary report, which along with other regional reports are sent to NACP for compilation and analysis.

Source: MoH, HIV/AIDS/STI Surveillance Report No. 23

Custodians and data sources in the Tanzanian health system

Data custodians and data sources in the health sector were mapped according to geographic governance levels in Tanzania (district, region, national) as well as beyond its borders (supranational). Health facilities in Kyela were included as they provide, together with the community and sentinel sites, vital data on the health of the population and on the performance of the health system at the hyperlocal level.

The findings reveal a multitude of health datasets at the supranational as well as several datasets at the national level (see Appendix 3). By comparison, very few datasets were found that are maintained by data custodians at the regional or local levels. Data on budgeting at the regional level by PO-RALG (PlanRep), and the regional and council health management plans are identified as data managed at the subnational level, and that are likely to contain data on the performance of the health system at the regional and council levels, but their value cannot be verified as these data were inaccessible to the research team.

CSOs and faith-based organisations either provide health services to local communities or conduct research in those communities. Both activities are likely to generate health data. In some cases, where those activities are funded by donors with open data policies, the data may become available at the supranational level via one of the many data repositories (such as the Dataverse, Dryad or those maintained by the donors themselves, most notably the Bill & Melinda Gates Foundation and the Wellcome Trust).

Flows of health data

In this section of the report, the flow of data in the system is discussed in relation to the map of data flows presented in Figure 3. It is acknowledged that mapping a complex system graphically will always fall short of reality which it attempts to represent. Nevertheless, it is hoped that even an oversimplified representation of a complex system can provide some useful insights into how data flows in the health data sector in Tanzania.

Vertical data flows from the hyperlocal level to the national level

Vertical data flows describe top-down and bottom-up flows. At first, the focus is on bottom-up flows, that is, from the hyperlocal level to the national level, because the hyperlocal level is where data originates in the health data system. This is followed by a description of the findings on top-down (return) flows of data from the national to the hyperlocal level.

The central source of health data are the 49 health facilities in Kyela District (Table 2). Patient data are captured as communities access services at health facilities located in villages and towns; and non-patient operational data are generated and captured as the health facilities deliver those services.

Figure 3: Data flows in the Tanzania health system

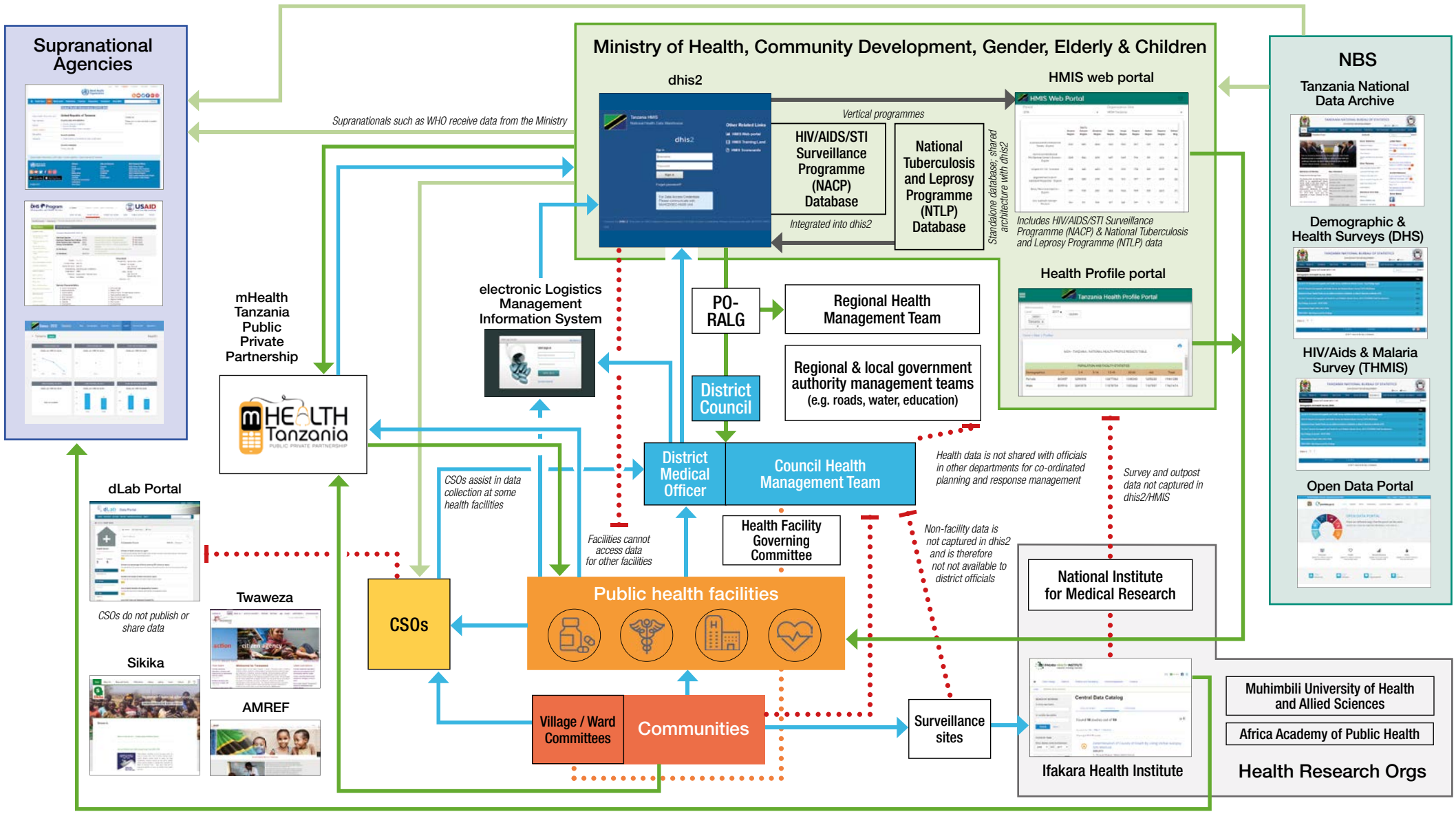


Table 2: Health facilities in Kyela District, by type and ownership

Type of facility	Ownership				
	Public	Private	Faith-based organisation	Parastatal	TOTAL
Hospital	1	-	1	-	**2**
Clinic	-	5	-	-	**5**
Dispensary	29	7	4	1	**41**
Health centre	1	-	-	-	**1**
TOTAL	**31**	**12**	**5**	**1**	**49**

Source: Opendata.go.tz, List of Operating health facilities – 2014, http://opendata.go.tz/dataset/85493939-6743-481b-8361-248d5f684153/resource/f9192849-9b32-4827-90fa-522ec1e84c1e/download/Operatating-Health-Facilities-with-location-2014.csv; updated using data from the Kyela District Statistics Portal as at 1 November 2018, http://kyeladc.go.tz/statistics

Figure 3 shows that at the most granular level, health data is collected at health facilities located in districts. All public and private health facilities (hospitals, dispensaries, clinics and health centres) capture patient and operational data, and submit these to the District Medical Officer's (DMO) office. Data on deaths, births, outpatients, inpatients, ante- and post-natal care, vaccinations, HIV/Aids treatment are just some of the data types that must collected at the facility level and captured in the forms prescribed by the Ministry of Health (MoH). The detailed, patient-level data captured by health workers are submitted in aggregated format for reporting purposes.

The national HMIS guidelines that have been developed and by the MoH set out the objectives of the HMIS system, describe the system in detail, list the required data collection tools (registers, tally sheets and summary forms) for every health programme, and contain instructions on how to use each data collection tool within the HMIS. The paper-based component of the HMIS system is referred to as MTUHA. The MTUHA comprises 15 books, of which Book no. 1 is the guideline and training manual of the HMIS system. The guidelines serve as a reference book for health service providers on how to collect, compile and report data.

Much of the data capturing is still done in paper format using prescribed printed forms provided by the MoH. Data captured in counter books are transcribed to the forms required by the MoH for processing by the District Medical Officer (DMO). It was found that 19% of health facilities capture data only in paper format with the remaining 81% using both paper and electronic data capturing methods. Data are typically converted into electronic format for submission to the DMO, although some still submit data to the DMO only in paper format (Silaa & Van Schalkwyk 2018).

Data submitted by health facilities to the office of the DMO are captured in the central dhis2 system by DMO office staff – a second instance of recapturing data in cases where paper records are submitted by health facilities. In addition to health data received from health facilities, DMOs capture other data in dhis2 specific to their district (e.g. demographic data) for the production of health indicators and reports from dhis2. These data are sourced from other government departments and agencies

such as the National Bureau of Statistics (NBS) by the district-level officials. While there is flow of data from the NBS to the DMOs, it was not established *how* officials in the DMO's office access and capture the non-health data obtained from NBS.

Data on commodities (i.e. medicines) is captured in the Electronic Logistics Management Information System (eLMIS) via MTUHA Book no. 4. In Tanzania, the eLMIS collects data from more than 6000 service delivery points (USAID 2015). Dhis2 is linked to eLMIS and extracts selected data from eLMIS (e.g. on tracer medicines) to produce health indicators. Interviewees commented that while the design of the eLMIS system looks good on paper, there are concerns about the comprehensiveness and quality of the data captured in eLMIS.

In addition to routine health data collection, the MoH has implemented the Integrated Disease Surveillance and Response (IDSR) reporting system developed by the World Health Organization Regional Office for Africa in 1998. IDSR was implemented to strengthen communicable diseases surveillance in the country. An assessment of IDRS in 2007 (Rumisha 2007) found that, in general, reporting systems are weak, both in terms of receiving all reports from all facilities in a timely manner, and in managing those reports at the district level. Routine analysis of surveillance data was not being done at facility or district levels, and districts were not monitoring the performance of their surveillance system.

Civil society organisations (CSOs) and other project-based initiates also participate in the collection of routine data at health facilities. Where relevant, this data is sent to the DMO for capturing into dhis2. CSOs are therefore active participants in the health data system but do not provide data directly to any central health data system administered by the government.

Both prior research (Tunga & Mushi 2016) and the interviews conducted by this research project show that CSOs are not publishing or sharing health-related data captured as part of specific projects that are not necessarily designed to assist health facilities in capturing routine health data. The sharing of health data among CSOs and between CSOs and the general public is therefore poor. Mostly data are used internally to produce reports, briefs and other outputs in support of advocacy work. Consequently, data sharing is restricted to disseminating printed materials or delivering presentations to relevant stakeholders.

Data does not flow from non-facility sources such as health surveillance sites or outposts to dhis2. Nor are survey or surveillance sentinel data captured in the dhis2 system, either via the District Medical Officer's office, the National Health Institute for Medical Research or Ifakara Health Institute, the latter comprising the corner-stone institution for government health research.

Survey data are also sourced at the population level in the form of population-based surveys such as the Demographic and Health Survey and the HIV/AIDS and Malaria Indicator Survey conducted by the National Bureau of Statistics (NBS); the Population and Housing Census also conducted by NBS; the Demographic Surveillance System

(DSS), which monitors vital statistics at sentinel sites located at various regions in country; sample vital registration with verbal autopsy, which operates under the sentinel panel of districts; and national vital registration systems and other specific health-related research. Results from surveys are typically presented in report-format to MoH and other interested government officials.

According to a district medical officer, the "[District Executive] accepted that village leaders can provide accurate data [when] we were distributing nets. They have been insufficient all the time due to underestimating population. After realizing that, we have been working with village leaders in identifying children under five and mothers. This information is always exact" (Bhatia et al. 2016: 12). In this case the medical officer is pointing out the problems of relying on inaccurate population data from the national level when distributing mosquito nets, and that the community is able to provide more reliable data to ensure an effective intervention.

However, any medical or health-related event that occurs within communities and does not involve a health facility, will not be captured in dhis2. In particular, out-of-hospital events such as births and deaths are often not recorded by health facilities. Such data is not included in any analysis done using dhis2 data, be it by the MoH, PO-RALG or the Council Health Management Teams (see below for more detail).

The MoH appears to be aware of the disconnect between research survey data, non-facility health data and decision-makers. It has proposed in its most recent policy document (MoHCDGEC 2017) the activation of a National Health Research Forum as a platform for knowledge exchange that will surface regional health priorities. The MoH also commits to an open access policy for research to make research more widely accessible. These are both positive steps in integrating research data into the health information system. However, other than to reference a "Data Dissemination and Use Strategy", the policy remains silent on exactly how research and non-facility data will be connected to and/or integrated into the health management information system.

The mHealth Tanzania Public-Private Partnership is one initiative that makes possible upward flows of data from isolated or disconnected health workers to national databases. The Partnership focuses on addressing ministry-defined public health priorities by supporting solutions that work in concert with initiatives underway at the MoH. mHealth initiatives include those that provide direct health communications to citizens via SMS and clinical decision information and reminder services for health workers. These services rely on existing data held by the MoH. But health data is also captured by mHealth initiatives by, for example, the IDRS system which allows health-facility workers to report disease surveillance data by making a free call from the field or by sending an SMS using any mobile phone. Real-time SMS and email alerts are then generated by the system for follow-up and action.

Vertical flows between hyperlocal communities and the district level

Health Facility Governing Committees (HFGCs) have been set up to create channels of

communication between communities, health facilities and the District Medical Officers office via a three-tiered community health 'team' consisting of (at the most local level) the Village Development Committee, the Ward Development Committee, and (liaising directly with the HFGC and the District Medical Officer) the Ward Executive Officer. Among other roles, the HFGCs are expected to raise issues and concerns in relation to community health concerns, discuss these with the respective health facilities, and then to report back to the community the agreed solutions provided by the health facility. The committees also collect information and concerns from local health facilities and share these with the community via the Ward Executive Officer. The HFGCs are expected to meet quarterly and to share copies of the meeting minutes with the District Medical Officer's office.

However, the District Medical Officer's office in Kyela District indicated the HFGCs have not been active in executing their responsibilities. And during the field visit to Kyela, members of the HFGC could not be interviewed because they had completed their three-year services terms of service and the district had not yet appointed new committee members (Silaa & Van Schalkwyk 2018).

The flow of data between communities, health facilities and the District Medical Officer therefore remains poor.

Return, vertical data flows: National level to hyperlocal level

According to the World Health Organization (n.d.): "Access to health information [in Tanzania] for all levels from the general community up to decision-makers, and the utilisation of the generated health information, have been inadequate and this is a major challenge to the health system. There are no standardised methodologies in place to ensure appropriate information is channelled to the right person at the right time and for the right purpose. Thus, there has been poor utilisation of the available information for knowledge strengthening and supportive evidence for decision-making."

District-level health management teams have access to dhis2 data and these data are accessed via reports generated by the dhis2 system. Bhatia et al. (2016) report, however, that analysis and planning at the district level cannot rely on dhis2 data alone and requires deeper analysis. The fact that community and other data sources are not available to district officials may hamper such analysis. For example, data from sentinel surveillance (including local data on births, deaths and burden of disease) are not used to calibrate and validate findings from facility-based data. And the fact that dhis2 does not include district-specific indicators (i.e. indicators relevant to specific district health needs), places further constraints on the availability and use of fine-grained, context-specific health data. A case in point in Kyela District are the seasonal cholera outbreaks that are not prioritised to the same extent as HIV and TB at the national and supranational levels. Data on cholera outbreaks are collected by several health facilities in Kyela District but there are no indicators for cholera on the national portals.

The flow of health data from the national and district governance level to the community is non-existent. Communities do not have any direct access to health data held by facilities nor do they have indirect access via intermediaries such as faith-based or civil society organisations. There are no formal mechanisms in place for health facility data to be accessed by ward-level governance structures, either via the Health Facility Governance Committee or directly from facilities. Only specific issues raised at a Health Facility Governance Committee that require the attention of the Ward Development Committee are communicated to the Ward Committee by the Health Facility Governance Committee chairperson (Silaa & Van Schalkwyk 2018). Due to the absence of mechanisms for sharing health data directly or for reporting to the community on a regular basis, village and ward committees lack access to health data.

CSOs also do not have direct access to the dhis2. The District Medical Officer for Kyela District confirmed that access to dhis2 is restricted to government organs and agencies. CSOs that work directly with the district can access health data by submitting in writing an official request which stipulates clearly the intended use of the requested data. However, according to the District Medical Officer, there are no guidelines for the release of data to CSOs or other stakeholders who may request health data from the district.

Horizontal data flows: National level

Horizontal data flows describe those data flows between government departments and agencies at the same level of governance. In this case, the intra-governmental flow of data between MoH and the National Bureau of Statistics (NBS) as well as the flows within divisions of MoH were mapped.

Data flows from NBS to MoH are either by request (usually for microdata) or by presentation (aggregated summary data), typically following the completion of one of the national surveys. Data are usually presented by NBS in reports made available in print and PDF formats. Data requests usually originate at a high level in government. For example, data requests to NBS often originate in parliament. In such cases, data requests are made by the permanent secretary in the President's Office to the Director of NBS.

Data flows within MoH show evidence of increased levels of coordination and integration. For example, collection of HIV/Aids data has been integrated into dhis2. According to the National Aids Control Programme's "National Guidelines on HIV and Aids Data Management" report, data is collected at facility-level using the CTC2 database, and dhis2 is the main repository for HIV and Aids data (MOHCDGEC 2017). Aggregated quarterly reports from CTC2 are captured in dhis2 while granular patient data remains in the CTC2 system. The HIV/Aids client records database has scaled to more than 900 facilities, improving record keeping for tracking HIV/Aids patients (Path 2017). The TB/polio database remains separate but shares architecture with dhis2. The eLMIS database also remains separate from dhis2, but dhis2 draws on data from eLMIS in order to produce key health indicators.

Horizontal data flows: Hyperlocal level

All dhis2 district-level health data is centralised by the District Medical Office, and health commodities data is centralized by the national Medicine Stores Department. While health facilities are the primary source of the national government's health data, health facilities can only access their own data on the national health management information system (dhis2) and on other relevant systems (such as eLMIS). They are therefore unable to access electronically the health data collected by neighbouring facilities in the district.

Limited horizontal access to health data at the facility level should not be interpreted as a lack of data sharing: 29 (90%) of facilities indicated that they are able to access health data from other facilities. Of those 29 health facilities, 94% indicated that they request data by telephone. In most cases (65%) requests are made only by telephone while in other instances, telephone calls are combined with a visit to the facility (10%) or with sending a WhatsApp message (19%). A small number of respondents (6%) indicated they make exclusive use of WhatsApp to request data from other health facilities.

Feedback loops, data quality and terminations

In its 2018 draft policy (MOHCDGEC 2017), the MoH states that the objectives for establishing a monitoring and evaluation system for the health sector in Tanzania include:

1. providing a mechanism for feedback to update the information system; and
2. providing reports to all stakeholders with necessary and sufficient data (MOHCDGEC 2017).

Figure 3 shows that only one feedback loop is in operation – the supply of facility data by the District Medical Officer's office to dhis2 and the use of the same data by district health planning officials, including the Council Health Management Team, for planning, response and resource allocation to health facilities from which the data originated.

Feedback loops are organic, social arrangements, and they require rewards to operate. Those rewards exist in the form of incentives for the initiator of the feedback loop, and for all subsequent data contributors and users. Incentives may take the form of material or non-material rewards. Bhatia et al. (2016) found that the key motivation for health officials at the facility level is to submit health data on time (most likely because timely submission is a key indicator in their performance evaluations) and that there are no incentives linked to the quality of the data submitted. Research by PATH suggests that data collection tools are designed to collect data for reporting purposes rather than to encourage or enable health workers to reuse their own data for planning and health-care service delivery (United Republic of Tanzania 2016). On this basis, upward flows appear to be efficiency- rather than quality-driven, and this has serious implications for the usefulness of the data for planning and resource allocation in return downward flows of health data in Tanzania.

Interviewees expressed concerns about the quality of the data transcribed as transcription is often done by nurses or administrators with poor data skills who are under pressure to submit the data on time. The MoH is aware of a range of issues contributing to poor data quality. From its own situational analysis conducted in 2016, it states that:

> a number of programmes and implementing partners had already identified data quality as a priority issue and were conducting data quality assessment and improvement activities. Challenges identified by stakeholders included: shortage of data collection and reporting tools at the facility level, lack of clear indicator definitions for routine data collection and reporting tools, human resource constraints (staff turnover, capacity), completeness of data collection tools, timeliness of reports, and inconsistencies between data collection tools and aggregate reports. Some stakeholders identified lack of data use as contributing to poor data quality. (MoHCDGEC 2016: 8)

Kimunai (n.d.) and Silaa and Van Schalkwyk (2018) report that some health facility staff would ask those with data skills to capture data on their behalf because they lack the skills to do so. "Data cooking" was cited by one interviewee as being common, as data capturers make up data to compensate for data gaps in the health facilities' records. Sato et al. (2017) measured aspects of motivation among health workers in rural Tanzania. They found that having a clear job description is the greatest motivation for health workers in executing their functions. It follows that if nurses and other health workers whose description does not include data transcription and capturing for reporting purposes (as opposed for patient management) are expected to undertake these data-related functions, then their motivation for completing these functions accurately will be low.

Dhis2 officials indicated that data quality tools have been made available to health officials to check the quality of the data submitted to and by the DMO, and that the MoH has been rolling out data training at the facility level to improve skills and capacity in data capturing. In the office of the District Medical Officer, an HMIS coordinator also works to validate the data in the system. Health facilities are contacted to verify data if any anomalies are detected by the HMIS coordinator.

However, respondents indicated that corrected data may not be updated in dhis2 because once the data submission deadline has passed, making changes to the data in the system raises "a lot of issues/concerns". The implication of this reluctance to update incorrect data after the submission deadline is that the corrected data will be held at the DMO's office but will not be recorded in dhis2. It was also indicated that some data go through more thorough checking processes because donor funding is available. This means that some types of health data (e.g. HIV-related data) receive more attention based on the funding priorities of supranational donors.

While training may help to prevent errors, and tools may flag obvious data errors, it is not clear how manufactured data can be prevented and detected when the incentives for accurate data collection and for capturing corrections are absent at the hyperlocal level. This raises serious concerns about the quality of the data in dhis2 used to conduct modelling, analysis and, ultimately, to make decisions about the allocation of health resources (see also United Republic of Tanzania 2016).

In many other instances, Figure 3 shows terminations in the flow of data, thus precluding the creation of feedback loops. Boerma (2013) argues that districts require health information systems that draw from multiple sources. And Ikonje (2014: ii) finds that in the case of the treatment of malaria in under five-year-olds in Tanzania, "data use and decision-making is an important component of quality improvement at all levels". To some extent, integration of data sources has taken place at the national level, but non-facility health data from surveillance sites, from the community and from CSOs do not flow into the health information system. Thus, there is no formalized system for District Medical Officers and health teams to access non-facility health data. And, as a consequence, there are no systematic local feedback loops between health officials and these sources of health data. As decentralization in Tanzania continues, local decision-makers will have a greater opportunity to use results data for local decisions. Systemic feedback between CHMTs and health facilities could go a long way to incentivise good performance and reporting.

At the same time, CSOs are either loathe or ill-equipped to share health data, and findings show that research institutes appear more inclined to share data with international donors than with either the national health ministry or local health officials. Health facilities in the districts are not able to access each other's patient and operational data via dhis2. If they were able to do so, this could improve district-level coordination in cooperation with the District Medical Officer and other stakeholders active at the hyperlocal level.

There is no mechanism for citizen-generated data, such as community-mapped health information, to be captured in any of the central health management information systems. Community reporting could play an important role in upward information flows via the IDSR or other mHealth initiatives so that communities are able join local and community health workers to create a wider network of disease reporters.

Open data

Several sources of open health data are available in the Tanzania health data system, although they vary in the extent to which they meet some of the requirements often prescribed for open data (such as being published under an open license, being machine readable and being available in bulk). Nevertheless, these datasets are all online and accessible without restriction.

Open data made available by the government comes from three main sources. The first is data published by the MoH predominantly from its health management

information system, dhis2, on two publicly accessible data portals: (1) the HMIS web portal and (2) the Health Profile portal.

The HMIS portal makes available indicator and raw data primarily on diseases but also includes demographic data (e.g. population by age), data on health personnel (e.g. health workers by cadre) and the completeness of health data collection across the system. The HMIS portal also includes data from two vertical health programmes: the HIV/AIDS/STI Surveillance Program (NACP) and the National Tuberculosis and Leprosy Program (NTLP). The NACP's data has been integrated into the dhis2 while the NTLP remains a separate database from which dhis2 extracts relevant data for producing indicators. The Health Profile portal's focus is on data and indicators related to health service delivery, including resourcing, medicines and infrastructure. Both web portals publish data disaggregated at the district level.

The second source of open data is from the National Bureau of Statistics (NBS). Data is published on the Tanzania National Data Archive and includes data from the health-related surveys conducted by NBS. The Tanzania Open Data Portal also falls under NBS and contains 12 health-related datasets mostly uploaded by the MoH.

A third source of open health data is the data published by the Ifakara Health Institute. At the time of writing only 8 of 109 datasets listed in its central data repository were either available as "Direct data access" or as "Public use data files".

Many of the open data originating from these government sources are republished in supranational open data sources such as those of the World Bank, the World Health Organisation and Unicef. Data is also published on the websites of those donor organisations that have invested in the Tanzanian health sector (for example, USAID and the Bill & Melinda Gates Foundation).

Both communities and health workers at the facility level can, provided they have the means and the skills to do so, access the HMIS public web portals that provide periodic data at the district level, but not at the level of facility. Using the HMIS web portals requires more than literacy; the portal describes data using sector-specific jargon and acronyms making it difficult for non-specialists to find and understand the data presented. The portals were also found to be slow and unreliable, and contained several incomplete data fields.

Findings show that 23 (74%) of the health facility supervisors interviewed were not aware of government's open health data portals and only 4 (13%) had made use of the portal in one way or another. Reasons given for using the health portals included accessing health guideline updates, familiarisation with 'hot cases' (disease outbreaks), for requesting medicines and for personal reasons. When asked whether they had ever used other, non-MoH data sources (e.g. WHO, UNICEF), 3 (10%) respondents indicated that they had used non-government sources, mainly to find new information pertaining to vaccination and disease management.

CSOs indicate that they struggle to access government health data, both at the national level and at the district level. One interviewee indicated that while government

welcomes the monitoring function performed by her organisation, government also responds to requests for data by saying that it does not want "our own data to be used to kill us". There are, however, examples of district health management teams and facilities responding positively to requests for data, expressing their willingness to engage with social accountability monitoring initiatives.

The Tanzania Data Lab (dLab) has been working with health CSOs on the publication of open health data and has made available a data portal to facilitate data publication. To-date hardly any CSOs have made use of the portal. CSOs that receive funding from donors who mandate data sharing as a condition of grants awarded, are publishing anonymised datasets either on the data repository of the funder (e.g. Wellcome Trust and Bill & Melinda Gates Foundation) or in an approved data repository (e.g. Harvard Dataverse, Figshare).

Conclusion

The findings from this research project on the flow of data in the Tanzanian health system show that there is a preponderance of health-related datasets at the supranational level and a dearth of data at the local level. Data collection is centralised via the Health Management Information System's dhis2 software at the national level and by the District Medical Officer at the local level. Local health facilities and other actors in the health system active at the local level cannot access granular, up-to-date data on all health facilities via the dhis2. Selected indicator data is, however, made available on public web portals.

Findings show that data flows are more definitively outwards and upwards from local communities and their health facilities than they are inwards or downwards. Those structures in place to facilitate the flow of health at the local level data between the district council, the health facilities, communities and CSOs are ineffective and do not enhance data flows.

There are several outward and inward data flows that terminate, including the flow of sentinel surveillance data. The consequence is a general absence of feedback loops resulting in stagnation in the flow of data in the system. A deficiency of feedback loops also precludes the possibility of improving on the poor quality of data in the system (Piovesan 2017; WHO n.d.) and limits the ability of system to improve on the relevance of the data collected to hyperlocal communities.

Strong upward flows feeding national-level indicators which are then used at the regional and local levels for planning and response purposes, may lack relevance at the hyperlocal level, and more so if local communities are unable to feed into the system data and information directly related to their day-to-day challenges. It is to the findings on the relevance and usefulness of health data to local communities that the following section turns its attention.

The flow of problem-relevant data to the hyperlocal level

The Data Zetu project facilitated ward-level listening campaigns in the Kyela District of Tanzania to identify problems or "pain points" faced by the communities in four wards in the District. The main health-related community pain-point identified was the outbreak of cholera, precipitated by an intersection of factors including seasonal flooding from March to May, poor water management infrastructure, a lack of health facilities and shortages of medicines (Data Zetu 2017b).

Despite an abundance of rivers and lakes in Kyela District, dug wells remain the main source of water for the majority of households in the communities consulted. In Kyela Ward, some streets are served by piped water from an underground well operated by local government. In all cases, the communities reported that water provision was unreliable, and the water quality was poor.

Communities also reported challenges in accessing health care in the district. Mwaya Ward consists of ten villages but there are health facilities in only three villages (see Appendix 5). In Matema Ward, the community reported that there is single health facility operated by a faith-based organisation and that it is too expensive and unable to serve the entire ward. The district hospital in Kyela cannot cope with number of referrals from district health facilities. Communities from all four wards expressed frustration at the lack of medicines and the pervasive practice of doctors at public health facilities referring patients to private pharmacies to obtain medicines.

The cholera situation is further exacerbated by infrequent water quality monitoring, lack of adequate sanitation facilities and poor hygiene practices. The WHO reports that:

> Assessments have shown that the main factors associated with the spread of the infection both in mainland Tanzania and Zanzibar are limited access to safe water and sanitation. Water supply institutions lack the capacity to chlorinate and conduct regular water quality monitoring and assessments. The coverage of improved latrines is also very low. Poor hygienic practices perpetuated by myths and misconceptions, and the lack of sanitation facilities in poor households have also been associated with the outbreak. (WHO 2016)

At the national level, the Tanzanian MoH is certainly aware of the issue of water quality in the country. From the 2018 draft Health Policy:

> The situation of water, sanitation, hygiene and food safety services in Tanzania has shown a significant improvement. [...] Despite the efforts taken by the government, Water, Sanitation, Hygiene and Food Safety problems remain of big concern in the country as observed in annual health reports that more than 60% of OPD diagnosis has been associated with poor sanitation and hygiene practices. This has

> been due to inadequate enforcement and overlapping of various laws and regulations (TFDA Act 2003 and Public Health Act 2009), weak coordination among stakeholders dealing with sanitation and hygiene issues. (MOHCDGEC 2017: 16-17)

Recent reports confirm the persistence and increase in cholera outbreaks in mainland Tanzania (WHO 2018) as well as the risk to communities in the Kyela District:

> Active cholera transmission has persisted in Tanzania Mainland, with Mbeya and Songwe Regions being the most active. … The closure of the fishing camps along Lake Rukwa on the Songwe Region border, forced the fishermen to move further south to Kyela district which borders Malawi's Karonga district. (WHO 2018)

In the sections that follow, the extent to which data identified during the Tanzania health data mapping is *relevant* to the specific health problem identified by the community in Kyela is determined. Thereafter, the *usefulness* of the relevant data is discussed in the context of possible solutions to the cholera pain point articulated by hyperlocal communities in Kyela.

These solutions are, in the main, hypothetical test cases based on the information gathered and insights gained while conducting this research project. What is presented is not meant to be an exhaustive list of solutions, nor are the solutions grounded in the sense that they have been explored with the community and other stakeholders who are likely to be instrumental in the success of the proposed solutions. Nevertheless, the list of solutions to the cholera problem does illustrate the number of different approaches possible, and the varying degree to which the data that is currently available could prove useful to alleviate the plight of communities in Kyela in relation to cholera.

Relevance of health data to community pain points

Figure 4 shows those datasets identified that are of possible relevance to the water quality and cholera community pain point in the Kyela District of Tanzania. Each dataset is described in detail in Appendix 4 of this report.

The datasets in orange boxes are those sourced from the datasets identified during the health-dataset mapping. The six datasets in green boxes are new datasets that were identified after the community pain point had been identified. The presence of these new datasets illustrates how the potential of certain datasets only becomes apparent once a problem has been articulated. Similarly, the absence in Figure 4 of many datasets identified in the broad health-related datasets mapping exercise shows how relevance only becomes apparent once a clear problem statement exists.

The datasets in Figure 4 are arranged concentrically to show their "closeness" to the community. In other words, datasets closest to the community are those located at

the ward level while those furthest from the community are at the supranational level. In this sense, the level represents the location of control over the data rather than the granular level of the data.

Figure 4 shows that as was the case for all health data, there is limited data at the ward or hyperlocal level, and at the district level. Only health facilities and the District Medical Office are known to hold relevant data at the ward and district levels. However, as one moves from the regional to the national and supranational levels, more relevant datasets are apparent, with most of the relevant datasets being at the supranational level.

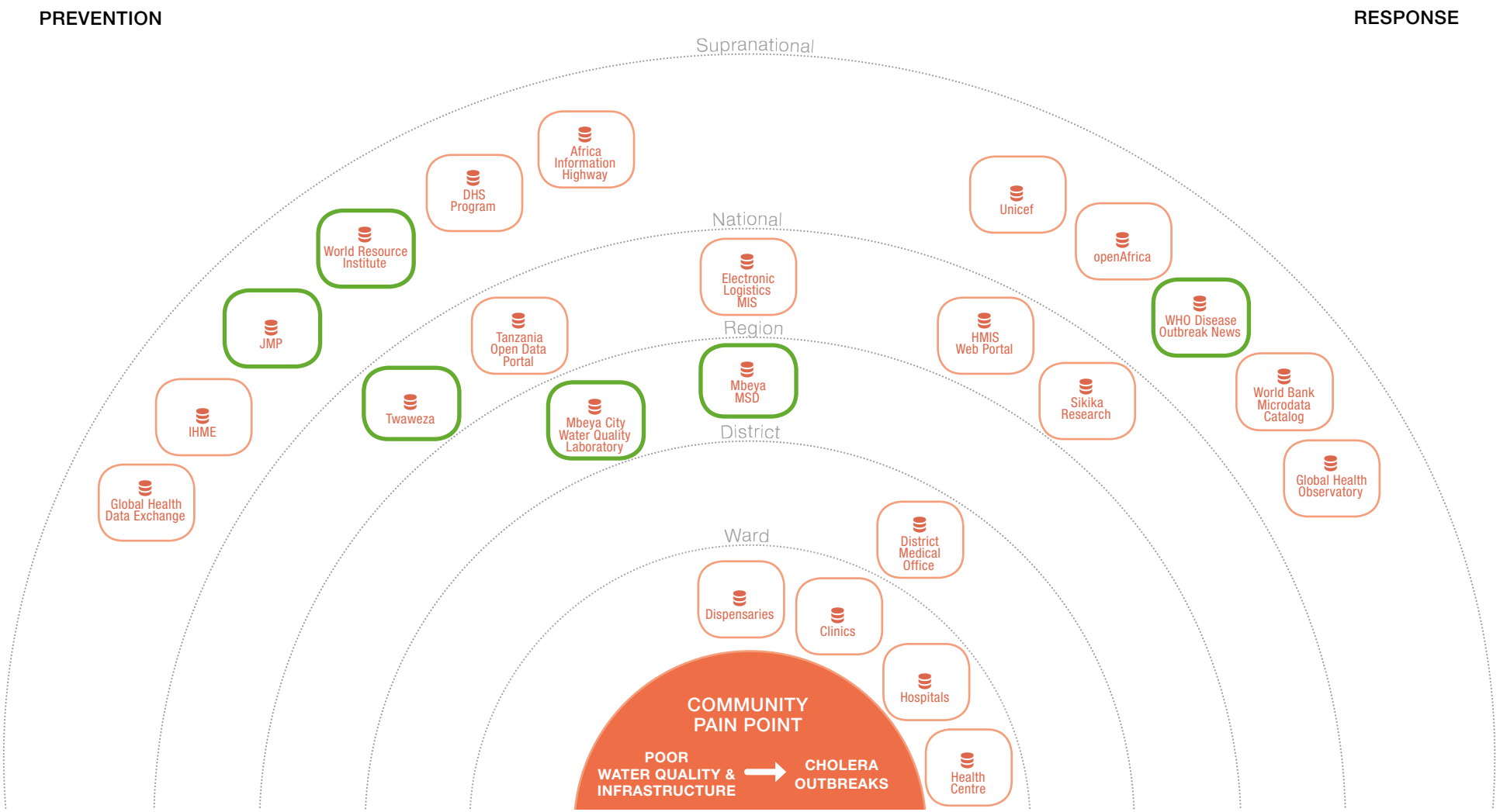
One could argue that Figure 4 masks the fact that there are, in reality, 49 relevant data sources at the hyperlocal level if each of the health facilities is treated as a single source of relevant data. It is certainly the case that one would expect to find relevant data at each health facility given that each facility is required to submit regular reports to the District Medical Office (as shown in Figure 3). From the health data stock-taking at the local level, it was established that health facilities do in fact collect data on cholera. Out of 31 public health facilities interviewed, 9 facilities reported that they experience annual outbreaks of cholera. During these outbreaks, cholera data are collected and transmitted but not are not captured and shared in the same way as routine health data. During a cholera outbreak, patient data are captured in an electronic mobile-based system (eIDSR) by health workers at facilities and in communities, and reported weekly to the national MoH.

Integrated Disease Surveillance and Response (IDSR) is a strategy developed by the World Health Organization Regional Office for Africa in 1998. The Ministry of Health in Tanzania has adopted IDSR for strengthening communicable diseases surveillance in the country. At the same time, relevant data from patients is captured in paper forms (such as the MTUHA no.5 OPD register and register no. 14) and submitted to the DMO's office for capturing into dhis2.

Each health facility could therefore be a valuable source of data. However, health data remains relatively inaccessible. While data is centralised at the District Medical Office, data are not available in full to each of the health facilities. Moreover, village committees are meant to be a source of hyperlocal data because the system has been designed such that villages collect data to feed into bottom-up planning processes and for monitoring performance. However, in reality, village committees do not collect data (Aiko 2018). Therefore, it remains the case that most of the relevant and accessible data is at the national and supranational levels.

The datasets in Figure 4 are also arranged according to whether they are relevant to water quality and therefore related to the prevention of cholera outbreaks (those located on the left-hand side), or whether they are relevant to the outbreak of cholera and therefore related to a response to such an outbreak (those located on the right-hand side). Datasets in the middle of the figure are relevant to both the prevention of and response to cholera outbreaks. For example, data on medicine stocks would include data on preventative commodities such as water purification tablets as

Figure 4: Data sources relevant to water quality and cholera in Kyela District, Tanzania



well as data on commodities needed to respond to a cholera outbreak such as oral rehydration medicines and antibiotics.

Figure 4 shows a fairly even distribution of data relevant to prevention and response at the national and supranational levels. At the regional level, data collected by the water quality monitoring lab in Mbeya City (the capital city of the region in which Kyela District is located), could play an important role in the prevention of cholera outbreaks. At the district and ward levels, however, data is mainly relevant to a response to cholera outbreaks. In the same way that the Medical Stores Department data is relevant to both prevention and response, so too is the data held by ward-level dispensaries. But clinics, hospitals and the health centre data, as well as the data received by the District Medical Office from those facilities, are more likely to contain data that alert health authorities to an outbreak once contamination and infection have already taken place.

Usefulness of relevant data to community pain points

It is important to distinguish between data's relevance and its usefulness to communities. Relevance is important as it allows for the identification of data sources and datasets that are linked directly to community pain points. Usefulness depends on additional factors, including but not limited to the quality of the data, how up-to-date the data is, its accessibility, the terms of its reuse, and the ease with which it can be interpreted – in short, the extent to which relevant data can be applied to resolve the problem.

How can the data sources identified be used to tackle the cholera problem in Kyela District? At the national level the response has been as follows:

- The Ministry of Health is collaborating with the WHO and other partners (including Unicef and the Centres for Disease Control and Prevention) to implement and monitor cholera control activities through the National Task Force.
- On 15 December 2017, the National Cholera Response Plan was reviewed and validated by the MoH and stakeholders.
- A Cholera National Task Force meeting was convened in December 2017 and agreement was reached to deploy a team to assist Uvinza District to respond to the increase in the number of reported cholera cases.
- The MoH has deployed a rapid response team to Ruvuma Region which reported cases of cholera for the first time in November 2017.
- The MoH has launched a community sensitisation and awareness campaign through local radio, television and social media.
- Unicef is supporting the promotion and distribution of chlorine tabs for household-level water treatment and storage. (WHO 2018)

Several other data-driven solutions could be pursued and these are explored below.

Water quality monitoring

The Water Quality Services Division of the Ministry of Water and Irrigation has a network of 16 Water Quality Laboratories across the country, one of which is located in Mbeya City. Water quality testing is done on an ad hoc basis by the Mbeya laboratory located in Mbeya City. Results data from water testing by the lab in Mbeya City have not been shared with the District Executive Director, civil society organisations active in the region, nor with the affected communities.

To improve the monitoring of water quality and the identification of sources of potable water during times of flooding, water sensors could be installed at key water collection points in the District. However, the lab in Mbeya has pointed out several challenges with the use of water sensors: (1) where to position the sensors; (2) the ability of sensors to record and measure relevant markers for cholera contamination; and (3) the possibility of new water sources becoming contaminated due to deteriorating water management infrastructure.

Regardless of the approach adopted to improve the testing of water quality at the hyperlocal level, it will be important for any data collected to be shared more widely than is currently the case and that whatever water quality data is made available, is shared in formats that are easy for community members of all ages to understand.

Chlorine disinfection of water

Chlorine disinfection is widely effective against waterborne bacteria is one of the most inexpensive and accessible methods for water treatment by households. The effectiveness of chlorine is limited, however, by turbidity. In Tanzania, there are no guidelines on water chlorination at household level, and limited data on whether dosing guidelines for higher turbidity waters are sufficient to produce potable water. A study conducted to assess the effectiveness of chlorination on rural water sources using two locally available chlorine-based disinfectants – WaterGuard and Aquatabs – found that water chlorination can be effective (Mohamed et al. 2015).

A 2016 survey found that 61% of households reported that they treat their water: 49% boil their water, 27% strain their water, and only 7% use chemical disinfectants such as WaterGuard (7%) to make their water safe for drinking (Twaweza 2017).

Unicef and the CDC developed a bulk water chlorination strategy for highly affected cholera communities in Tanzania. Water vendors with water trucks were enrolled in the programme, supported by chlorine tablets. Findings indicated a low incidence of cholera cases in targeted areas as a result of these efforts (Unicef 2016).

On 25 April 2018, Msufini Tanzania Limited (MTL) officially started the construction of the first chemical plant in Tanzania located in Mlandizi, about 54 km north-west of Dar es Salaam. The plant will convert un-iodized salt into chlorine. The plant is set to be functional by March 2020 and will produce chlorine for the disinfection of water (Tanzania Invest 2018).

Given the above, how could the data sources identified be put to use? Data on the availability of health commodities (such as oral rehydration kits and certain antibiotics in the case of cholera) can help medical staff to respond more effectively to outbreaks. Data on the availability of oral rehydration medication supplied by the Medical Stores Department via its Mbeya zonal office to health facilities in Kyela District could be used to track the supply of oral rehydration kits at health facilities in or near the communities affected by cholera outbreaks as well as at the zonal medical store. Data on oral rehydration medication is published online by the MoH and is available at the district level. But the data is only published quarterly and is not disaggregated to the facility level. This places limitations on the usefulness of the data.

Data on the supply and availability of chlorine tablets and antibiotics, both important in the prevention and treatment of cholera, could also be useful. However, it is not known whether the MoH distributes chlorine tablets via the MDS nor whether the prescribed antibiotics for the treatment of cholera are available as part of the tracer medicine data published on the MoH's web portals.

The WHO's Procurement Services provides cholera kits through its regional offices and Unicef is supporting the promotion and distribution of chlorine tabs for household-level water treatment (WHO 2018). These organisations may be able to provide supplementary data on the distribution and availability of water treatment products in Kyela District.

Infrastructure

Aside from water quality, the other big issue according to WaterAid Tanzania, is one of infrastructure (Hicks 2014). At the governmental level, water and health are in separate "silos" – dealt with by different departments, with little connection between them. This means, for example, that large sums are spent treating waterborne diseases such as typhoid and cholera which could be prevented by investing in infrastructure to ensure the reliable supply of potable water.

The MoH acknowledges the shortcomings:

> Despite the government efforts in strengthening Water, Sanitation, Food safety and Hygiene services, weakness still exists. [...] Majority of Health and their Environmental determinants require a well-coordinated as well as intersectoral collaboration with other Public and Private sectors. Provision of these services is the key component towards prevention of diseases prevalent in the country, e.g. cholera, typhoid, dysentery and diarrhoea. Moreover, this sub-sector faces weak coordination and linkage with other sectors and stakeholders leading to duplication of efforts in tackling the existing challenges, unsatisfactory participation of the community and the private sector. (MOHCDGEC 2017: 36)

While it is not clear what data the Ministry of Water and Irrigation (MoWI) collects

because the initial data mapping focused on health data, it is known that the MoWI has published 19 datasets on the Tanzania Open Data Portal.

Increased data sharing between MoH and MoWI either through formal data sharing routines and the publication of open data could improve coordination to ensure the provision of potable water to those communities most affected by cholera outbreaks.

Education, training and communication

According to a Unicef country report on Tanzania (Unicef 2016), selected local government authorities enhanced their capacity to plan and implement sustainable water, sanitation and hygiene services. Unicef supported the provision of water, sanitation and hygiene (WASH) services in 54 schools, allowing 64,300 children to access improved and inclusive WASH services. Unicef also supported training for 216 teachers on improved WASH practices and teachers were encouraged to set up school WASH clubs. A total of 54 school WASH clubs were established as a peer-to-peer strategy for hygiene promotion. Using the community-led total sanitation approach, 48,600 households from 600 sub-villages in six districts of Mbeya, Mbarali, Iringa, Mufindi, Njombe and Makete were "triggered".

56 regional, district and ward health officers from five cholera hotspot regions (Mbeya, Rukwa, Katavi, Njombe and Iringa) were also trained in the use of social mobilisation strategies. The capacity built, and the stock of educational materials developed, will, according to Unicef, contribute to building the resilience of regions and local governments for future outbreaks (Unicef 2016).

Unicef also reports that in 2016, "some 136,177 unique mobile phone subscribers accessed the 'Internet of Good Things' at no cost, receiving mobile-ready resources and applications on topics and issues such as maternal health, hygiene, emergency information on diseases such as yellow fever and cholera, HIV and sexual health advice for adolescents, child online protection, and positive parenting techniques."

The mHealth Tanzania Public-Private Partnership (PPP) is a partnership between the CDC Tanzania, MoH, health NGOs and private sector companies, including mobile network operators. The partnership seeks to leverage economies of scale and pool expertise and resources from across the partners and to harness mobile phone technology for sustainable health programmes. The partnership's main goal is to support the implementation of Healthy Pregnancy, Healthy Baby Text Messaging Service (HPHB) (or *Wazazi Nipendeni*), which offers free and safe antenatal, motherhood and infant healthcare information to Tanzanians. However, the project is expanding to other health areas.

Amref Health Africa is implementing a project in Arusha, Kilimanjaro, Kagera, Kigoma, Katavi and Mbeya which seeks to train community health workers in lay case definitions of epidemic-prone diseases and events for accurate and timely notification and reporting.

Available data on water quality and the incidence of cholera cases could form a

valuable component of education, training and communication campaigns. Time series data combined with targets and incentives, can be powerful tools to motivate children, teachers and health workers to change behaviour patterns.

Community monitoring

As the research presented in this report has attempted to demonstrate, increased coordination through data sharing at the national level cannot be assumed to lead to the provision of services at the hyperlocal level to those communities most in need. Monitoring the completion and quality of water management infrastructure projects using contract data as well as onsite monitoring could be one approach to improving infrastructure in the district. The Tanzania Media Foundation (TMF) trains journalists in investigative data journalism with a strong focus on the use of contract data for public sector monitoring and accountability. It could play an important role in catalysing data use for monitoring purposes.

Community monitoring can also be a valuable source of new, relevant data. As many of the solutions above illustrate, useful data may either not exist or, if it does it exist, it is inaccessible. One approach to circumvent the challenges of tightly controlled government data is for communities to collect their own data. The purpose of such citizen-based monitoring approaches is to better inform communities about conditions in their immediate environment and to provide them with evidence needed to petition public institutions for improvements in service delivery.

An example of citizen-based monitoring in Tanzania is the *Sauti za Wananchi* initiative implemented by Twaweza. *Sauti za Wananchi* monitors service delivery in the education, health, and water and sanitation sectors. The initiative uses mobile phones to gather citizen perspectives about service delivery. Once survey data has been collected and analysed, direct advocacy and media publicity is used to publicise results, promote national policy messages and generate national dialogue on key issues. Twaweza has developed relationships with government departments making it easier to leverage results to advocate for policy change (ECOSOC 2015).

An independent assessment of the programme's effectiveness shows that the results from the initiative have gained traction with parliament, media and civil society. Twaweza does, however, reports concern about the lack of government responsiveness to the monitoring results which, in turn, causes apathy and lack of trust in the process among citizens who are eager to see results (ECOSOC 2015). One would hope that equivalent levels of apathy would not be encountered at the local government level given the paucity of data at that level and the proximity of this level of governance to communities.

Tancred et al. (2018), in a project that sought to address maternal and neonatal health problems faced by local communities through community-centred quality monitoring, show that village volunteers had the capacity to learn and apply techniques to address local problems in rural Tanzania. Despite data challenges, there was some evidence of changing social norms around maternal and neonatal health.

In the case of Kyela District, citizens could elect to monitor the quality of water delivered by the officially registered community-owned water supply organisations (COWSOs). COWSOs deal with the management and operational maintenance of water supply projects in rural areas and are accountable to the Ministry of Water and Irrigation. According to 2013/14 data published by the Ministry, there were four such COWSOs in Kyela District at the time.

The Construction Sector Transparency initiative (CoST) monitors public infrastructure projects by collecting data from awarded contracts and on subsequent project expenditure and delivery. CoST Tanzania undertook the monitoring of four public infrastructure projects in 2013. It is not clear whether CoST is still active in the country. If it is, the community could cooperate with CoST to monitor infrastructure projects in the District, particularly those related to water management. If the initiative is no longer active, the community could still undertake monitoring activities and may benefit from the methods developed by the CoST initiative.

Intermediation

Data is a single contributor in the array interconnected components needed to solve complex social problems. Any solution is likely to consist of several components that will need to work symbiotically to catalyse positive change. Making data useful for hyperlocal communities will, in addition to resources, incentives and enabling regulatory environments, require the intervention of one or more intermediaries who act on the data in the interest of the community (Van Schalkwyk et al. 2016).

To illustrate: Data relevant to the reduction of cholera in Kyela includes data that describes the availability of rehydration therapy kits supplied by the Medical Stores Department via its Mbeya Unit to health facilities in the district. The community may already know when cholera outbreaks are most frequent but may lack the skills or facilities to track the supply of oral rehydration kits at both health facilities and at the zonal medical store. An intermediary organisation such as a health CSO may have the facilities and skills to interpret the data, and could work with the community to track the supply of rehydration kits. However, the CSO may not have access to the data. The data is published online by the MoH but only quarterly and for the entire District. The CSO may therefore need to work with another intermediary who possesses the social capital to gain access to the medicine stocks data at a granular level (Linnander et al. 2017). But even if these two intermediaries are able to access the data in order to track and monitor effectively the availability of rehydration kits ahead of the rainy season, they may lack the ability to influence anyone in a position of authority at the Medical Stores Department (MSD) (both national and zonal) to ensure the re-stocking of hydration kits. Or they may lack the knowledge that the WHO's Procurement Services supplies cholera kits through its regional offices or that Unicef is supporting the promotion and distribution of aqua tabs for household level water treatment and safe storage in affected households (WHO 2018). They may therefore need to bring in another intermediary with personal or political connections to advocate effectively for the re-stocking of oral rehydration kits at critical times.

The Data Zetu project was designed on the understanding that if data was to lead to impact, intermediaries will play an important role once community pain points had been identified. Data Zetu uses the term "custodian organisation" to describe the intermediation role. The project seeks to engage only a single custodian organisation for each of the community pain points. This approach of linking intermediaries to specific pain points acknowledges the domain expertise (and possible sector-specific connections) required by the custodian organisations if they are to successfully deploy data for impact. That Data Zetu did not consider multiple intermediaries, may be based on the view that custodian organisations will facilitate and coordinate connections with other partner organisations or individuals if required, or could simply be explained by the limited capacity and resources available to the project. Either way, the possible solutions presented above show the many possible intermediaries or custodian organisations that could play a vital role not only in connecting hyperlocal communities to data, but in ensuring that relevant data is used to effect positive change at the hyperlocal level.

Conclusion

This section has examined how existing data that flows through the Tanzania health data system may be both relevant and useful to local communities in solving a problem identified by those communities during the Data Zetu listening campaigns. It was found that most data still reside at the national or supranational levels. What data there is at the local level (i.e. data held by health facilities themselves as well as aggregated data held by the District Medical Officer), remains inaccessible. Findings also showed that the value of data in relation to the needs of local communities comes more sharply into focus when those needs have been clearly identified.

In terms of the specific problem identified by communities in Kyela District – i.e. outbreaks of cholera – it was established that relevant data is collected at health facilities as part of the routine MTUHA reporting system as well as by means of a real-time, digitised reporting system during times of outbreaks. However, health facilities, NGOs and communities are unable to access this data to inform decisions and strategies in response to cholera outbreaks or in efforts to prevent future outbreaks. Health facilities can access their own reported data but this was found to result in individual rather than collective responses to the crisis. Where data are made publicly accessible by the national ministry (on two data portals), relevant cholera data is not available among health indicators that favour existing vertical programmes on HIV/Aids, TB, malaria and others.

Finally, this section has attempted to show that while there is a dearth of data at the local level, there is relevant and usable data elsewhere in the data ecosystem and there are opportunities for new or improved data collection efforts at village and ward level. Either way, gaining and leveraging insights gained from the data will require co-ordinated efforts by multiple actors to mobilise actionable solutions that will improve the lives of local communities.

Concluding comments

The broad question that this research posed was: 'How does data flow in the Tanzanian data ecosystem?' To answer this question, a subset of questions was formulated:

1. What are the available datasets in the ecosystem?
2. How are the datasets arranged in relation to one another in the ecosystem?
3. How does data flow between datasets?
4. Where are the breaks and blockages in the flow of data?
5. Are the needs of the community connected to the available data?

The project sought to answer these questions by (1) providing both a visual and descriptive account of a sector-specific data ecosystem, and (2) linking the data in the ecosystem to the needs of communities at the hyperlocal level.

It was found that the preponderance of health datasets is located at the supranational level and that there are limited health data at the community or district level. There is evidence of improved coordination of data systems as expressed in the horizontal flows of data between projects and departments within the MoH, but a lack of incentives is hampering the quality of data in the system. Flows are most often outwards from local health surveys and sentinel sites to the supranational level, rather than being multi-directional so as to include flows to the national, regional or district levels. Blockages in the flows of data remain: there is no horizontal access to data at the health facility level; and there are no direct community feedback mechanisms or connections to extra-governmental survey and CSO data that could provide useful in supplementing administrative, commodity and patient data collected from health facilities.

A data system with too many breaks and terminations also limits the possibility of improving the relevance of health data for hyperlocal communities. Strong upward flows feeding national-level indicators which are then used at the regional and local levels for planning and response purposes, lack relevance at the hyperlocal level, and more so if local communities are unable to feed into the system data and information directly related to their day-to-day challenges. As Aiko (2018: 22) concludes: "District Strategic Plans and the Ward Development Plans may also inform the types of metrics for which data needs to be collected. Due to limited resources for statistical activities in the districts, however, the data may not always be collected or made publicly available. Instead, most of the time the data that is collected is limited to what is prescribed in the existing MIS … and [in] community-based data collection forms and templates developed by ministries, of which the districts may not have room of adapting in any way to accommodate local data needs."

Following the identification of a specific health-related problem faced by the community in the Kyela District, several relevant datasets were identified from within a sample of health-related datasets. As in the case of the broader health dataset

mapping, most datasets were located at either the national or supranational levels. The identification of a specific problem facing a local community revealed relevant datasets previously overlooked, indicating the value of problem articulation in the identification of relevant and potentially useful data.

The usefulness of the datasets identified was tested by proposing a series of potential solutions and it was found that while availability and accessibility to useful data remain a major challenge, there are opportunities to leverage existing datasets for positive change. However, any response to the problem is likely to require the use of multiple datasets, the sourcing of new data and the intervention of multiple intermediaries which may be in scarce supply at the hyperlocal level.

To return to the idea of the space of flows being disconnected from the space of places, the findings show sufficient blockages and unidirectional, outward flows to support the suggested systemic cleavage between hyperlocal communities and globally networked actors. The findings highlight the need for network switchers, that is, those who are able to connect the space of flows with the space of places, and the need for network (re)programmers who are able to challenge the programs of existing global networks as new data flows are activated.

References

Andrason A & Van Schalkwyk F (2017) Opportune niches in data ecosystems. *SSRN*.

Bentley CM & Chib A (2016) The impact of open development initiatives in lower- and middle-income countries: A review of the literature. *EJISDC* 74(6): 1-20.

Berdou E & Shutt C (2017) *Shifting the Spotlight: Understanding Crowdsourcing Intermediaries in Transparency and Accountability Initiatives. Making All Voices Count Research Report*. Brighton: Institute of Development Studies.

Bhatia V, Stout S, Baldwin B & Homer D (2016) *Results Data Initiative: Findings from Tanzania*. Washington DC: Development Gateway.

Boerma, T (2013) Public health information needs in districts. *BMC Health Services Research* (Suppl 2):S12. http://www.biomedcentral.com/1472-6963/13/S2/S12

Boex J, Fuller L & Malik A (2015) *Decentralized Local Health Services in Tanzania: Are Health Resources Reaching Primary Health Facilities, or Are They Getting Stuck at the District Level?* Washington DC: Urban Institute.

Canares M, (2016) Open Data and Subnational Governments: Lessons from Developing Countries. *Journal of Community Informatics* 12(2)

Castells M (2017) Afterword 2017. In J Muller, N Cloete & F van Schalkwyk (eds), *Castells in Africa: Universities and development*. Cape Town: African Minds.

Castells M (2010) *The Rise of the Network Society. The Information Age: Economy, society & culture*. Volume 1 (revised edition). Oxford: Blackwell

Castells M (2009) *Communication Power*. Oxford: Oxford University Press

Castells M (1998) *End of Millennium. The Information Age: Economy, society & culture*. Volume 3. Oxford: Blackwell

Castells M & Himanen P (2014) Models of Development in Global Information Age: Constructing an Analytical Framework. In: M Castells & P Himanen (eds), *Reconceptualising Development in the Global Information Age*. Oxford: Oxford University Press. pp. 7-25.

Data Zetu (2017a, September) Listening campaign guide. https://docs.google.com/document/d/1lpSr24DrO50bvU1k6qNp5wIPhQ2mTRH0fYOwmNPC7tI/edit

Data Zetu (2017b, 12 October 2017) Pain Points Prioritization Guide: Data Zetu, Kyela. https://docs.google.com/document/d/1GBm5KQVtWzPmqaqdcP3h6CLxDOy_P-5QAr8rHFaxKHw/edit?ts=59e811f3#heading=h.gjdgxs

ECOSOC (2015) Citizen-based Monitoring of Development Cooperation to Support Implementation of the 2030 Agenda. *2016 Development Cooperation Forum Policy Briefs* No. 9. New York: Development Cooperation Policy Branch, ECOSOC.

Global Integrity (n.d.) Fiscal Governance Treasure Hunts: Kickstarting and accelerating data-driven learning cycles. Unpublished document. Washington DC: Global Integrity.

Harrison TM, Pardo TA & Cook M (2012) Creating Open Government Ecosystems: A Research and Development Agenda. *Future Internet* 4(4):900-928.

Heimstädt M, Saunderson F & Heath T (2014). From Toddler to Teen: Growth of an Open Data Ecosystem. Journal of eDemoracy and Open Government 6(2). DOI: https://doi.org/10.29379/jedem.v6i2.330

Helbig N, Cresswell AM, Burke GB & Luna-Reyes L (2012) *The Dynamics of Opening Government Data: A White Paper*. Center for Technology in Government, University at Albany.

Hicks C (2014, 3 November) Wateraid in Tanzania: The babies who die for want of clean water. *The Telegraph*. https://www.telegraph.co.uk/news/health/children/11198148/Wateraid-in-Tanzania-The-babies-who-die-for-want-of-clean-water.html

Ikonje A (2014) Strengthening data management and use in decision making to improve health care services: Lessons learnt. Dissemination Workshop Report. Tibu Homa Project for the United States Agency for International Development. Mwanza, Tanzania: University Research Company.

Kilewo EG & Frumence G (2015) Factors that hinder community participation in developing and implementing comprehensive council health plans in Manyoni District, Tanzania. *Global Health Action* 8(1): 26461. doi: 10.3402/gha.v8.26461

Kimunai E (n.d.) Availability of health data. Blog: My One Year of Service – Dodoma, Tanzania.

https://myoneyearofservice.wordpress.com/2014/09/16/availability-of-health-data/

Linnander E, Yuan CT, Ahmed S, Cherlin E, Talbert-Slagle K & Curry LA (2017) Process evaluation of knowledge transfer across industries: Leveraging Coca-Cola's supply chain expertise for medicine availability in Tanzania. *PLoS ONE* 12(11): e0186832. https://doi.org/10.1371/journal.pone.0186832

Maluka S, Kamuzora R, San Sebastian M, Byskov J, Ndawi B & Hurtig A-K (2010) Improving district level health planning and priority setting in Tanzania through implementing accountability for reasonableness framework: Perceptions of stakeholders. BMC Health Services Research 10(322):1-13. doi:10.1186/1472-6963-10-322

Mori AT, Kaale EA, Ngalesoni F, Norheim OF, Robberstad B (2014) The Role of Evidence in the Decision-Making Process of Selecting Essential Medicines in Developing Countries: The Case of Tanzania. *PLoS ONE* 9(1): e84824. doi:10.1371/journal.pone.0084824

Ministry of Health and Social Welfare, United Republic of Tanzania (2015, October) *National Strategic Plan V for Tuberculosis and Leprosy Programme 2015–2020*. Dar es Salaam: Ministry of Health and Social Welfare.

Ministry of Health and Social Welfare, United Republic of Tanzania (2015, August) *Health Sector Strategic Plan July 2015–June 2020 (HSSP IV)*. Dar es Salaam: Ministry of Health and Social Welfare.

Ministry of Health and Social Welfare, United Republic of Tanzania (2013, November) *HIV/AIDS/STI Surveillance Report No. 23*. Dar es Salaam: National AIDS Control Programme.

Ministry of Health and Social Welfare, United Republic of Tanzania (2011, July) *Comprehensive Council Health Planning Guidelines*. Dar es Salaam: Ministry of Health and Social Welfare.

Ministry of Health and Social Welfare, United Republic of Tanzania (2008) *Mapping of partners and financial flows in the medicines procurement and supply management system in Tanzania*. Dar es Salaam: Ministry of Health and Social Welfare.

Ministry of Health, Community Development, Gender, Elderly and Children (MOHCDGEC), United Republic of Tanzania (2017, October) *The National Health Policy (Sixth Draft Version)*. Dar es Salaam: Ministry of Health, Community Development, Gender, Elderly and Children.

Ministry of Health, Community Development, Gender, Elderly and Children (MOHCDGEC), United Republic of Tanzania (2016, June) *The National Road Map Strategic Plan to Improve Reproductive, Maternal, Newborn, Child & Adolescent Health in Tanzania (2016-2020): One Plan II*. Dar es Salaam: Ministry of Health, Community Development, Gender, Elderly and Children.

Ministry of Health, Community Development, Gender, Elderly and Children (MOHCDGEC), United Republic of Tanzania (2016, November) *National Guidelines for Health Data Quality Assessment*. Dar es Salaam: Ministry of Health, Community Development, Gender, Elderly and Children.

Ministry of Health, Community Development, Gender, Elderly and Children (MOHCDGEC), United Republic of Tanzania (2016) *HIV/AIDS/STI Surveillance Report No. 24: Tanzania Mainland 2013-2014*. Dar es Salaam: National AIDS Control Programme.

Ministry of Health and Social Welfare, United Republic of Tanzania (2009) *Proposal to Strengthen Health Information System [HIS]*. Dar es Salaam: Ministry of Health and Social Welfare.

Mohamed H, Brown J, Njee RM, Clasen T, Malebo HM & Mbuligwe S (2015) Point-of-use chlorination of turbid water: results from a field study in Tanzania. *Journal of Water and Health* 13(2): 544-553. DOI: 10.2166/wh.2014.001.

National Aids Control Programme, Ministry of Health and Social Welfare, United Republic of Tanzania (2017, January) *National Guidelines on HIV and Aids Data Management*. Dar es Salaam: Ministry of Health and Social Welfare.

NTLP (National Tuberculosis & Leprosy Programme), Ministry of Health, Community Development, Gender, Elderly and Children (2017, 27 December) NTLP goes more digital. NTLP. https://ntlp.go.tz/posts/ntlp-goes-more-digital/

Path (2017) *Journey to Better Data for Better Health in Tanzania 2017-2023*. Data Use Partnership.

Piovesan (2017) Beyond standards and regulations: Obstacles to local open government data initiatives in Italy and France. In F van Schalkwyk, S Verhulst, J Walker, G Magaelhes & J Pane (eds), *Social Dynamics of Open Data*. Cape Town: African Minds.

Ravallion, M (2016) 'Are the world's poorest being left behind?', *Journal of Economic Growth* 21(2): 139-164.

Rumisha SF, Mboera LEG, Senkoro KP, Gueye D & Mmbuji PK (2007) Monitoring and evaluation of Integrated Disease Surveillance and Response in selected districts in Tanania. Tanzania Health Research Bulletin 9(1): 1-11.

Sato M (2018, 28 March). Personal communication.

Sato M, Maufi D, Mwingira UJ, Leshabari MT, Ohnishi M & Honda S (2017) Measuring three aspects of motivation among health workers at primary level health facilities in rural Tanzania. *PLoS ONE* 12(5): e0176973. https://doi.org/10.1371/journal.pone.0176973

Scrollini F (2017) Open your data and will 'they' build it? A case of open data co-production in health service delivery. In F van Schalkwyk, S Verhulst, Magalhaes G, Pane J & Walker J, *The Social Dynamics of Open Data*. Cape Town: African Minds.

Stalder F (2006) *Manuel Castells: The theory of the network society*. Cambridge: Polity

Tancred T, Mandu R, Hanson C, Okuga M, Manzi F, Peterson S, Schellenberg J, Waiswa P, Marchant T & The EQUIP Study Team (2018) How people-centred health systems can reach the grassroots: experiences implementing community-level quality improvement in rural Tanzania and Uganda. *Health Policy and Planning* 33(1): e1–e13. https://doi.org/10.1093/heapol/czu070

Tanzania Invest (2018, 27 April) Tanzania First Chlorine Plant USD 112 million. https://www.tanzaniainvest.com/construction/chlorine-plant-construction

Tunga M & Mushi J (2016) *Towards engaging civil society organisations (CSOs) in the open data agenda: A case study of selected CSOs in the health sector in Tanzania*. Washington DC: World Wide Web Foundation

Twaweza (2017) Clean and Safe? Water, Sanitation and Hygiene. *Sauti za Wananchi Brief No. 40*. https://www.twaweza.org/.../SzW-TZ-2017-Water-EN-FINAL-web.pdf

Unicef (2016) Unicef Annual Report 2016: Tanzania. Unicef. https://www.unicef.org/about/annualreport/files/Tanzania_(United_Republic_of)_2016_COAR.pdf

United Nations Economic Commission for Africa (ECA) (2017) Africa Data Revolution Report 2016. Addis Ababa: United Nations Economic Commission for Africa.

United Republic of Tanzania (2016) Tanzania Digital Health Investment Roadmap 2017–2023. Dar es Salaam: Government of the United Republic of Tanzania. https://www.healthdatacollaborative.org/fileadmin/uploads/hdc/Documents/Country_documents/Tanzania_Digital_Health_Investment_Road_Map.2017_to_2023.pdf

United Republic of Tanzania, President's Office, Regional Administration and Local Government (April 2018) Administrative Area Registry System (AReS) Requirements Document. In Request for Proposal #2018-016 Development of Administrative Area Registry System (AReS). PATH.

USAID (2015) *eLMIS Selection Guide*. Arlington, VA: USAID Deliver Project.

Van Schalkwyk F & Cañares M (2019) Open government data for inclusive development. In ML Smith & RK Seward (eds), *Making Open Development Inclusive*. MIT Press, Cambridge MA.

Van Schalkwyk F, Canares M, Chattapadhyay S & Andrason A (2016a) Open Data Intermediaries in Developing Countries. *Journal of Community Informatics* 12(2): 9-25

Van Schalkwyk F, Verhulst S & Young A (2017) South Africa's Medicine Price Registry. In S Verhulst & A Young, *Open Data in Developing Economies: Toward Building an Evidence Base on What Works and How*. Cape Town: African Minds.

Van Schalkwyk F, Willmers M & McNaughton M (2016b) Viscous Open Data: The Roles of Intermediaries in an Open Data Ecosystem. *Journal of Information Technology for Development*, 22: Sup 1: 68-83.

Verhulst S, Noveck B, Caplan R, Brown K & Paz C (2014) The Open Data Era in Health and Social Care. NYLS Legal Studies Research Paper No. 2563788. https://ssrn.com/abstract=2563788 or http://dx.doi.org/10.2139/ssrn.2563788

Verhulst S & Young A (2017) *The Impact of Open Data in Developing Countries*. Cape Town: African Minds.

Wickremasinghe D, Hashmi IE, Schellenberg J & Avan BI (2016) District decision-making for health in low income settings: a systematic literature review. *Health Policy and Planning* 31: ii2-ii24

World Health Organization (2018, 12 January) Cholera – United Republic of Tanzania. Disease Outbreak News. http://www.who.int/csr/don/12-january-2018-cholera-tanzania/en/

World Health Organization (2016, 22 April) Cholera – United Republic of Tanzania. Disease Outbreak News. http://www.who.int/csr/don/22-april-2016-cholera-tanzania/en/

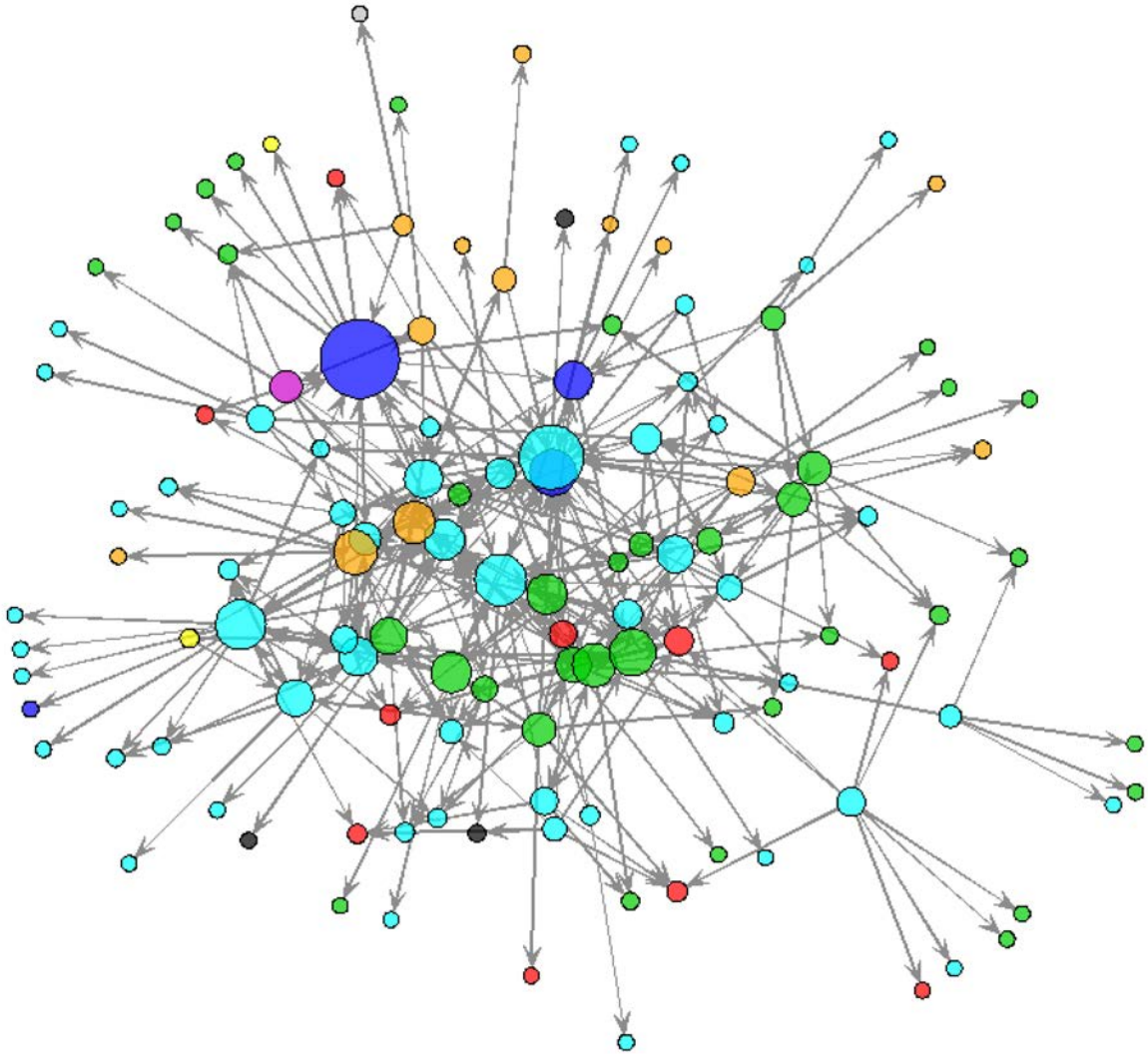
World Health Organization (n.d.) Tanzania Country Profile: Health information, research, evidence and knowledge. http://www.aho.afro.who.int/profiles_information/index.php/Tanzania:Health_information,_research,_evidence_and_knowledge#Data_sources_and_generation

Yale University (2012) An Innovative Public-Private Partnership: Medical Supply Solutions in Tanzania Overview. http://nexus.som.yale.edu/ph-tanzania/

Key informants interviewed

- Rose Aiko, Data Ecosystems Consultant, Global Partnership for Sustainable Development Data
- Kathryn Alexander, Senior Associate, Development Gateway
- Omar Bakari, Project Co-ordinator, Data Zetu
- Taha Jiwaji, Zaid Mkwanga and Nuru Magwaza, Code for Tanzania
- Aman Grewal and Mahadia Tunga, Tanzania Data Lab (dLab)
- Leonard Mpande, dhis2 team, Department of Computer Science, University of Dar es Salaam
- Jumanne Rajabu Mtambalike and Joyce Msigwa, Sahara Sparks
- Josephine Nyongi, Programme Manager, Sikika
- Emanuel Petro, HIV/Aids Prevention Co-ordinator for the Mbeya Region; DREAMS Project co-ordinator, Mbeya
- Yunus Rafiq, Researcher / Ifakara Health Institute
- Rosemary W Silaa, Public Health Consultant, Tanzania
- Hannah Wheatley, Civicus and formerly Yale Project on Medicines Distribution in Tanzania

Appendix 1: Key stakeholder and use network for the Tanzania health system



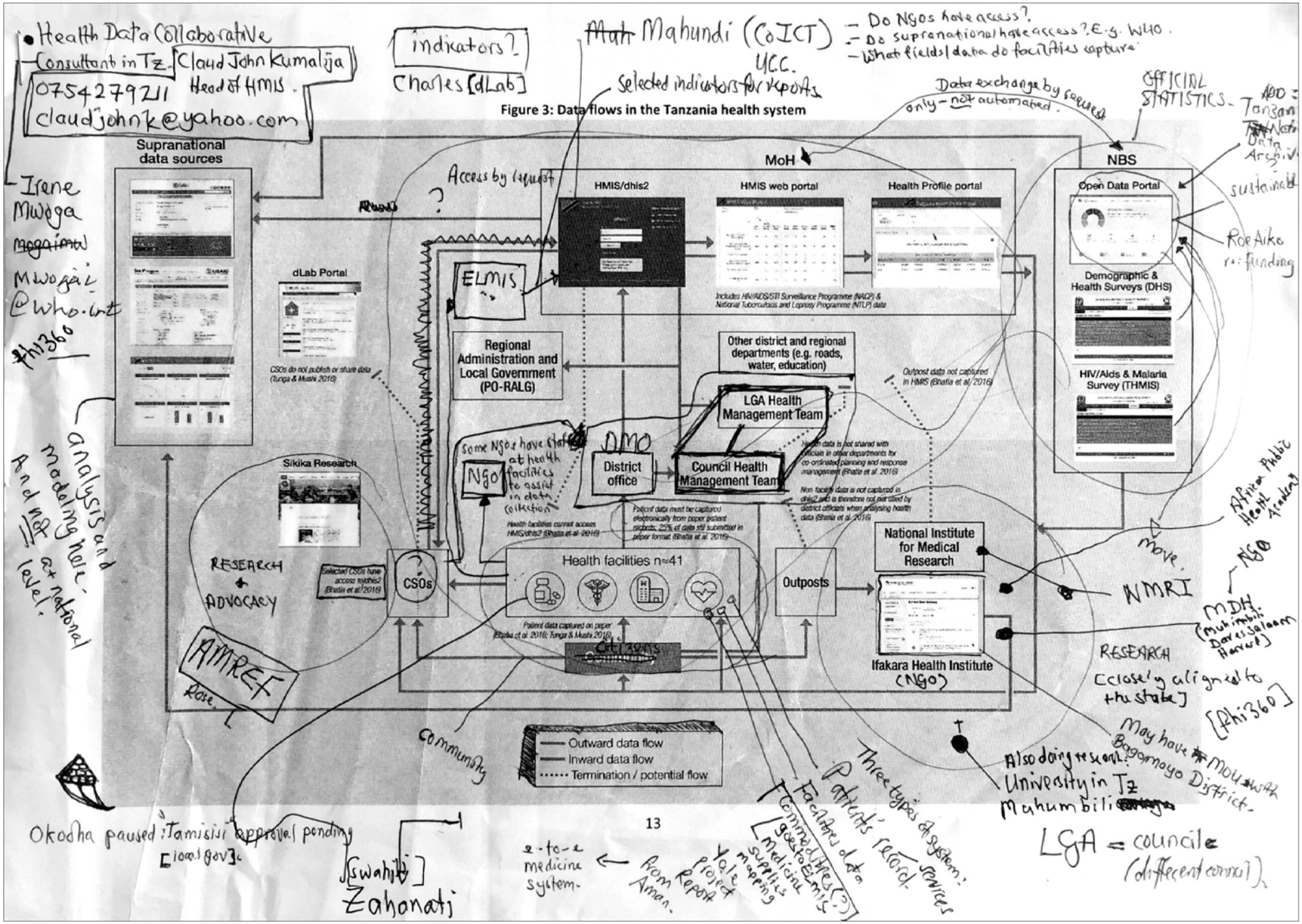
[Source: PATH 2017: 121]

MOH-District Health Plans

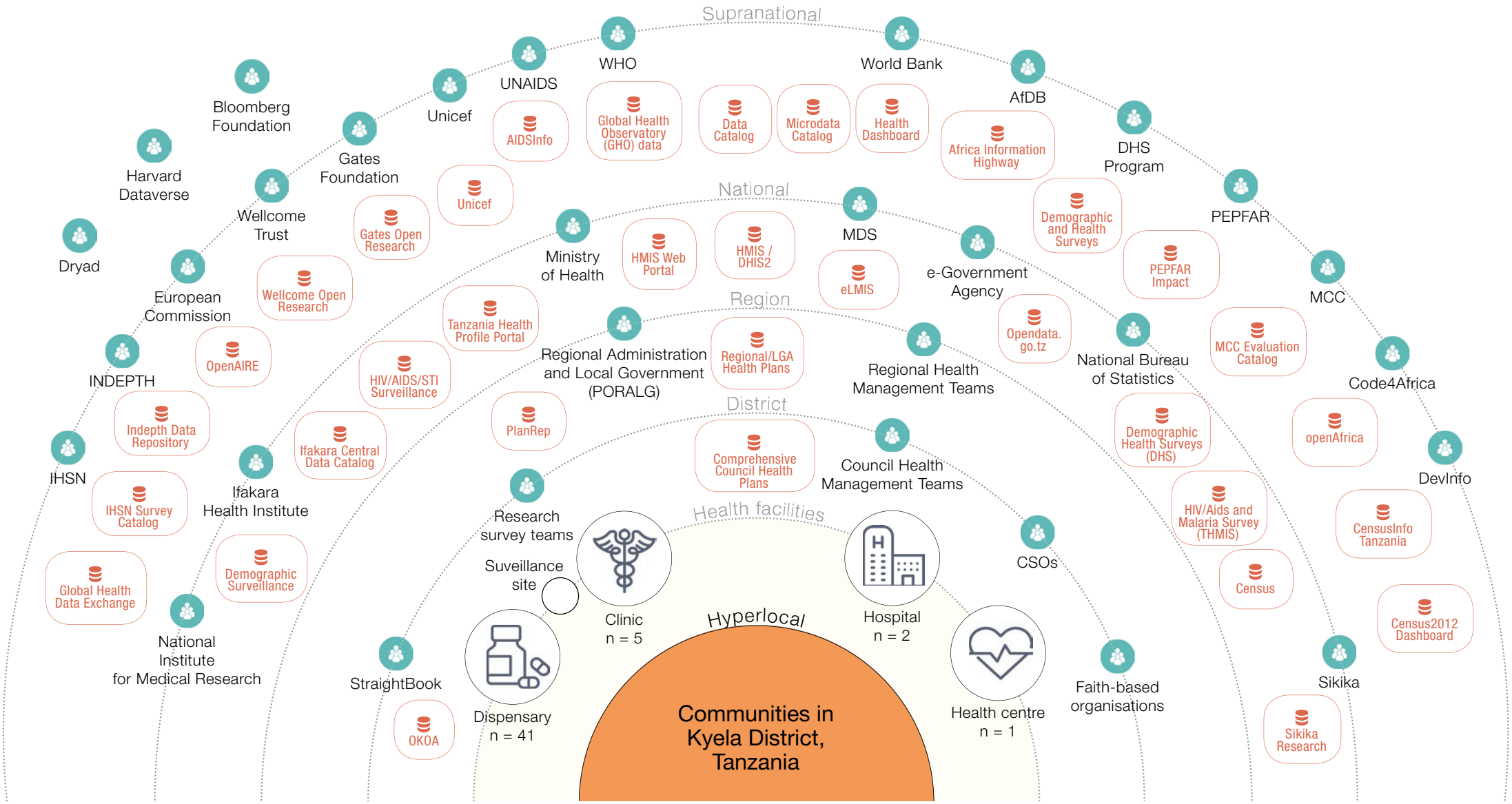
Notes:

1. Health facilities play a key, central role and have many connections throughout the network illustrated above. Some Ministry units and government agencies are centrally located in the network but others sit on the periphery with little connection to the core network. Implementing partners are dispersed throughout the network.
2. Peripheral players are connected with only one other member of the network. While part of the network, these organisations will have less influence. Two organisations are represented as yellow dots in the networks; these are regional or local government entities, and both are located at the periphery and relatively disconnected from the network.
3. The MOH-District Health Planning is equally peripheral and disconnected from other actors in the network.

Appendix 2: Example of health data flows mapping method



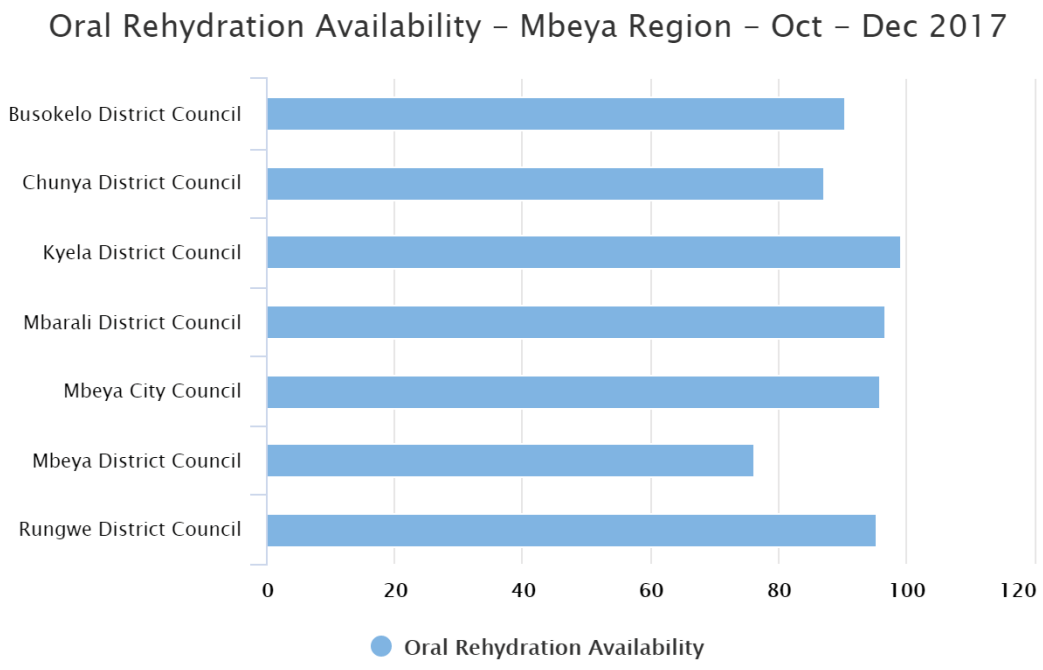
Appendix 3: Custodians and data sources in the Tanzania health system



Appendix 4: Descriptions of data relevant to water quality and cholera Kyela District, Tanzania

Source: MoH HMIS Web Portal

Prevention (water quality) or response (cholera treatment): Response
Level: National
Data: Tracer medicines
URL: https://hmisportal.moh.go.tz/hmisportal/#/indicator/tracer_medicine
Description: This is data from dhis2 published on the MoH's HMIS portal. Data is available in annual or quarterly periods at the level of district. As at 30 April 2018, the most recent data available was for the period October to December 2017. Data can be viewed in tabular, chart or map formats, and can be downloaded in Excel format. Data can be accessed and downloaded freely but is restricted to personal use.
Relevant datasets related to the monitoring of the availability of medicines required to respond to cholera outbreaks include: oral rehydration availability. According to the CDC, rehydration is the corner-stone of cholera treatment.[14]
NOTE: Antibiotics are also used in severe cases of cholera; it is possible that the database could be used to monitor the availability of antibiotics used in the treatment of cholera (e.g. erythromycin and doxycycline) if these are reported on in the HMIS database.

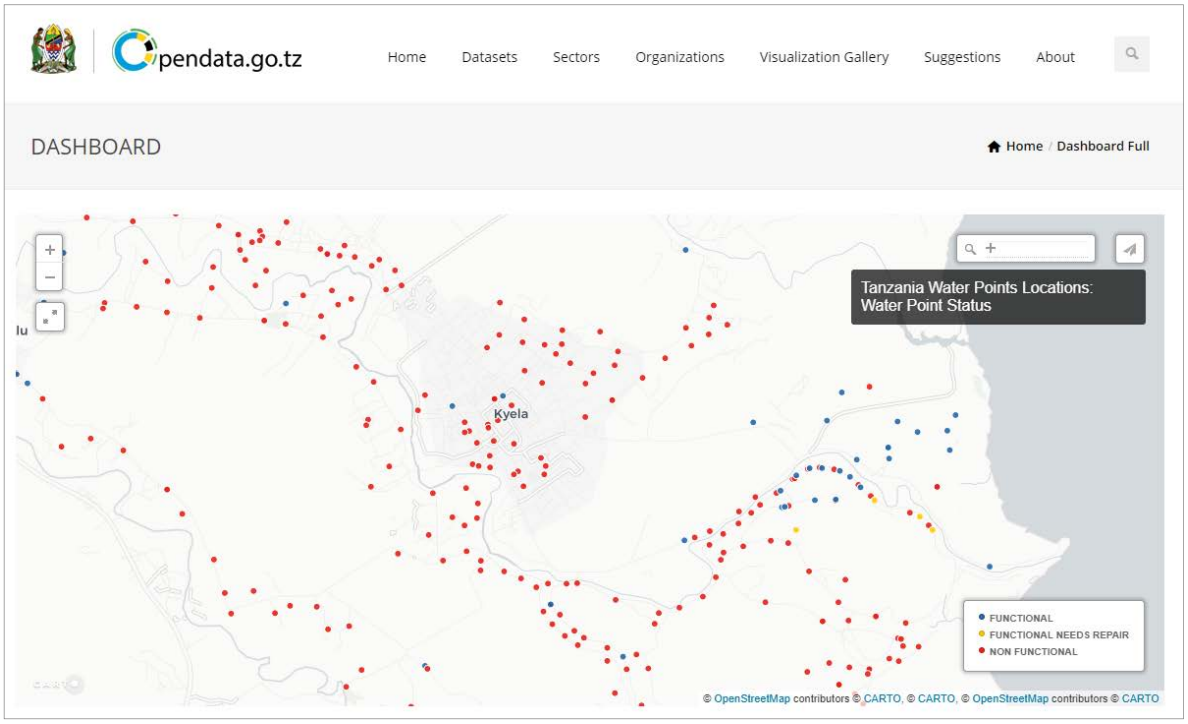


14 https://www.cdc.gov/cholera/treatment/rehydration-therapy.html

Source: Electronic Logistics Management Information System
Prevention (water quality) or response (cholera treatment): Prevention (?) & Response
Level: National
Data: medicine stock levels: chlorine tablets (?); oral rehydration; antibiotics
Data on commodities (i.e. medicines) is captured in the Electronic Logistics Management Information System (eLMIS). In Tanzania, the eLMIS collects data from more than 6 000 service delivery points. The central-level, web-based eLMIS interfaces with software used to process orders for delivery by the Medical Stores Department (MSD) in Tanzania. MSD delivers essential medicines, medical supplies and laboratory reagents to more than 5,600 health facilities (hospitals, Health centers and dispensaries) across the country. MSD zonal stores are located in Dar es Salaam, Mtwara, Mbeya, Iringa, Dodoma, Tabora, Mwanza, Moshi as well as Tanga and Muleba Sales Points. The MSD has seven community outlets at Muhimbili National Hospital in Dar Es Salaam, Sekou-Toure Hospital in Mwanza, Mbeya nearby Mbeya Regional Referral Hospital, within Mount Meru Hospital in Arusha, Chato District Hospital in Geita, Ruangwa District Hospital in Lindi and Mpanda District Hospital in Katavi. The outlets operate seven days a week and offer both generic and brand medicines to individual patients, NHIF beneficiaries, and public and registered private hospitals. eLMIS data is not publicly accessible although selected data from eLMIS is published on the HMIS Web Portal (see above).

Source: Tanzania Open Data Portal
Prevention (water quality) or response (cholera treatment): Prevention
Level: National
Data: Water point status
URL: http://opendata.go.tz/dataset/hali-halisi-ya-vitoa-maji-kwa-mikoa-yote-tanzania/resource/15b43e6d-9ef9-41cc-a8aa-d35ddef19c59
Description: This data shows the total number of water points, number of functional, not-functional and functional-needs repair in all Tanzania. This is data from the Ministry of Water published on the National Open Data Portal. Data is available in annual periods at the level of district. As at 30 April 2018, the most recent data available was for the period 2015-2016. Data can be viewed in tabular format and can be downloaded in csv format. Data can be accessed and downloaded freely under a royalty-free license.

Source: Tanzania Open Data Portal
Prevention (water quality) or response (cholera treatment): Prevention
Level: National
Data: Water point locations
URL: http://opendata.go.tz/en/indicator/a2fab64e-47f7-11e5-847d-0e5e07bb5d8a
Description: This data shows the total number of water points, number of functional, not-functional and functional-needs repair in all Tanzania. This is data from the Directorate of Rural Water Supply in the National Ministry of Water and Irrigation published on the National Open Data Portal. Data is available in annual periods at the level of district. No data is provided to indicate the period covered by the data. Data can be viewed on a map (from a link in the visualisation gallery).



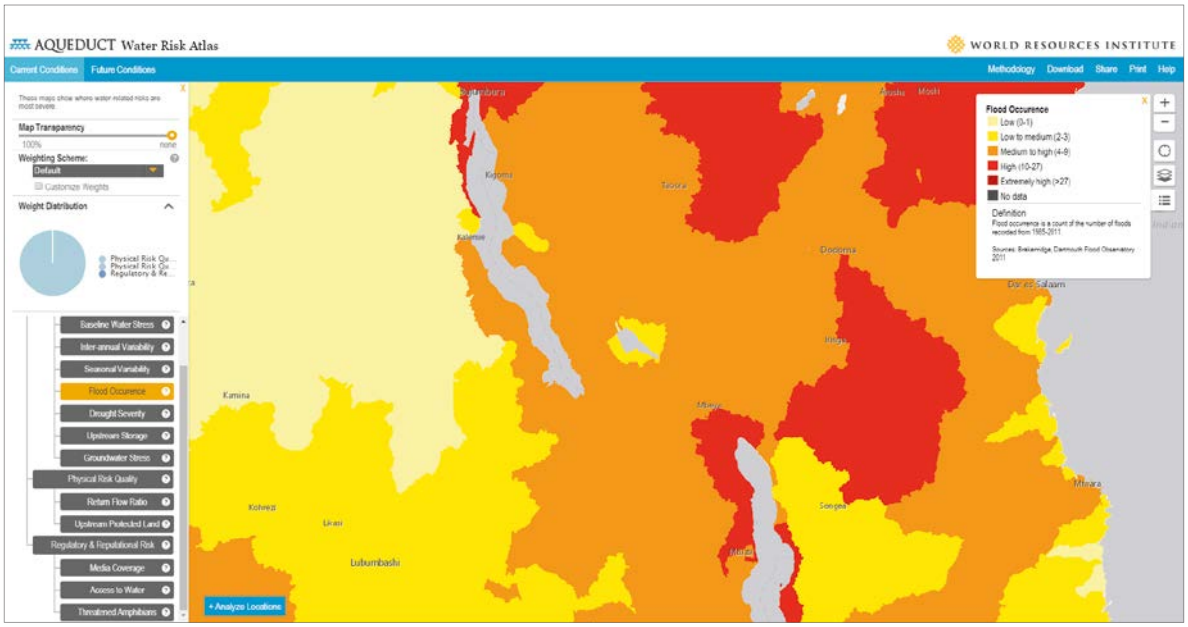
Source: DHS Program
Prevention (water quality) or response (cholera treatment): Prevention
Level: Supranational
Data: drinking water sources and treatment; sanitation; hand washing
URL: https://www.dhsprogram.com/data/dataset/Tanzania_Standard-DHS_2015.cfm?flag=0
Description: The Tanzania Demographic and Health Survey (DHS) 2015-2016, also known as the Tanzania Demographic and Health Survey and Malaria Indicator Survey (TDHS-MIS), is part of phase 7 of the Demographic and Health Survey series. Topics commonly covered in DHS include: child and maternal health, family planning, nutrition, health behaviour and knowledge, health care access and use, and immunization. For the 2015-2016 Tanzania DHS, 13,266 women and 3,514 men ages 15-49 were successfully interviewed from 12,563 households. This is data from the

DHS Program, a programme funded by USAID to provide technical assistance to more than 300 surveys in over 90 countries to advance global understanding of health and population trends in developing countries. The data was collected by the Tanzanian National Bureau of Statistics. Data is available at the national and regional levels. Data is not visualised and can be downloaded in parts in dta, dat, sas7bdat and sav formats. Registration and login is required to download the data. Dataset access is only granted for legitimate research purposes.

Source: World Health Organization Global Health Observatory
Prevention (water quality) or response (cholera treatment): Response
Level: Supranational
Data: Cholera – number of reported cases; Children aged <5 years with diarrhoea receiving oral rehydration therapy (%)
URL: http://apps.who.int/gho/data/node.country.country-TZA?lang=en
Description: This data shows the total number of reported cases of cholera per annum, as well as the proportion of children under 5 years of age with diarrhoea who received oral rehydration therapy. This is data from the World Health Organization and although the source is not indicated is most likely sourced from the Tanzanian dhis2. Data is available in annual periods at the national level only. There are many gaps in the data: only 2011 data is available for cholera incidence and only 2010 data is available for the proportion of children under 5 years of age with diarrhoea who received oral rehydration therapy. As at 30 April 2018, the last update was on 24 November 2017. No data is provided to indicate the period covered by the data. Data is presented in tabular format and can be downloaded in part or in full in csv, Excel, xml and json formats.

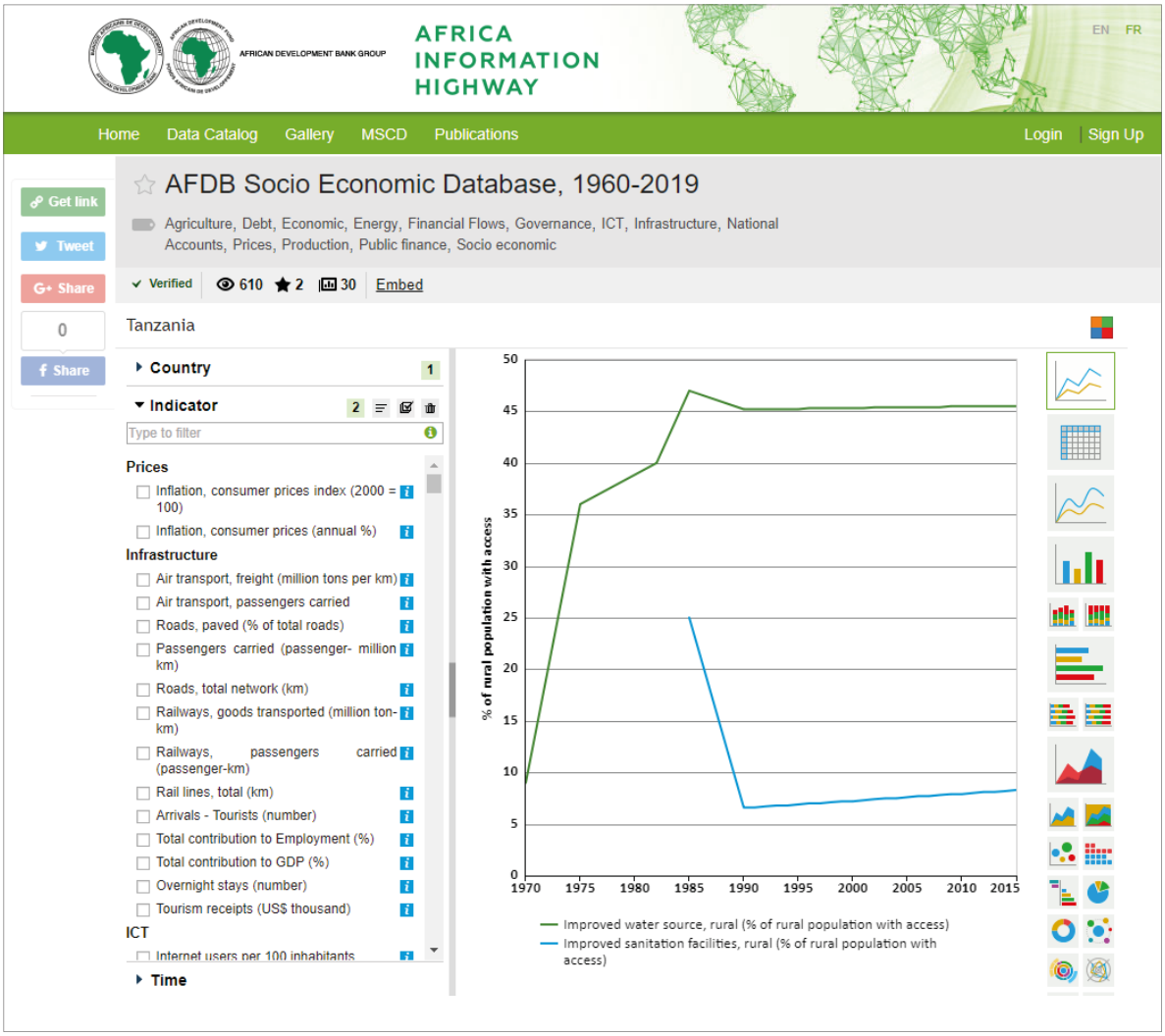
	United Republic of Tanzania													
Indicator	2015	2014	2013	2012	2011	2010	2009	2008	2007	2006	2005	2004	2003	2002
Cholera - number of reported cases					942.00									
Number of under-five deaths (thousands)	119	121	124	126	128	131	135	138	141	145	149	154	160	166

Source: World Resource Institute
Prevention (water quality) or response (cholera treatment): Prevention
Level: Supranational
Data: flood occurrence and prediction
URL: http://www.wri.org/our-work/project/aqueduct/ ; http://www.wri.org//resources/maps/aqueduct-global-flood-analyzer
Description: This data shows current and predicted flooding. This is data from the World Resource Institute collected and modelled using the extant literature and data on global water resources. Data is available in annual periods at the national level, although it is possible to view data on a map at the district level. Data is from several sources dating between 2000 and 2014. As at 30 April 2018, the last update was in April 2015. Data is presented in map format and can be downloaded as shape files or ESRI Geodatabase files.



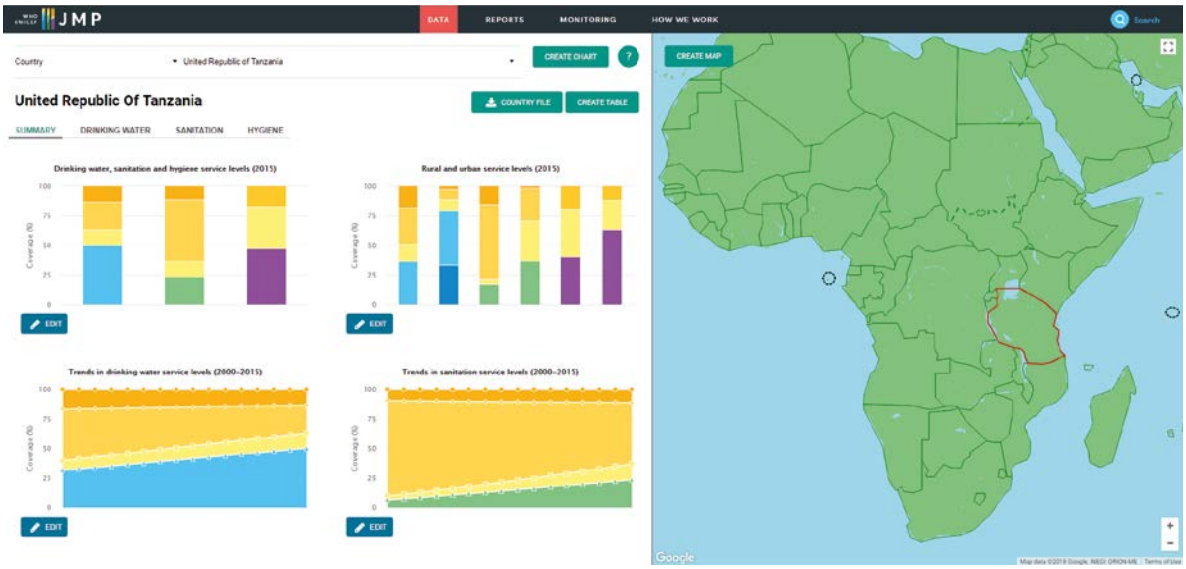
Source: World Bank Micodata Catalog
Prevention (water quality) or response (cholera treatment): Prevention
Level: Supranational
Data: sanitation
URL: http://microdata.worldbank.org/index.php/catalog/2232/study-description
Description: This data is from a study on handwashing and sanitation interventions in Tanzania conducted from 2009 to 2012 across 181 rural wards in 10 districts followed by an endline survey of 3619 households and 5 768 children under 5 years of age. This is data from a joint World Bank and Inter-American Development Bank research project funded by the Bill & Melinda Gates Foundation. Data files: household; individual; community_participation; handwashing_facilities; animals; latrine_expenses; structured_observations; under_5_mortality; listing; community. The data format is unknown are can only be accessed by registering and logging in to Microdata Library.

Source: African Development Bank Africa Information Highway
Prevention (water quality) or response (cholera treatment): Prevention
Level: Supranational
Data: improved water source, rural; improved sanitation, rural
URL: http://dataportal.opendataforafrica.org/bbkawjf/afdb-socio-economic-database-1960-2019
Description: This data shows improvements in water sources and sanitation. This is data from the African Development Bank and the source is indicated as the African Development Banks data portal which could not be accessed. Data is available in annual periods from 1970 to 2015 at the national level only. As at 30 April 2018, the last update was on 5 February 2018. Data is presented in multiple chart and tabular formats, and can be downloaded in part or in full in csv, Excel, OData and json for the raw data, and in PDF, Excel, PowerPoint or png formats for visualised data. No use license is specified.



Source: openAfrica
Prevention (water quality) or response (cholera treatment): Response
Level: Supranational
Data: cholera incidence
URL: https://africaopendata.org/dataset/cholera-outbreaks-bulletin-for-tanzania-february-2018
Description: This data shows reported cholera cases and deaths by district. The data was uploaded by Code for Tanzania. The source of the data is not specified. Data is available in weekly periods for February 2018. As at 30 April 2018, no data was available for other periods. The data was uploaded on 6 April 2018. Data is presented in tabular format and is available for download and use under a Creative Commons Attribution license in Excel format.

Source: Joint Monitoring Programme for Water Supply, Sanitation and Hygiene (JMP)
Prevention (water quality) or response (cholera treatment): Prevention
Level: Supranational
Data: drinking water; sanitation; hygiene
URL: https://data.unicef.org/wp-content/uploads/2015/12/Drinking-Water-Sanitation-Hygiene-Database-July-2017.xlsx ; https://washdata.org/data
Description: This data shows the percentage of the total population with access to dinking water, sanitation facilities (toilets), and hygiene facilities (water and soap). The data is from the WHO/Unicef Joint Monitoring Programme for Water Supply, Sanitation and Hygiene (JMP). The JMP global database includes nearly 5 000 national datasets enabling the production of estimates for over 200 countries, areas, and territories. Data is collected from household surveys and from national ministries and statistics agencies. Data is available in annual periods at the national level only, although reporting is broken down in urban and rural areas. The latest data for Tanzania is for 2015. As at 30 April 2018, the last update was in July 2017. Data is presented in the form of charts or tables and can be downloaded in full in csv format.

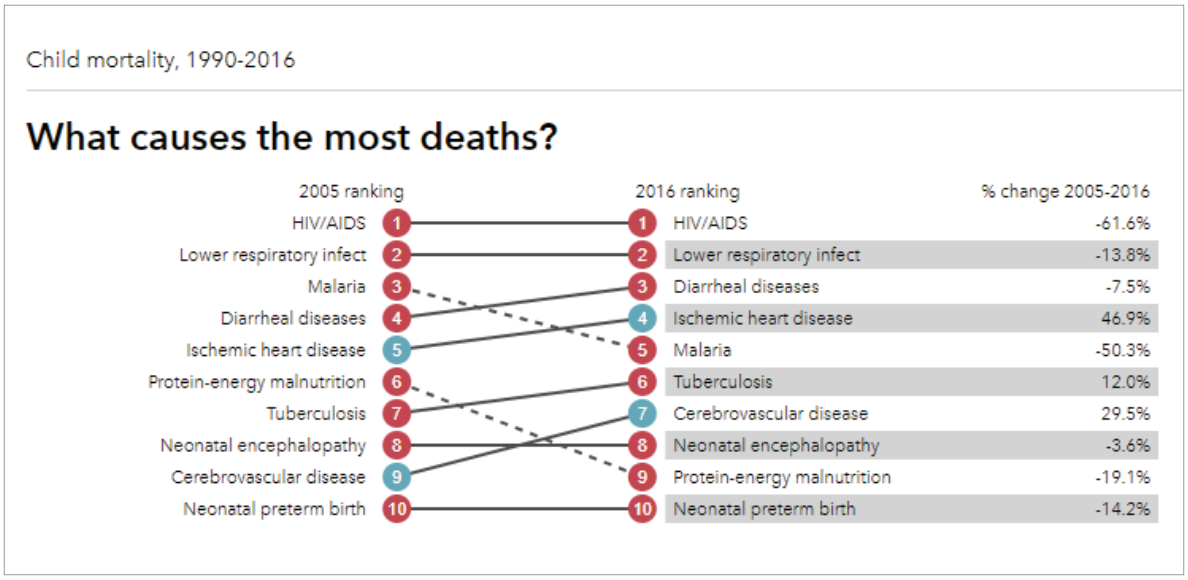


Source: Unicef
Prevention (water quality) or response (cholera treatment): Response
Level: Supranational
Data: diarrhoea treatment
URL: https://data.unicef.org/topic/child-health/diarrhoeal-disease/
Description: This data shows (1) diarrhoea treatment with oral rehydration supplement (ORS): percentage of children under age 5 who had diarrhoea in the two weeks preceding the survey and who received oral rehydration salts (ORS packets or pre-packaged ORS fluids); and (2) diarrhoea treatment with oral rehydration therapy and continued feeding: percentage of children under age 5 who had diarrhoea in the two weeks preceding the survey and who received oral rehydration therapy) and continued feeding. The data is prepared by Unicef and sourced from the national demographic health surveys and the health management information system. Data is available for the years 2004, 2010 and 2016 at the national level only, although reporting is broken down in urban and rural areas (as well as by gender, mother's educational level and wealth). As at 30 April 2018, the last update was in July 2017. Data is not presented in any form and can be downloaded in full in Excel format.

Source: WHO
Prevention (water quality) or response (cholera treatment): Response
Level: Supranational
Data: cholera outbreak reports
URL: http://www.who.int/csr/don/archive/disease/cholera/en/
Description: This data comprises reports from the WHO on cholera outbreaks. Data is therefore mainly text but does occasionally include non-machine-readable data presented in charts. The data is prepared by WHO and sourced from the ministry of health and other partners. Data is available from 1996 to the present and reports are posted on an ad hoc basis. As at 30 April 2018, the most recent report was for 29 March 2018.

Source: Global Health Data Exchange (GHDx)
URL: http://ghdx.healthdata.org/geography/tanzania
Description: As a repository, the GHDX provides access to several survey datasets already described above such as the 2015-16 Tanzania Demographic and Health Survey and Malaria Indicator Survey (2015-16 TDHS-MIS).

Source: Institute for Health Metrics and Evaluation
Prevention (water quality) or response (cholera treatment): Prevention
Level: Supranational
Data: sanitation; hygiene
URL: http://www.healthdata.org/tanzania
Description: This data is for selected health indicators displayed as charts, infographics and maps. The data is from the Institute for Health Metrics and Evaluation (IHME), an independent global health research centre at the University of Washington. Data is collected from multiple sources available from the Global Health Data Exchange. Data is available in annual periods at the national level only from 1990 to 2016.



Source: Sikika
Prevention (water quality) or response (cholera treatment): Response
Level: National
Data: Health surveys and health budget monitoring
URL: http://sikika.or.tz/index.php/publications/analytical-works
Description: Sikika conducts ad hoc surveys of health-related issues. Some of those surveys may be of relevance to the community in Kyela. An example of such a survey is the "Report on Availability of Essential Medicines, Medical Supplies and Bed Capacity in Hospitals in Tanzania Mainland" published in 2013. Sikika also monitors government's budgeting and expenditure on health care. Sikikia does not, however, publish the raw data from its surveys and analyses. The community in Kyela would need to engage with Sikika to obtain data that is both current, location-specific and relevant to the issue of the treatments of cholera. With such data, the community in Kyela could monitor and advocate for more effective health service delivery in the District.

Source: Twaweza
Prevention (water quality) or response (cholera treatment): Response
Level: National
Data: mobile phone health surveys
URL: https://www.twaweza.org/.../SzW-TZ-2017-Water-EN-FINAL-web.pdf
Description: Twaweza conducts ad hoc national-level mobile phone surveys, including on health-related issues. An example of such a survey is the 2016 Water, Sanitation and Hygiene survey conducted by its Sauti za Wananchi programme. The aggregated data from the survey are published as text and charts in a PDF report. Twaweza does not, however, publish the raw data from its surveys and analyses.

Appendix 5:
Facility visited in Kyela District by ward and type

Name of facility	Type/level of service	Ward
Bujonde	Dispensary	Bujonde
Busale	Dispensary	Busale
Ikama	Dispensary	Ikama
Ikolo	Dispensary	Ikolo
Ikombe	**Dispensary**	**Matema**
Ilopa	Dispensary	Ikama
Ipande	Dispensary	Ipande
Ipinda	Health Centre	Ipinda
Itope	Dispensary	Itope
Kajunjumele	Dispensary	Kajunjumele
Kandete	Dispensary	Ipande
Kapamisya	**Dispensary**	**Mwaya**
KatumbaSongwe	Dispensary	KatumbaSongwe
Kilasilo	Dispensary	Ikimba
Kisale	Dispensary	Ipinda
Kyela District Hospital	**Hospital**	**Kyela**
Lema	Dispensary	Busale
Lugombo	**Dispensary**	**Mwaya**
Lukwego	Dispensary	Lusungo
Lutusyo / Muungano	Dispensary	Ikolo
Makwale	Dispensary	Makwale
Mwalisi	**Dispensary**	**Ngana**
Ndobo	Dispensary	Ndobo
Ndwanga	Dispensary	KatumbaSongwe
Ngana	**Dispensary**	**Ngana**
Ngonga	Dispensary	Ngonga
Ngekye	Dispensary	Mababu
Njisi	Dispensary	Njisi
Njugilo	Dispensary	Ipande
Sinyanga	Dispensary	Nkokwa
Tenende	**Dispensary**	**Mwaya**

About the authors

François van Schalkwyk is a researcher at SBC4D. He holds a doctorate degree from Stellenbosch University, as well as masters degrees in education (University of the Western Cape) and in publishing studies (Stirling University). François has undertaken several research projects on open government data in Africa and has also contributed to the drafting of data policies on the continent.
https://www.researchgate.net/profile/Francois_Van_Schalkwyk3

Rosemary W. Silaa is a public health consultant based in Dar es Salaam, Tanzania. She holds a masters degree in business administration from the Eastern and Southern African Management Institute (ESAMI), a bachelors degree in pharmacy, and an advanced diploma in purchasing and supply. Rosemary has several years' experience working in the Tanzanian health sector across different levels of service provision as well as at the policy level.
https://www.linkedin.com/in/rosemary-silaa-98880436/