Aalto-yliopisto
Aalto-universitetet
Aalto University

Master's Programme in Water and Environmental Engineering

Monitoring and Evaluation in the Rural Water Sector in Ethiopia

Case of a water project in achieving Sustainable Development Goal 6

Eelis Hemberg

Master's Thesis
2021

Aalto-yliopisto
Aalto-universitetet
Aalto University

Author Eelis Hemberg

Title of thesis Monitoring and Evaluation in the Rural Water Sector in Ethiopia – case of a water project in achieving Sustainable Development Goal 6

Programme Water and Environmental Engineering

Thesis supervisor DSc (Tech) Marko Keskinen

Thesis advisors MSc (Tech) Anni Juvakoski and MSc (Tech) Arto Suominen

Date 19.07.2021 **Number of pages** 50 / 3 **Language** English

Abstract

Access to sustainable and safe drinking and household water is a clear precondition for tackling poverty and empowering local communities. But how can progress in these fronts be effectively measured? Definitions, nuance, and classifications are important in development cooperation, and seemingly small adjustments to these can affect the lives of millions of people. There is a trend in monitoring and evaluation (M&E) systems towards more responsive and real-time follow-up of results of interventions. This is increasingly the case after the adoption of the Sustainable Development Goals (SDGs) in 2015, which increased the ambition level of data collection as well. This study has analyzed three different strata (micro, meso and macro) covering the M&E system in the case study context of the COWASH Phase III project in Ethiopia's rural water sector. The aim of the study has been to ascertain the influence of SDG 6 (clean water and sanitation) on M&E processes, whether these processes are encouraging effective data collection and responsive approaches in project interventions, and whether the M&E systems at different strata are integrated with one another seamlessly. These research questions were investigated through seven semi-structured interviews with key informants. Indicators for rural drinking water access coverage indicators were also analysed on the three strata of analysis. A novel analytical framework and indicator correspondence analysis method were devised for the study by modifying existing methods for stakeholder analysis. Based on the results, there is evidence of the several M&E systems functioning well in isolation, but there seems to be rather little vertical integration between the different systems across the strata, with persisting needs for closer methodological harmonization. Closer attention must be placed on improving the capacity of actors at the local level in improving M&E capacity at the level where beneficiary data originates. There should be greater discussion about the ways in which the SDGs are utilized in the rural water sector in Ethiopia, and the role of local communities in their use should be better incorporated in their implementation. Adoption of novel data collection methodologies must be executed carefully, and attention should be paid to the capacity of systems to utilize data from new sources (such as citizen-reported big data through mobile phone-based technologies) prior to these types of practices being widely adopted for monitoring purposes in project interventions.

Keywords monitoring, evaluation, development cooperation, development policy, Ethiopia, Sustainable Development Goals, WASH, rural development, water sector, drinking water

Aalto-yliopisto
Aalto-universitetet
Aalto University

Tekijä Eelis Hemberg

Työn nimi Seuranta- ja arviointimenetelmät Etiopian maaseudun vesisektorilla – vesihanke tapaustutkimuksena kestävän kehityksen tavoite 6:n saavuttamisessa

Koulutusohjelma Vesi- ja ympäristötekniikan maisteriohjelma

Vastuuopettaja/valvoja TkT Marko Keskinen

Työn ohjaajat DI Anni Juvakoski, DI Arto Suominen

Päivämäärä 19.07.2021 **Sivumäärä** 50 / 3 **Kieli** englanti

Tiivistelmä

Kestävä ja turvallinen kotitalouksien vesihuolto on köyhyyden vähentämisen ja paikallisyhteisöjen voimaannuttamisen selvä edellytys. Mutta kuinka edistystä voidaan näiden osalta tehokkaasti mitata? Määritelmät, nyanssit ja kategoriat ovat kehitysyhteistyössä merkityksellisiä, ja näihin tehdyt näennäisesti pienet muutokset voivat vaikuttaa miljoonien ihmisten elämään. Seuranta- ja arviointijärjestelmien (M&E) kehityssuuntana on ollut herkempi ja reaaliaikaisempi toimenpiteiden seuranta. Tätä on edesauttanut kestävän kehityksen tavoitteiden (SDG) hyväksyminen vuonna 2015, mikä on myös lisännyt myös datajärjestelmiltä vaadittua kunnianhimoa ja vaatimustasoa. Tämä tutkimus on tarkastellut kolmea analyysitasoa (mikro, meso ja makro), jotka kattavat seuranta- ja arviointijärjestelmän COWASH Phase III-hankkeen tapaustutkimuskontekstissa Etiopian maaseudulla. Tarkoituksena on ollut selvittää, missä määrin SDG 6 (puhdas vesi ja sanitaatio) on vaikuttanut tapaustutkimuksen seuranta- ja arviointitoimenpiteisiin, tukevatko käytössä olevat toimenpiteet tehokasta datan käyttöä ja projektin kehittämistä hanketyössä, sekä onko M&E-järjestelmiä integroitu toisiinsa saumattomalla ja johdonmukaisella tavalla. Työn metodi on perustunut pääasiassa seitsemään puolistrukturoituun haastatteluun. Lisäksi on tarkasteltu indikaattorien vastaavuutta juomaveden palvelutason kattavuutta mittaavien indikaattorien osalta kaikilla kolmella analyysitasolla. Tutkimusta varten on kehitetty uudenlainen analyyttinen viitekehys sekä indikaattorien vastaavuusanalyysimetodi aikaisimpia sidosryhmäanalyysin menetelmiä mukauttaen. Tulosten perusteella on näyttöä siitä, että useat käytössä olevat M&E-järjestelmät toimivat hyvin erillään, mutta vertikaalista integraatiota eri tasojen välillä on varsin vähän ja metodologiseen harmonisointiin on yhä tarvetta. Huomiota tulee kiinnittää paikallistoimijoiden seuranta- ja arviointikapasiteetin parantamiseen paikallistasolla lähellä hyödynsaajadatan tuottamista. Kestävän kehityksen tavoitteiden hyödyntämisestä Etiopian maaseudun vesisektorilla tulee käydä laajempaa keskustelua, ja paikallisyhteisöjen rooli tulee huomioida tarkemmin niiden toteuttamisessa. Uudenlaisten datankeräysmenetelmien käyttöönotossa tulee toimia harkitusti, ja järjestelmien kykyyn hyödyntää uusia datalähteitä (kuten kansalaisten raportoimaa mobiilipohjaista massadataa) tulee kiinnittää huomiota ennen tällaisten datalähteiden käyttöönottoa projekti-interventioissa.

Avainsanat seuranta, arviointi, kehitysyhteistyö, kehityspolitiikka, Etiopia, kestävän kehityksen tavoitteet, WASH, maaseudun kehitys, vesisektori, juomavesi

Aalto-yliopisto
Aalto-universitetet
Aalto University

Contents

Acknowledgements

One of the works that most influenced me during my bachelor's studies at McGill University was Amartya Sen's *Development as Freedom*. In his book, Sen writes: "freedoms are not only the primary ends of development, [but] they are also among its principal means" (Sen 2002). Although written over two decades ago, characterizing development as a complex web of mechanisms to increase people's freedom is very fitting. Methods devised in service of development should seek to widen people's choices and agency. The indicators, tools and frameworks discussed at length herein may at times seem abstract and theoretical, but they are about empowerment and freedom for *people*. This philosophical consideration should be a humbling principle for all working in this field, and it certainly has been for me.

This has been a journey of unforeseen circumstances in challenging times, and I could not have achieved this without the many guides on my path. I am very grateful to my supervisor associate professor Marko Keskinen for his guidance and mentorship, especially during this thesis work but along my master's degree journey overall as well. I am also extremely thankful to my thesis advisors Anni Juvakoski from the Water and Development Research Group (WDRG) and Arto Suominen from Ramboll Group, who have provided valuable insights into various aspects of this work. This thesis would not have been possible without financial support from the Maa- ja vesitekniikan tuki ry foundation and the Water and Development Research Group at Aalto University School of Engineering. Special thanks to Abebaw Getachew and the COWASH project staff, experts at the Ministry for Water, Irrigation and Energy of Ethiopia, Maria Kruskopf from Niras Consulting, and WASH expert Antti Rautavaara, who have assisted greatly in formulating this study. Lastly, my interest in Ethiopia was initially kindled by associate professor Jon Unruh at McGill University, whose fascinating lectures during my undergraduate studies have stuck with me ever since.

I dedicate this work to my parents, who have always been my greatest champions and role models – thank you for inspiring my passion towards Africa, global development, and service to others. Thank you, kiitos, merci and tack to all my friends in Otaniemi, Montréal and around the world for your love and friendship, and for always encouraging me to be who I am. I also want to (and will) specifically thank my dear friends in the Polytech Choir, who will doubtless carry out a *thorough* proof-reading of this work post-publication. Thank you for giving me a voice!

Ympäri maailmaa kierrettyäni on hienoa tietää, että minulla on aina henkinen koti Otaniemessä. Etiopialainen sananlasku sanoo: ”elämällä ei ole merkitystä ilman kotia”. Siispä kannan teekkarilakkiani vastaisuudessakin aina suurella ylpeydellä!

Helsinki, 19th July 2021 ሰኞ ሐምሌ 12 ቀን 2013 በሄልሲንኪ ውስጥ

Eelis Hemberg ኤሊስ ሀምበርግ

Abbreviations and key terms

Abbreviations

ACP	African, Caribbean, and Pacific Group of States
CWA	Consolidated WASH Account
CMP	Community-managed project
COWASH	Community-lead Accelerated WASH project
CTA	Technical Centre for Agricultural and Rural Cooperation ACP-EU
DFID	Department for International Development (United Kingdom)
GTP II	Growth and Transformation Plan, Phase II (2016-2021)
GoE	Government of Ethiopia
IGO	Inter-governmental organization
IFAD	International Fund for Agricultural Development
JMP	Joint Monitoring Programme
KPI	Key Performance Indicator
LDCs	Least developed countries
M&E	Monitoring and Evaluation
MIS	Monitoring Information System
MDG(s)	Millenium Development Goal(s)
MoW(I)E	Ministry of Water, Irrigation and Electricity (Ethiopia)
MoFA	Ministry of Foreign Affairs (Finland)
NGO	Non-governmental organization
NWI	National Water Inventory
OWNP	One Wash National Programme
OECD	Organisation for Economic Cooperation and Development
O&M	Operations and Maintenance
PDRE	People's Democratic Republic of Ethiopia (1987-1991)
PMDB	Planning, Monitoring and Reporting Database (COWASH)
RBM	Results-based monitoring
SDG(s)	Sustainable Development Goal(s)
SDG 6	Sustainable Development Goal 6 ("Ensure availability and sustainable management of water and sanitation for all")
SNA	Social Network Analysis
SSA	Sub-Saharan Africa
SWAp	Sector-wide approach
WASH	Water, sanitation, and hygiene
WASHCO	Water supply, sanitation, and hygiene committee

Administrative divisions in Ethiopia

Gott	Administrative area equivalent to a village or community (unofficial, traditional)
Kebele	Administrative area equivalent to a municipality, comprising of several villages (*gott*)
Woreda	Administrative area equivalent to a district, comprising of several *kebele*
Zone	An intermediary administrative unit composed of several *woredas*, usually without financial autonomy
Regional state	First level administrative division, subdivided into zones – highest unit of sub-national administration

1 Introduction

According to an Ethiopian proverb "you think of water when the well is empty" (Rodarte 2003). Indeed, water is so foundational for societies that it can be seen as one of the most important prerequisites to development. Ensuring access to safe and sustainable drinking and household water for all is instrumental in tackling poverty and vulnerability, and investment in these services brings proven direct and indirect benefits to communities (Calow, Ludi, and Tucker 2013). But how can it be ensured that progress is made in terms of providing access to this resource for all? How can we measure how progress is made on this front globally, nationally, and locally? These are the elemental questions that the international development community, national governments, and local communities have grappled with for several decades.

Definitions, nuance, and classifications around water and development are incredibly important, and seemingly small adjustments to these can alter the lives of millions of people. In many ways, monitoring and evaluation methodologies form the backbone of effective water sector development cooperation projects, since they allow for important quantitative and qualitative determinations on the effectiveness of interventions within the scope of the activities of a project, and the methodologies in use have implications when results are aggregated and abstracted at higher levels. As such, it is of paramount importance to ensure that the indicators and frameworks used to monitor and evaluate development progress are relevant and well-integrated into the prevailing objectives and mechanisms at all levels of analysis. Consequently, there has been an increasing recognition in the recent decades that there must be a profound paradigm shift in the field of monitoring and evaluation (M&E) from a devotion to data repositories with periodic revisions towards more performance-oriented and responsive practices (da Silva Wells, van Lieshout, and Uytewaal 2013).

This shift is of particular importance due to the evolution of the global commitments and goals surrounding the sector, and the ways in which progress is measured therein. Not only has the rural water sector seen significant change during the last decades due in part to factors such as climate change, economic development, and population dynamics in many countries of the Global South, but there has also been an increasing recognition of the social equity and governance-related complexities in securing access to drinking water in an equitable and sustainable manner. This is most prominently reflected in the adoption of Sustainable Development Goal 6 in 2015 as the key global framework for addressing these challenges in the water sector, situated in a part of a decades-long continuum of global initiatives that have led to the inception of this much more sophisticated global M&E context.

It is increasingly evident that M&E practices should not be seen as separate mechanisms to be conducted 'alongside' project management, nor should they be seen as mere means of 'data acquisition', but that there should be an aspiration towards continuous and iterative improvement of the project's approach that is informed by effective analysis of results and project data (da Silva Wells, van Lieshout, and Uytewaal 2013; Thomson and Koehler 2016). In the past it may have been as sufficient to collect data based on set parameters, to measure progress and then draw on these successes in a periodic fashion to improve the project's approach along set intervals. Modern multidimensional and diverse indicators, in

contrast, require a more holistic approach of corrective measures and continuous evaluation to ensure responsiveness.

In the latter type of system, M&E cannot be seen as separate from the project's management and service delivery rationale, but there should be constant reflection and feedback to ensure that corrective measures are taken where appropriate. This is also more realistic than before in part due to emergence of new technologies facilitating more real time and responsive M&E practices, such as big data collection from the public through mobile phone technology, remote sensors used to monitor facilities or remote sensing data used to monitor surface water availability (Thomas et al. 2018).

An indirect consequence of this development is that there is likely a certain expectation of ambition placed on improving the M&E systems in place at different scales, and the requirement to improve the integration of these systems across scales from the micro strata of project execution to macro strata of national WASH frameworks, to macro-level global aggregation of SDG reporting. It could be argued that achieving meaningful progress towards SDG 6 requires not only sufficient resources and planning across scales, but also concerted efforts to improving how the results are measured and how well these observations feed into enhancing project effectiveness.

While the emergent paradigm shifts in water-sector M&E are clear, there persist many challenges in Sub-Saharan Africa (SSA) in general and Ethiopia in particular in achieving the ambitious goals of SDG 6. Aggregated UN Joint Monitoring Programme (JMP) figures show a considerable reduction of the share of the population relying on unimproved water sources in rural areas from 48.34% in 2000 to 27.74% in 2017, but Ethiopians residing in urban areas are still almost thrice as likely be covered by a basic or safely managed drinking water service than those residing in rural areas (WHO/UNICEF JMP 2021). This fact is in parallel to the continent as a whole, as only 45.3% of the rural population in SSA had access to basic or safely managed drinking water sources, compared to the global average of 78.02% (WHO/UNICEF JMP 2021). Indeed, despite signs of progress towards universal access to improved water sources for all in Ethiopia, there remains a clear service inequality, both between regions and by socioeconomic factors like household wealth (Damtew and Geremew 2020). This further underlines the importance of well-designed, effective, and responsive M&E systems in providing an accurate picture of progress in the sector.

1.1 Case study context: COWASH project and Ethiopia

The Community-led Accelerated WASH (COWASH) project is an intervention pursuing a variety of WASH-sector initiatives in mostly rural areas in Ethiopia. The focus of this research paper is Phase III of this project in 2016-2021. The project's Phase III has contributed to improved community and institutional water, sanitation supply coverage in select target *woredas* in five target regions of Ethiopia's 10 regional states. COWASH Phase III has also worked with increased functionality and sustainability of WASH facilities, along with women's empowerment and leadership in WASH-related activities. The project has received an extension in the form of Phase IV, which commenced during the time of writing in early 2021. (Ministry for Foreign Affairs of Finland 2016; Impact Consulting Oy Ltd 2019).

The COWASH project is an interesting case study of WASH interventions in Ethiopia in part because of its long existence in several forms, as well as its longstanding commitment to decentralization of initiatives through the Community Managed Project (CMP) approach. This approach is based around a service delivery methodology whereby "investment funds for physical construction or rehabilitation are transferred via regional micro-finance institutions directly to communities" and "communities are responsible for the water supply development process through planning, procurement, implementation and maintenance"; this service delivery methodology is established as one of the four official rural WASH financing modalities in Ethiopia (Suominen and Rautiainen 2016). The roots of CMP can be traced to previous iterations of bilateral cooperation between Ethiopia and Finland, and it has evolved as an approach to rural WASH interventions in conjunction with various previous project interventions such as the Rural Water Supply and Environmental Program (RWSEP) Phases I-IV from 1994 to 201 one WASH1 and the COWASH project since 2011 (Behailu, Suominen, and Katko 2015).

Additionally, there is also an interesting and rather unique national sector-wide consultation process that has been ongoing in Ethiopia, which has culminated in a nationwide National WASH Inventory (NWI) being carried out in two phases to provide baseline data at the household level (Welle 2013). The One WASH National Programme (OWNP) has been operational since 2013, which is an initiative to create a sector-wide approach (SWAp) to results-based monitoring in the WASH sector nationwide, along with "harmonizing government and donor approaches to planning, procurement, implementation and financing" and building a "platform on which a closer partnership can be built between planners, implementers, development partners and others to achieve common goals". (Wilson et al. 2018)

1.2 Purpose and scope of the study

This thesis seeks to explore the challenge of reconciling the increased level of M&E ambition stemming from the SDGs, with the difficulty of implementing this ambition into practice and operationalizing them through project interventions. This thematic is approached in the case study context through the following research questions:

1. *How has the increased level of ambition set forth by SDG 6 influenced the monitoring and evaluation practices in place?*
2. *How effective are the M&E procedures at encouraging effective data collection, responsiveness, and improvement of project interventions?*
3. *How well integrated are the indicators and M&E frameworks in use at different strata, and how are challenges of data aggregation being addressed?*

The study has been constrained to the rural water sector in Ethiopia. This is both because the rural areas are considerably more disadvantaged compared to urban settings in the country, and since addressing these areas is the chief focus of the COWASH Phase III project and the associated M&E environment, which is in turn used as a case study to elaborate on broader sectoral trends and draw conclusions. The study also intentionally focuses on the household water supply

component of the case study project, expressed in SDG 6.1 (with the exclusion of the sanitation, institutional WASH, and women's empowerment components expressed in other targets). This is because there are several behavioural and societal aspects in the M&E methods for these types of interventions that make it rather complicated to analyze within the constraints of a thesis.

The study uses a three-stage stratification to categorize the different levels in use in the case study context – the micro (project), meso (national) and macro (global) strata cover a vertical cross-section of the path of project data in the case study context, from the most local level of service delivery all the way to the global level of analysis. It was deemed appropriate to use a simple three-stage classification in order to bring clarity to the web of interactions in the M&E system of the case study context, and to help in analyzing the hierarchical data transformations and interactions therein. The stratification, along with the justification for using this stratification, have been further elaborated in Section 3, with a summary in Table 2.

The research questions have been explored through a case study methodology, relying mostly on semi-structured interviews with key informants (N=7) covering most of the reporting and data acquisition workflow in the M&E systems related to the COWASH project. Furthermore, a mostly paper-based indicator correspondence analysis has been carried out, in order to analyze the connections and interdependencies in the data acquisition and results reporting hierarchy of the case study context. For these, two analytical frameworks with relevant methodological rationales were devised, further elaborated in sections 3 and 4.

Although the M&E system and practices are analyzed holistically in terms of performance and effectiveness, the main focus of the indicator correspondence analysis in particular will be placed on the rural water supply access coverage components and indicators on each stratum. In terms of the SDG indicator framework, this entails a focus on Target 6.1 "by 2030, achieve universal and equitable access to safe and affordable drinking water for all" and the associated global SDG Indicator 6.1.1 "proportion of population using safely managed drinking water services" (UN Water 2017b). The corresponding focus area in the national OWNP framework is the "Rural and Pastoral WASH", sub-component 1 "Rural Water Supply" (Wilson et al. 2018), and in the context of the Growth and Transformation Plan (GTP) Phase II it is the sectoral target 4.4 "Rural potable water supply coverage as per GTPII standards" (Government of Ethiopia 2015). In the context of the Result Framework of the COWASH Phase III project, this is mainly captured by Outcome 1 "Increased climate resilient community and institutional water supply coverage (GTPII standards, including water quality) in the target *woredas* in 5 program Regions by 2019", and specifically by the associated Output 1.1 "Access to new and improved water sources for communities".

The study is particularly timely and relevant since the research takes place at a pivotal moment on all three strata of analysis. At the global stratum, the SDGs have begun to be operationalized, and the five years of the new global sustainable development framework call for greater knowledge about the contributing and hindering factors at play across the different spheres of development cooperation work. At the national stratum of Ethiopia as a country, the operationalization of the OWNP is underway, with a new Monitoring Information System under development. The COWASH project's Phase III concluded during the time of writing, and the new

Phase IV begun in early 2021. As such, the findings are important both in directly informing the case study project, as well as providing insights to stakeholders at different levels and better understanding the flows of information across different strata. In some ways, this thesis serves as a “meta-evaluation” of the M&E system used in the COWASH project, and its relationship with the M&E systems used by associated institutions at different levels.

2 Literature review

There is a rather long history of water sector development cooperation in Ethiopia, and as a natural consequence, the evolution of M&E theory and practice has been studied relatively extensively in this context. However, there are a few main reasons why the current body of research is not entirely sufficient in consistently describing the monitoring and evaluation processes in the rural Ethiopian water sector in the contemporary context.

First, the constant evolution of both international and domestic frameworks governing M&E in the sector and the relatively recent adoption of the SDGs as an influencing factor mean that there is still not a lot of research into how the recent developments have changed M&E practices and implementation.

Furthermore, while there exists research on the individual local, national, and global trends and processes concerning the rural Ethiopian water sector, there does not seem to be any consistent review of how the process of information aggregation and building of data composites from local results occurs *across* these different scales. As such, it is difficult to consistently evaluate how effective this process is at present, and what changes may be required to attain more accurate and reliable results.

The main service delivery methods of the COWASH project, in turn, have been quite consistently discussed, including directly in the case study context of COWASH and its predecessor projects. This is valuable information, as it complements project documentation in understanding the case study context and its evolution. The individual strata of the prevailing M&E framework have also been relatively well documented in isolation, and quite a lot of white paper research, reports and other documentation is available on the individual parts of the monitoring system. The connections between each stratum, however, remain rather vague, and seem to have been little discussed in conjunction with one another.

2.1 Macro stratum: SDG 6 and the global M&E framework

2.1.1 Sustainable Development Goals and underlying M&E theory

The global regulatory and measurement framework related to water resources has evolved considerably over time. Particularly important in terms of international development community's commitment was the replacement of the Millenium Development Goals with the new Sustainable Development Goals and Agenda 2030 in 2015. Although SDG 6 exists in a sequence of various other international high-level water-sector development commitments, including International Drinking Water Supply and Sanitation Decade from 1981 to 1990, the New Delhi Statement from 1991 to 2000 and the MDGs from 2001 to 2015, there are also marked differences in the way in which the SDGs approach this thematic (Sadoff, Borgomeo, and Uhlenbrook 2020).

The Millenium Development Goals (MDGs) represented "a historic and effective method of global mobilization to achieve a set of important goals worldwide" – they also mobilized concern around pressing issues such as poverty, inequality, and environmental degradation under a set of eight clear and manageable goals (Sachs 2012). As 2015 and the end of the MDGs was approaching, there was considerable discussion about the direction of the agenda that development cooperation should

take in terms of the global policymaking framework (Ziai 2016). After a considerably more participatory process of consultation involving civil society, private sector stakeholders, and a larger number of governments, considerably a considerably more ambitious framework for international development cooperation was devised (Feeny 2020). Sachs (2020) notes that the conception of the SDGs was informed the notion that sustainable development (combining economic development, environmental sustainability, and social inclusion) was a globally acknowledged consideration, and there was broad agreement that this "triple bottom line approach to human wellbeing" had to be at the heart of global development efforts (Sachs 2012).

While the MDGs were blamed for not being ambitious enough, the SDGs stepped up the level of ambition considerably. Sadoff et al (2020) describe the shift to the SDGs as sectorally significant, since it quite fundamentally expanded and altered how progress in water sector development is framed, and subsequently how this progress is monitored and measured. In the previous MDG framework, water was addressed under MDG Target 7C, entitled "halve, by 2015, the proportion of the population without sustainable access to safe drinking water and basic sanitation" (United Nations 2015). While the goal was reached globally in terms of drinking water access, there was also a quite unidimensional focus on water access at the expense of water resource management and governance (Sadoff, Borgomeo, and Uhlenbrook 2020). Sadoff et al write that "the MDGs' focus on access reflected a world where the limiting factors to delivering water services were related to infrastructure, capital or management, not governance and the scarcity and variability of the water resource" (Sadoff, Borgomeo, and Uhlenbrook 2020). Experiences from the MDGs informed a much more multidimensional understanding of "sustainable water and sanitation for all" present in the SDGs.

The challenges of indicator selection are set into the context of the MDG-SDG transformation by Thomson and Koehler (2016), who argue that a key problem of the one-dimensional nature of the MDG monitoring framework was that the indicators were overly focused on measuring progress towards the set goals, instead of thinking about the indicators as themselves being seen as drivers of better performance (Thomson and Koehler 2016). Under the MDG framework, target 7C was met, but the monitoring framework relied heavily on "improved" water sources, while ignoring quality, quantity and access of these water sources – as such, the SDG framework was altered to be more sensitive to this, and the SDG 6.1 goal on drinking water was formulated as "by 2030, achieve universal and equitable access to safe and affordable drinking water for all" (UN Water 2017b; Thomson and Koehler 2016).

Thomson and Koehler also point out that while new technological innovations have provided opportunities in developing more responsive monitoring systems, automated and technologically refined systems alone cannot suffice. There should also be a conscious transition from more traditional M&E paradigms focused on cyclical alterations and lessons learned, and an embrace of a "surveillance-response" paradigm, where monitoring provides a "service ladder" to attain SDG 6.1. Under this type of operation, a system driven by operational feedback loops combined with more responsive technological applications can provide a more consistent overview of performance. While periodic monitoring provides a "snapshot in time" and may obscure temporal variability. Considering both the opportunities and challenges of both consistent and intermittent M&E practices, the authors

argue for a combination of the two as an achievable alternative, as this provides for the more nuanced and responsive data collection needs that effective monitoring in the age of the SDGs requires (Thomson and Koehler 2016). To formulate this notion another way, modern M&E is not necessarily separate from operations and management, but the former should be used to inform and improve the latter throughout, which requires new types of methods and practices.

A natural implication of widening the scope and type of water-sector commitments are challenges in formulating suitable indicators and monitoring methods. While the MDGs could target efforts on monitoring access to water resources, the SDGs include more systemic and transformative measures, backed by a wide set of quantitative indicators. However, Sadoff et al claim that the ambitious measures are not at present backed by ambitious actions, and the world consequently is not on track to achieve clean water and sanitation for all – for instance, the paradigm of water engineering must be rethought from linear and centralized delivery systems to more robust and flexible systems, and management practices need to become more adaptive and integrated (Sadoff, Borgomeo, and Uhlenbrook 2020). To place this consideration into the context of the topic of this thesis, devising great indicators alone does not alone suffice, but there has to be cross-sectional and fundamental change in planning, governance, and monitoring to materialize this shift and achieve the goals.

Indicator selection criteria have been explored by Schwemlein et al (2016) in an article published at the inception of the new SDG framework. Although well thought out indicators are crucial in the WASH sector in terms of quantifying significant information and creating useful abstractions of complex phenomena (Hammond et al. 1995), in water sector development cooperation, consistent indicator selection methods are not always used in project interventions (Schwemlein, Cronk, and Bartram 2016). After a systemic review of 20 articles from various fields and classification of various indicator selection frameworks, Schwemlein et al propose a more standardized six-step indicator selection process. After defining the purpose and scope for monitoring, selecting a conceptual framework for categorizing indicators logically should take place; thereafter, selection of candidate indicators takes place based on existing literature, after which selection criteria (such as measurability, reliability, data availability, among others) are determined (Schwemlein, Cronk, and Bartram 2016). Finally, indicators are scored against these criteria and final indicators are selected based on these results (Schwemlein, Cronk, and Bartram 2016).

The authors note that the benefit of selecting “clear, consistent indicators in WASH” allow for “comparisons in status across space and time”, such as between projects, programs, or countries (Schwemlein, Cronk, and Bartram 2016). Indeed, it could be argued that this type of methodology is a prerequisite for any process in which information is collated from various sources, but particularly in a highly ambitious global framework such as the SDGs. In addition, as Schwemlein et al point out, “indicators do not necessarily reveal the entire situation of a project or program and data must be interpreted with care”, further underlining the fact that project indicators are always abstractions of reality (Schwemlein, Cronk, and Bartram 2016). The limitations of indicators should be recognized, particularly in terms of comparability between different contexts or instances.

The transferal and subsequent operationalization of the SDGs to lower scales is discussed by Janoušková et al (2018), who also argue that "a procedurally well-designed, conceptual indicator framework for selecting and/or designing indicators" is required in order to assess SDG results on a global level (Janoušková, Hák, and Moldan 2018). Since the SDGs are not legally binding commitments, governments are responsible for many aspects of their practical implementation. While the SDGs provide for a consistent and firmly embedded policy framework, there exists a considerable risk of ambiguity when SDG results are aggregated on a globally. An example of problematic aggregation, they argue, is SDG indexing and ranking of such figures. While appealing due to their advantage of simplifying communication of progress, divergent methods between countries and data aggregators cast a lot of doubt towards the usability and sensibility of these types of figures as policy support instruments (Janoušková, Hák, and Moldan 2018).

Da Silva Wells et al (2013) underline the importance of capacity development, transparency, and sectoral reviews in developing a WASH sector M&E system that both collects reliable data and strives for continuous improved performance. Although written during the era of the MDGs, da Silva Wells et al provides a good look into what sector-wide continuous improvement entails in this context. Due to the water sector being an operating environment with considerable uncertainty due to associated rapid social and environmental change, monitoring can provide early signs of problems and successes, when coupled with a continuous cyclical model where monitoring processes inform adaptation of future implementation measures (da Silva Wells, van Lieshout, and Uytewaal 2013). Continuous learning processes are central to adaptive management practices and must involve a large spectrum of stakeholders – one practical implementation method for this is the Joint Sector Review (JSR), "sector-wide reviews that are led by national government and involve all major stakeholders are increasingly recognized as key to improving WASH coordination and planning" (da Silva Wells, van Lieshout, and Uytewaal 2013). When planned with the appropriate steps and infrastructure, as well as coupled with multi-stakeholder iterative development platforms and emancipatory capacity building efforts at all levels, such reviews can produce good outcomes, evidence of which exists throughout Africa and elsewhere (da Silva Wells, van Lieshout, and Uytewaal 2013). However, there must also be the recognition that monitoring is not synonymous with mere information management, but "the value of extensive data and information is limited if it is not reflected upon and used in decision-making" (da Silva Wells, van Lieshout, and Uytewaal 2013).

2.1.2 Practical implementation of SDG 6 in the WASH sector

While it is crucial to understand the underlying theory of the SDGs and associated M&E systems that have been developed in conjunction with it, the technical side of the practical operationalization of the goals is also an important consideration. This thematic has been elaborated mainly in donor literature, while academic research seems to focus more on the underlying theory of the monitoring of the SDGs.

Thomas et al (2018), in a World Bank-published book, review some of the underlying theory and rationale of the indicators and data sources used for monitoring SDG 6.1 in greater detail. They also explore possible improvements arising from emergent technologies and methods (Thomas et al. 2018). Since global data

collection for monitoring requires "timely and reliable data gathered in a cost-effective manner", household surveys and censuses will likely remain primary data sources for monitoring of SDG 6.1. Since this was the primary data collection method for MDG 7, it makes sense to build upon the institutional knowledge and successful measures built under the previous framework. However, "in order to address the ambition of the SDG targets, other data sources will be progressively integrated" in the JMP monitoring framework; these include administrative sources and data collected by regulators, along with more novel methods such as in situ sensors, water quality testing, and remote sensing earth observations (Thomas et al. 2018). In addition, while the MDG 7 focused on "improved" water sources as its core element, the SDG 6.1 monitoring methodology expands this to four elements: as such, "safely managed drinking water" consists of "1. Improved drinking water source that is 2. Located on premises, 3. Available when needed, and 4. Compliant with fecal (and priority chemical) standards" (Thomas et al. 2018). For the first three of these steps, data can be effectively collected from survey data either explicitly or implicitly, whereas the fourth can be monitored by a combination of survey results and data collected by regulatory authorities on coverage, quality, and access (Thomas et al. 2018).

Thomas et al also review some of the technological applications that have recently emerged as plausible tools for WASH-sector monitoring. In-situ monitoring through the use of microbial or chemical water quality sensors, as well as remote sensing through satellite spectral imagery, "may offer some contribution to addressing some of the challenges of information asymmetry and data gaps in developing communities including unreliable survey data and relying on spot checks to assess performance" (Thomas et al. 2018).

In the Integrated Monitoring Guide for SDG 6 sub-component G1 (good practices for country monitoring systems) published by UN Water, the topic of operationalizing SDG 6 into practice through cross-sectional cooperation both between different levels and across different sectors has been discussed. The report states that while the goals are defined as "global and aspirational", it is clear that interventions should be tailored to national circumstances through available resources, existing capacity, and level of urgency around different issues in each country context (UN Water 2017a). For effective policymaking at the national and sub-national level, "the global indicators are still useful, but the data need to be disaggregated spatially and temporarily, by sector, subcomponents and different socioeconomic strata" (UN Water 2017a). The report calls for inclusiveness of stakeholders, such as civil society and private sector actors, in evaluating progress nationally (UN Water 2017a).

The discourse around the monitoring and evaluation SDGs is also best placed into a context of results-based monitoring (RBM), which has been increasing in importance already during the MDG era. In a World Bank-published book, Kusek and Rist (2004) outline the importance of RBM in fostering accountability and transparency as increasing responsiveness to internal and stakeholders in demonstrating results must be prioritized, citing various imperatives for reform, including the MDGs (Kusek and Rist 2004). However, pursuing the RBM approach is sometimes difficult to achieve in a context-sensitive fashion. On the issues around results-based management systems in the case of multilateral organizations in the Sub-Saharan African context is provided by Lockwood (2015), who notes that it

may be challenging to pursue a results-based approach to monitoring and evaluation that is "accountable to both development partners and beneficiaries", all the while "retaining a focus on efficiency and effectiveness (Lockwood 2015).

2.2 Meso and micro strata: M&E in the rural Ethiopian water sector

2.2.1 Ethiopia in the current global M&E framework

Academic research that contextualizes the Ethiopian rural water sector in the wider global framework described in Section 2.1 does exist, although rapid economic development, intersecting societal trends and rapid evolution of the key governing frameworks both nationally and internationally makes it sometimes difficult to ascertain one individual core narrative. However, there are also some interesting development paradigms and trends that arise from existing literature set in the regional context at hand.

In a book stemming from a United Kingdom Department for International Development (DFID) funded research program in Ethiopia, Butterworth et al (2013) explore the issues around WASH sector monitoring at length, focusing on the rationale behind the processes and approaches taken on different levels in terms of data collection and utilization (Butterworth et al. 2013). The authors, focusing on the rural water supply, provide an overview of the general process of M&E across scales, and the advantages and disadvantages of the major processes in place at the time of writing in 2013. Butterworth et al note that there was a significant discrepancy in the estimates of rural water supply coverage measured by the figures of the JMP's methodologies and the figures reported by the Ethiopian Ministry of Water and Energy (MoWIE). While estimates on use of improved water facilities by the JMP and the water access coverage figure reported by MoWIE in the year 2010 had little divergence in urban settings (91.5% and 97% coverage respectively), there was a huge divergence in the figures for rural areas, with the national MoWIE figures reporting an estimate more than double that reported by the JMP; this trend of much higher estimates in national figures compared with those by the JMP has persisted from the 1990s (Butterworth et al. 2013).

Butterworth et al observe that key reasons for this discrepancy were, at the time of writing in 2013, methodological differences in compiling the estimates related to data sources, as well as definitions of access and coverage. While the international JMP figures were compiled from a wider variety of sources including "household surveys, regional inventories, updates used by the government to prepare official sector reports and the Ethiopian National WASH Inventory (NWI)" with a standardized set of criteria for evaluating the percentage of people with *access* to *improved* water sources, the Ethiopian national estimates relied on the NWI and followed more crude federal guidelines relying on *coverage* or "potential of access" that assumes that schemes serve a certain population and number of users, resulting in the possible scenario where systems that were non-functional or partially functional at the time of inventory were included in the figure (Butterworth et al. 2013). Furthermore, the definitions of "rural" areas were different in the national and international reports (15 liters per capita per day and 20 liters per capita per day, respectively), resulting in further challenges in comparability (Butterworth et al. 2013)

The important conclusion that Butterworth et al note from this divergence between national and international figures is that while the development of better practices for WASH sector monitoring is a vital consideration for better sector-wide performance, the authors foresee that there will likely not be a single all-encompassing M&E system in place, but different parallel systems at the global, national, and local levels (Butterworth et al. 2013). Since the methodologies in use for each are fundamentally different and thus results generated will also differ, there is a requirement for "navigation of the interface between international and national monitoring" and "critical analysis on the use of different methods and a better understanding of the perspectives of the organizations that generate the results" are required in the future (Butterworth et al. 2013). These remarks by Butterworth et al are interesting, as they demonstrate that the challenges pertaining to the development of M&E systems and ultimately the aggregation of results across different levels has been present for several decades in Ethiopia prior to the adoption of the SDGs as the guiding framework in 2015, and reconciling these differences is likely to remain firmly on the agenda in the future.

A working paper co-written by UNICEF, the think-tank IRC, and consulting firm Akvo (2016) provides a "generic organizing framework for a functional national WASH sector monitoring and evaluation (M&E) system" (IRC 2016). According to this framework, "the purpose of national monitoring and evaluation is to enable effective decision-making – at all levels within a country – through the use of continuous, reliable, and relevant data and indicators which can be processed, analyzed and used to inform decisions" (IRC 2016). The organizing framework they propose includes 12 components organized into three spheres. The first sphere (people, partnership, and planning) is concerned with "human resources, partnerships and planning to support data collection and data use" (IRC 2016). These components act as a precondition for the second sphere (collecting, verifying, and analyzing data) achieved through methodologies, consisting of routine monitoring, surveys, national and sub-national WASH databases, supervision/auditing, and evaluation/research (IRC 2016). These then enable the third sphere of access and use, which is used for decision-making support (IRC 2016). The authors of the report also remark that the implementation of steps should not be sequential, but all need to be implemented on an acceptable national standard to function effectively – all components also need not to be implemented on all levels, of the system (IRC 2016). While the IRC framework is not country-specific, it identifies several issues regarding national M&E system planning and implementation that act as important preconditions for effectively monitoring WASH interventions nationally.

2.2.2 Emergence of CMP as a major service delivery method

The Community Managed Project (CMP) method is an increasingly important service delivery mechanism in water-sector development cooperation projects in Ethiopia (Behailu, Suominen, and Katko 2015). Beginning in the 1990s, it started to become clear that the conventional top-down management approach had important shortfalls, as there was a failure to consider infrastructure development and the social aspects of development in conjunction with one another – these were increasingly recognized as mutually connected in the global development community since the 1990s in a variety of forums (Behailu, Suominen, and Katko 2015). The evolution of CMP also runs parallel to the wider societal context of the fall of

the Etiopian military government in 1991, which resulted in a more decentralized governance mechanism compared to those of the *Derg* military junta and People's Democratic Republic of Ethiopia (PDRE) administration from 1974 to 1991 – the post-communist era has also seen in greater focus placed on rural water and sanitation systems in the country (Behailu, Suominen, and Katko 2015).

This evolution of the delivery mechanism in use can be seen in the predecessor of the COWASH project, the Rural Water Supply and Environmental Program (RWSEP), from 1994 to 2011. Initially the implementation of the RWSEP programme relied on directly funded delivery methods, both administered and managed on the *woreda* level. Participatory rural appraisal was important in appropriately directing development efforts to where there need for them were identified. Capacity development in local communities in terms of financing, management and construction were prioritized, and a multisectoral approach to a wide range of issues was pursued. Eventually, the CMP approach was scaled up into the One Water National Program (OWNP) under development, and the COWASH project was instituted as the successor to facilitate this development (Behailu, Suominen, and Katko 2015). Setting this evolution in the wider context of the theoretical literature, the CMP methodology can be seen as attempt towards more participatory, sector wide capacity development initiatives called for by da Silva Wells et al, or the adoption of a "service-response" paradigm described by Thomson and Koehler.

Senbeta and Shu (2019) draw on the social empowerment implications of selecting novel community-managed bottom-up approaches in the rural water sector context instead of conventional top-down practices: "the communities through representative user WASHCOs are responsible for the overall process of their water supply development, including planning, construction management, financial responsibility, implementation, operation and maintenance management of water services in community-managed project approaches" (Senbeta and Shu 2019). The authors found that "sustainability outcomes of water services were affected by project implementation management modalities under the study", with the CMP approach bearing a correlation with high levels of institutional performance and service functionality in the study area. Although not a conventional approach in the SSA context, "community management of water infrastructures generates higher institutional, managerial, financial, and technical sustainability performance compared to projects managed by local government, NGOs, and charity organizations" (Senbeta and Shu 2019). As such, there is evidence that CMP as a service delivery method is effective in work towards SDG 6, and they argue that governments and policymakers ought to focus "besides participation, on the empowerment of user communities in the construction of water points" (Senbeta and Shu 2019).

Person et al (2017) contextualize the governance and social empowerment components of projects and M&E systems through the lens social capital (networks, norms, and trust that facilitate cooperative behaviors). Evidence suggests social capital "influences a community's ability to manage communal water resources" particularly in the rural water sector due to its "relationship to cooperative behavior and the social nature of rural water governance" (Person et al. 2017). The study included a household survey analysis of 20 *gotts* using a Social Capital Assessment Tool to quantify social capital in these target communities with a social capital index, through identifying the presence of six social capital domains (groups and networks; collective action and cooperation; trust and solidarity; information and

communication; social cohesion and inclusion; empowerment and political action). The results of Person et al show that information and communication is "significantly associated with governance and is recognized in the literature as critical to water sustainability" (Person et al. 2017). These have implications in the current debate over water supply management approaches, including the decentralized methodology employed in the CMP approach. The flow of information and transparency in particular is identified as an issue of reciprocal trust which can be a major barrier to effective execution of water projects in the Ethiopian context (Person et al. 2017).

Giné Garriga et al (2015) have proposed a "monitoring framework to compile, analyze, interpret and disseminate water, sanitation and hygiene information" in decentralized service delivery contexts. According to the authors, planning of data collection processes must be concerned with two questions: "the data must be analyzed to produce outcomes that are relevant to the policy question, and the analysis must be disseminated and transmitted to policymakers" (Giné Garriga, Jiménez Fdez. de Palencia, and Pérez Foguet 2015). They underline the fact that, in order to truly develop pro-poor policies through local government entities, success "depends upon real efforts to strengthen the capacity of decentralized authorities"; in this regard, it is important that local authorities can truly utilize existing M&E frameworks, and these actors are awarded sufficient resources for doing so (Giné Garriga, Jiménez Fdez. de Palencia, and Pérez Foguet 2015).

2.2.3 Impacts of the federal governance structure and decentralization

One thematic that has several implications for the administration of bilateral development cooperation interventions in Ethiopia, including those in the rural water sector, is the federal government structure and strong degree of regional devolution of power arising therefrom.

Although a large country of over 117 million inhabitants in 2021 (United Nations DESA 2019), the roots of Ethiopia's devolved governance structure are in its history. Since Ethiopia's transition from the *Derg* military junta and PDRE to a federal democratic system of administration in 1991, decentralization of power was seen as a "precondition for transition to civilian rule" (Yilmaz and Venugopal 2008). This was in part due to the emergence of the Ethiopian People's Revolutionary Democratic Front (EPRDF) as the ruling political party, which has its roots in ethno-regional parties (Yilmaz and Venugopal 2008). There have been recent developments towards a more unitary, national, and arguably authoritarian form of governance after the dissolution of the EPRDF and formation of a unified, national party called Prosperity Party under prime minister Abiy Ahmed (Gedamu 2021). Although the developments with regards to ethno-regional conflict and governance structure change are very much ongoing at the time of writing, it is still clear that the present governance system awards considerable discretion to regional and local actors, something that has fundamentally shaped how rural water project interventions are structured and executed in Ethiopia.

Table 1 summarizes the most significant units of administration for the case study context. It is evident that there is a high degree of devolution in sectoral governance, with regional authorities wielding a considerable degree of autonomy in implementing national interventions. Furthermore, a lot of the implementation capacity has been further devolved especially to the *woreda* administrative level,

which has an important role in terms of practical implementation of initiatives, including development of rural infrastructure like water points (Yilmaz and Venugopal 2008). However, Ayenew writes that there is still a considerable degree of regional divergence between regions both in terms of administrative capacity and governance structure, and there are several unresolved jurisdictional matters in the relationship between local government entities (Ayenew 2002). There is also still considerable reliance on the national government that may hinder building effective federalism (Ayenew 2002).

One aspect of administrative structure and capacity that ought to be briefly mentioned in relation to local-level execution of initiatives in the CMP service delivery methodology is the legalization of Water, Sanitation and Hygiene Committees (WASHCOs). This has been identified in the COWASH Phase IV project document as a concern since the Government of Ethiopia cannot directly finance WASHCOs "as long they are not generally legalised and audited following the GoE system" (Impact Consulting Oy Ltd 2019). This further underlines the fact that building capacity at the lowest levels of WASH sector governance is still evolving and under development, and it is expected in project documentation that all WASHCOs will not be legalized until 2030 (Impact Consulting Oy Ltd 2019).

Table 1: Levels of administration in the Ethiopian federal system, as applied to the case of the rural water sector (adapted from Ayenew 2002; Yilmaz and Venugopal 2008; Butterworth et al. 2013; Besah et al 2016; Impact Consulting Oy Ltd 2019; UN DESA 2019)

Administrative level	Purpose and role of the administrative unit	Size in relation to adjacent levels of administration
Federal government	Nationwide policy coordination and technical guidance to lower levels of administration. Aggregation and processing of results. Ministry-level coordination with both regional actors and global partners.	Total population of 117,876,000
Regional state	Ethno-linguistically based first-level administrative subdivisions. Highest level of devolution in the federal system, with most political power devolved. Implementation of national development policies and all matters devolved by the national government.	10 regional states in total, with populations ranging from 200,000 to 25 million.
Zone	An intermediary level of government, usually technical without financial autonomy and political representation (with high degree of regional variation). May take the form of intermediary governance divisions or oversight bodies for woredas. Sometimes disregarded in WASH project interventions.	Typically, around 10,000 to a few million inhabitants. Around 3-12 in each regional state and 68 zones in total.
Woreda	An administrative unit equivalent to district. Dual accountability: upward to zonal and regional authorities and downward to the population. Management of woreda-level development plans and constructing local infrastructure, local national resource management.	Typically, around 10,000 to 300,000 inhabitants. On average around 85 woreda in a regional state and 770 in total.
Kebele	An administrative unit equivalent to municipality, lowest level of federal administration and most local official administrative unit. Functional division between kebele and woreda administrations not clear nationwide, with regional variation. Less constitutional formality than regional states and woredas.	An average of around 5,000 inhabitants. Usually around 30 kebele in one woreda, Innumerable (several thousand) kebele in total.
WASHCO	Water, sanitation and hygiene committee, local organizations used especially in the CMP service delivery approach for water management and maintenance. In charge of operations and management of systems at the most local level of service delivery, closest to beneficiary communities.	Approximately 5-7 members in one WASHCO. Approximately 220,000 WASHCO members in the COWASH project.

3 Research material and methods

The main method of academic inquiry in this thesis is the qualitative case study methodology, using the COWASH Phase III project and its associated national and global M&E mechanisms as a case study. Research was conducted through semi-structured interviews of key informants from different strata related to the project (Section 3.1). An analytical framework was devised to better situate these observations at different scales, based on a stratification arising from the literature (Section 3.2). In addition, existing indicators on different strata of analysis were cross-referenced to evaluate their compatibility in facilitating data aggregation in practice (Section 3.2).

3.1 Research methodology and strata of analysis

Selection of the case study methodology as the approach to qualitative inquiry was informed by Creswell (2013) and Yin (2009), who have exhaustively described the characteristics of this method and its possible alternatives. The COWASH project and the wider M&E system is a rather clear entity to situate within a wider context, and it provides a good instrumental case providing insights on the wider phenomenon of M&E in the Ethiopian water sector (Creswell 2013; Yin 2009). The analysis method can provide an in-depth understanding of processes and meaningful assertions of underlying trends (Creswell 2013; Yin 2009). Initially, there was also the possibility to include some ethnographic or observational components to the inquiry through the inclusion of a case study component, but this was unfortunately not possible due to circumstances described below. Although not the only possible alternative, case study research is well established in analyzing governance and social phenomena (Creswell 2013; Yin 2009), and this long precedent in similar fields of research underlines its appropriateness for this thesis. In other words, it was deemed that this type of inquiry would be efficient in uncovering information on cross-strata information flows and shed light on cross-institutional M&E interactions.

The levels of analysis have been simplified into three strata to provide a useful abstraction of the different settings on which data aggregation occurs in the case at hand. These have been adapted from Butterworth et al (2013), who have analyzed M&E systems in the Ethiopian WASH sector in their work with a three-level stratification. Butterworth et al use a three-strata classification system of 1. global level (measurement of progress across countries); 2. national level (facilitation, regulation, and monitoring of WASH across the country and the role of policy) and 3. the *woreda* level (service delivery, responding to unmet needs and system failures) (Butterworth et al. 2013). These have been adapted for the purpose of this thesis as macro, meso and micro levels, with the scopes of each elaborated in Table 2 – the main change has been the inclusion of the COWASH project as the micro stratum as a whole, not just the local administrative level.

The core reason for altering this division of the strata has been that the COWASH Phase III project is highly dispersed to the local level, although there is an ongoing effort to scale up the project's CMP service delivery methodology to the national policy framework (Ministry for Foreign Affairs of Finland 2016). Therefore, analyzing this system as the micro stratum is quite logical, and evaluating the

project as a whole also becomes easier if the entire project is included as a single stratum. Furthermore, since the facilitating and regulating role of the national institutions, along with the global institutions' concern with global process, are fitting in the case study context, there was seen no need to alter them.

Table 2: Three strata of analysis, with key stakeholders, formal institutions, and governance frameworks on each level

Stratum	Limits of the stratum	Key stakeholders	Key formal institutions and governance frameworks
Macro	Global and international processes of SDG monitoring and evaluation; inputs from national governments	JMP; MoFA	SDG 6 associated documentation; JMP data custodian technical M&E documentation and methodological notes
Meso	Ethiopian national water-sector M&E and governance frameworks	MoWIE; GoE	NWI, OWNP, GTP II
Micro	COWASH Phase III project governance; Regional, zone and local (*woreda*) levels of rural water system governance, individual beneficiary community level (*kebele, WASHCO*)	COWASH project staff and technical experts; Project-affiliated staff on lower levels of governance (regional state, *woreda*, *kebele,* WASHCO)	COWASH Phase III project governance mechanisms; regional and local administration structures

The main method of data collection for this thesis has been qualitative, semi-structured interviews with key informants (N=7) representing various roles on each of the three strata in this project above the regional level of administration (see Table 3 in Section 4). The main purpose of the interviews was to determine the practical nature of data aggregation in the case study context and the practical implications on effectiveness of the M&E system, along with generalizable trends for the sector as a whole.

The interview guide used for participants has been included in Annex I. Using the observations of Puusa (2020) as a reference, the semi-structured interview was selected as the best methodology for conducting the interviews, seeing that the rigidity of selecting pre-determined answers or criteria was not seen as necessary in obtaining replicable results and entirely omitting the role of the researcher (Puusa 2020). On the contrary, there was an interest in uncovering certain elements related to the M&E system that the informants are involved with in their own capacity, since there were likely to be nuanced observations that are omitted from formal documentation at each stratum of the M&E system and different stages of the data aggregation processes (Puusa 2020). The semi-structured interview methodology also allowed to slightly curate the questions to each informant based on their position within the wider M&E framework, since the informants are involved with the monitoring and evaluation of the COWASH Phase III project in wildly varying ways, and some only indirectly. Additionally, since these types of interview methods are well established in research into organizations and management (Puusa 2020), they were seen as appropriate methods into analyzing the functionality of M&E systems as well.

Initially, there was an intention in the project's planning phase to conduct a field visit to Ethiopia, with visits to select *woredas*, and to use these localities as example project sites to obtain additional information on the practical nature of

data collection on the ground. However, due to the global COVID-19 pandemic, which was ongoing throughout the length of the research for this thesis, travelling to conduct field research was rendered impractical. As such, the project has focused much more on the M&E systems on the upper micro and meso strata of analysis, namely the project's mechanisms and the associated national NWI II and OWNP mechanisms, as data for these could be more feasibly collected in the prevailing circumstances. In other words, the perspective was by necessity shifted less to the local processes of service delivery, and more to the national and global aspects of project work.

3.2 Analytical framework and indicator correspondence analysis

In order to clearly present the results obtained from the semi-structured interviews, it was deemed important to place these results in a simplified model of the data aggregation workflow. This was achieved by constructing a novel analytical framework and a supplementary indicator correspondence analysis methodology, to provide greater insight into the interactions in monitoring and evaluation across different strata. Both of these methods have been influenced by knowledge mapping and Social Network Analysis (SNA) methods to stakeholder analysis, as elaborated below.

The analytical framework was constructed based on the stratification introduced above and the study objectives, in order to assist in inspecting the data aggregation workflow and situating associated shortfalls, illustrated in Figure 1. Since a core purpose of the study is to provide a better general outlook into the effectiveness of data aggregation and identify possible shortfalls, this framework acts as a frame of reference in locating the precise scales at which these might occur.

The stratification presented in Section 3.1 was used to divide the data aggregation workflow of the analytical framework into three strata. While another type of division in the strata could have been possible, it was thought best to utilize a clear division as introduced above, which allows clear assertions to be made as to the interactions between institutions and actors on different strata.

The analytical framework seeks to identify the sources of data especially at the micro stratum, and the transformations the data undergoes upon aggregation at different strata as it is collected onto higher strata in data repositories. In the context of the case study project, this occurs at three stages, with three main data aggregation stages identified. In essence, the micro stratum acts as a data source for data associated to service delivery with some data processing and reposition taking place. The meso and macro strata, meanwhile, variously aggregate data for different end use purposes, with national reporting, MoFA progress reporting and JMP estimation for SDG 6 reporting being the ultimate end uses of the data beyond the purposes of the project itself. These have been used to produce a network diagram, essentially acting as a process knowledge map.

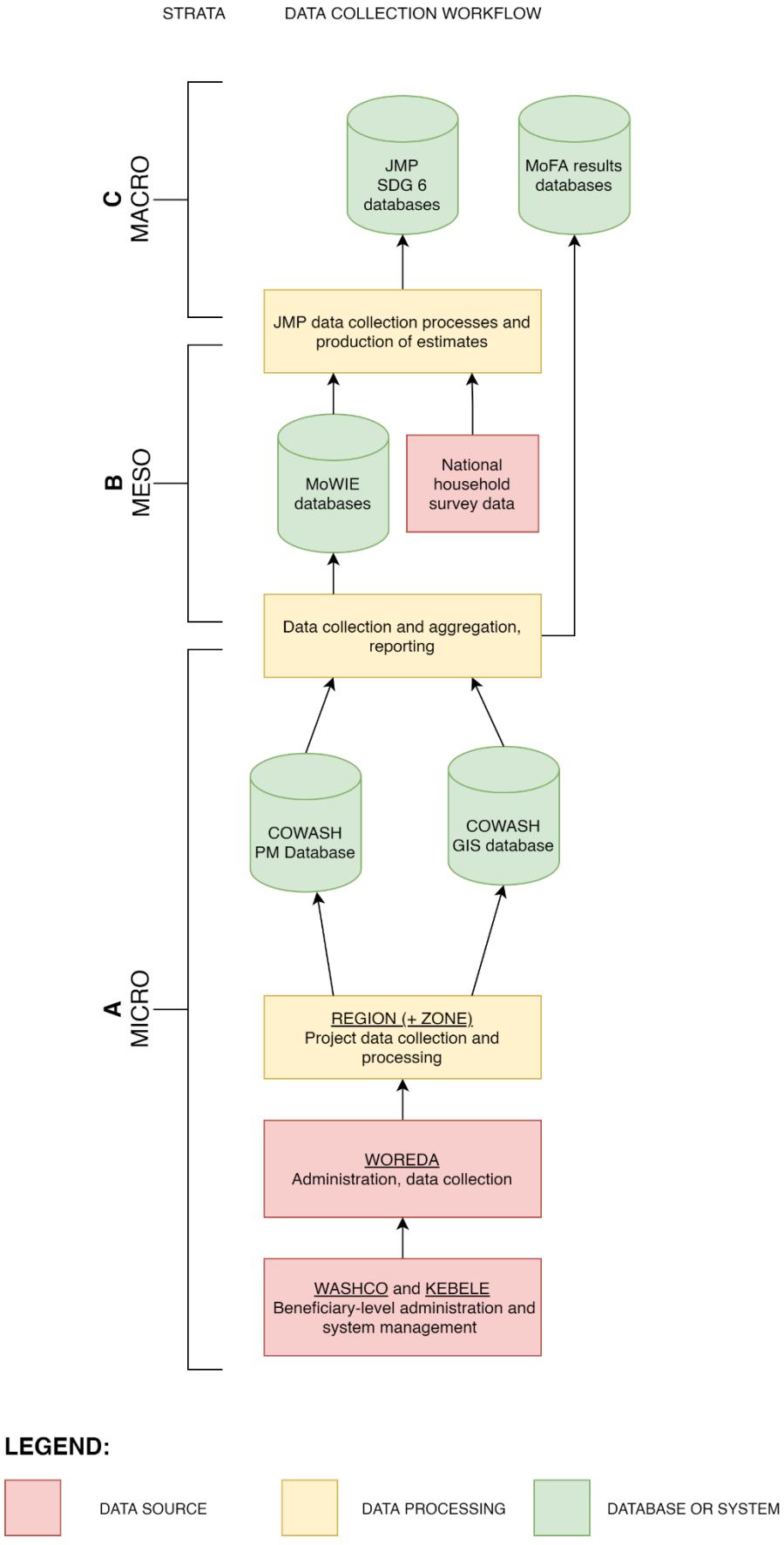

Figure 1: Analytical framework: data aggregation workflow of COWASH Phase III situated in broader data collection for national and global M&E systems

The observations obtained from the semi-structured interviews with key informants were supplemented with an indicator correspondence analysis of the indicators in use across the different strata of analysis, and how the practical application and data aggregation at different strata changes how the data flows across the different strata. The data for these was mostly acquired from available documentation at the different strata and concentrated on rural water access coverage indicators only (SDG 6.1.1 and equivalents), since this was seen as suitable in the limited scope of this work in providing insight into the rural household water sector thematic specifically. The analytical framework of indicator correspondence types (Fig. 2 in section 4) used to classify transformations is elaborated in greater detail in the results, but the indicators were evaluated for their degree of correspondence to one another to ascertain how the aggregation of indicator data occurs in practice across different strata.

The main reason for inspecting the indicators against each other, and their correspondence across scales, arises from the fact that the SDGs require rather sophisticated data to be collected vis-à-vis different M&E frameworks, but the operationalization of 'SDG-compatible' data falls on lower strata where policy framework planning, operationalization and practical execution of interventions occurs. Furthermore, if there is a desire to streamline data collection and move towards more responsive M&E practices, there is a definite need to effectively follow progress across these different scales or strata and using compatible indicators or other mitigation methodologies will certainly become necessary.

The analytical framework and indicator correspondence analysis models have been partially influenced by the knowledge mapping and social network analysis (SNA) methodologies of stakeholder analysis, which have been described by Reed et al in the context of natural resource management (Reed et al. 2009). The analytical framework and indicator correspondence visualizations have elements of a SNA-informed sociogram, although the network has been abstracted to a much more simplistic form in both. Furthermore, the methodology has been altered to focus more on information flows than mere relationships between the actors in the system, particularly in the case of indicator correspondence classifications. This type of abstraction has been done to ensure that the analysis remains appropriate for the scope of the work, and to focus on the M&E processes in particular over organizational coherence between stakeholders in general.

However, the use of several methodological elements of SNA and knowledge mapping in the analysis was seen as useful in formulating the analysis methodology, since the SNA-knowledge mapping interface provides an opportunity to extend the "who knows who" of SNA to visually represent "who knows what", and it also allows for "identifying the dominant flows of knowledge" and "identifying knowledge bottlenecks" (Reed et al. 2009). The analysis of information exchange mechanisms corresponds well to the desired outcome of evaluating the M&E systems in place (Reed et al. 2009). The research methodology described above is rather similar to the practical adaptation of knowledge mapping into the development cooperation context used by USAID, and the methodological steps "identify and categorize knowledge assets" and analyze "barriers and constraints to fulfilling goals and objectives"; the methodology also seeks to identify "decision milestones", "knowledge requirements" and "routes for access and retrieval of knowledge"

(USAID 2003). In major contrast to this, however, the study analyzes the process of knowledge transfer *between* distinct organizations, not as much the transfers by actors within those organizations, and the analytical focus has been more on data transformations between distinct stakeholders than the process of knowledge management by the stakeholders individually.

4 Results

Semi-structured interviews for the qualitative analysis were conducted with the seven informants in May and June 2021, as detailed in Table 3 – the informant IDs have also been colour-coded for the reader's convenience. The results of this analysis in terms of observations by strata have been summarized in Fig. 2. This assists in summarizing the results in the three-step stratification of the analytical framework used. Reference is made to the letter-number designations used in this figure when the interview results are discussed in the text.

Furthermore, the interview findings have been analyzed in greater detail in the context of the each of the research questions, in order to evaluate the themes of SDG 6 integration (section 4.1) and M&E system effectiveness (4.2) in the case study context. Some of the interview conclusions have also been incorporated together with the indicator correspondence analysis (4.2) to provide further insight to this analysis.

Table 3: List and description of the informants interviewed for the study

ID	Description of informant	Organization represented by the informant	Stratum	Date of the interview and medium of conducting the interview
1	COWASH project regional M&E specialist	COWASH Project, one of the Regional WASH Coordination Offices	Micro	21.5.2021, interviewed using interview guide by Abebaw Getachew, notes, no audio recording.
2	COWASH national M&E specialist	COWASH Project national office, Addis Ababa, Ethiopia	Micro	17.5.2021, interviewed using interview guide by author, teleconferencing medium, audio recording
3	Civil servant working with national WASH sector M&E	Ministry for Water, Irrigation and Energy, Addis Ababa, Ethiopia	Meso	14.6.2021, interviewed using interview guide by author, teleconferencing medium, audio recording
4	Civil servant working with national WASH sector M&E	Ministry for Water, Irrigation and Energy, Addis Ababa, Ethiopia	Meso	19.6.2021, interviewed using interview guide by author, teleconferencing medium, audio recording
5	Former Senior Specialist, water policy	Department for Development Policy, Ministry for Foreign Affairs of Finland, Helsinki, Finland	Macro	18.5.2021, interviewed using interview guide via email by author
6	Senior Advisor on Development Policy	Department for Development Policy, Ministry for Foreign Affairs of Finland, Helsinki, Finland	Macro	12.5.2021, interviewed using interview guide by author, teleconferencing medium, audio recording
7	Member of the Joint Monitoring Programme (JMP) team	World Health Organization, Geneva, Switzerland	Macro	11.6.2021, interviewed using interview guide by author, teleconferencing medium, audio recording

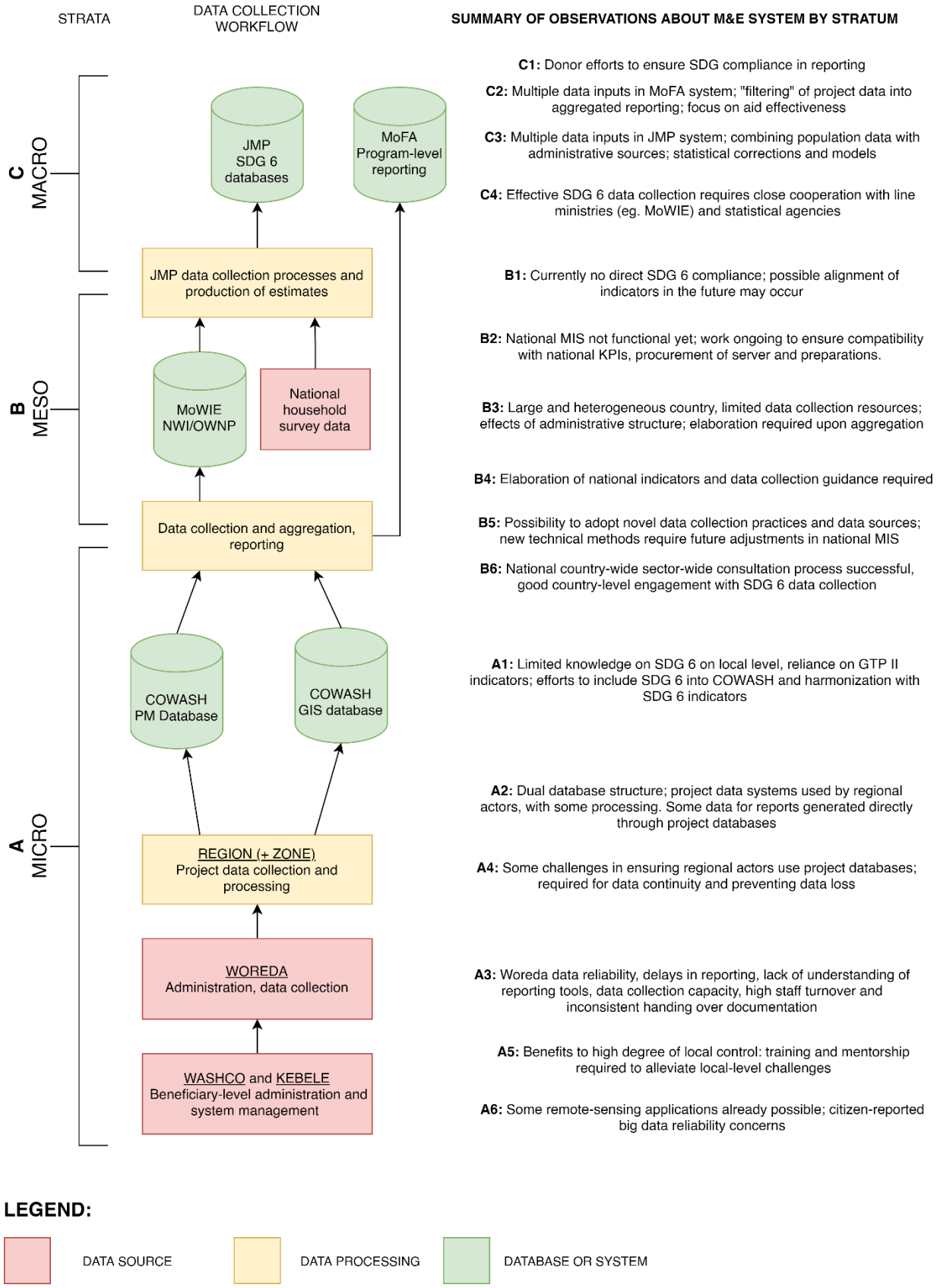

Figure 2: Main findings and themes raised in semi-structured interviews, contextualized by strata in the analytical framework

4.1 Integration of SDG 6 in the case study context

Based on interview results, there are varying degrees of SDG 6 integration on the micro, meso and macro strata. Overall, it seems that there is a clear impetus towards harmonizing existing systems with SDG 6 wherever this is possible. However, different institutions are at wildly different stages of adopting SDG 6-compliant indicators, and informants seemed to slightly disagree as to the degree to which this is harmonization is desirable.

Informant 1 (regional M&E specialist) stated regional-level actors (micro stratum) are not very knowledgeable on SDG 6, and the implementation and data analysis are still based on GTP II indicators (Fig. 2 A1). The effects of the SDGs on data collection and analysis are thus not yet clear on the regional level according to the informant, since the data mostly processed at this stage still uses the national GTP II system, and it is difficult to tell what the effects of the SDGs are.

Informant 2 (national M&E specialist) said that there have been efforts to integrate SDG 6 into the reporting system used by the COWASH project, and currently there is partial compliance with these indicators (Fig. 2 A1). Since the project has to report data both to national authorities and routine monitoring reports, there is a strong incentive to harmonize the data collected with SDG 6. For instance, disaggregation by SDG service level in data collection has been adopted to ensure compliance. However, since there are limited resources when it comes to monitoring, there is an inherent cost to adopting the SDGs directly. The donor (MoFA) opinion has influenced the move towards SDG 6 compliant data wherever this is deemed possible.

At the national level (meso stratum), the situation when it comes to SDG indicator harmonization is rather complex (Fig. 2 B1). Informant 3 noted that the indicators used by the Government of Ethiopia (GTP II) are currently not directly SDG compliant and speculated that achieving some of the SDG 6 targets might be difficult, even though there has been progress made. Informant 4 also noted that the national indicators are not directly SDG compliant, but also stated that this is not necessarily a problem since the JMP has its own data collection methodology so the two do not need to be in direct alignment with one another. However, according to this informant, there has been discussions that the national norms and service levels will move towards more SDG-compliant direction when they are revised in the future, and the GTP II rural service levels are already rather close to those noted by the SDGs when it comes to results generated with these to different methodologies.

On the MoFA system (macro stratum), Informant 5 noted that a lot of effort in the MoFA has been placed into ensuring that the indicators are SDG 6 compliant, and that the indicators used at different stages of the data system in the MoFA system are as close to the SDGs as possible (Fig. 2 C1). However, there are necessarily delays in this transformation since it requires considerable time to change the indicators in ongoing interventions. In general, however, informant 5 noted that the efforts have been rather effective, although data between interventions is still currently not entirely comparable horizontally, as there are different types of indicators in use in different projects and programmes. These observations were confirmed by Informant 6, who stated that the indicators currently used across the board have been rather consistently harmonized with the SDGs, and the water

sector additionally has comparatively well-established globally agreed indicators than those used by some other sectors.

4.2 Effectiveness of the M&E systems in use across scales

Interview results show that there are many differences in monitoring systems and tools used across the three strata, and as can be expected, there is evidence of differences in technical capacity and resolute planning efforts between the strata. In general, actors at higher strata have more knowledge of effective data collection practices and recognition of the importance of effective M&E systems, and the systems are generally more thought out and elaborate at higher strata as more data aggregation takes place. This means that there seems to be greater capacity on higher strata to account for errors and inaccuracies in data originating from lower strata.

This does not categorically mean, however, that the systems at higher strata are necessarily more complex, although there are by necessity more data inputs into the systems at the higher strata. It seems that institutions on each stratum have designed systems that are quite functional for the requirements and purposes of that stratum, but there is insufficient coordination between the strata.

4.2.1 M&E system on the micro stratum (COWASH project)

Informants on the micro stratum (informants 1 and 2) raised generally quite similar issues regarding the functionality and effectiveness of the M&E system in the COWASH project. This was expected, as both informants are project M&E staff.

In the context of the COWASH project's regional M&E work, Informant 1 (regional M&E specialist) noted that the regional administrative level provides a results-based report using a set outline to the federal office, in which progress is evaluated against quarterly and annual plans (Fig. 2 A2). Project data systems (Planning, Monitoring and Reporting Database / PMDB, web-based Facility Database) is used for this reporting, with the PMDB mainly used to store activity and indicator data. Data is mostly processed using Excel spreadsheets and Microsoft Word, with some data generated from the Facility Database (such as women's participation in WASHCO membership and key positions).

The most important challenges in terms of data quality and sources of error noted by Informant 1 (Fig. 2 A3) are lack of reliable data from *woredas,* delays in report preparation to regional staff, lack of understanding of the reporting tools and lack of capacity in data collection, high staff turnover, as well as inconsistent handing over documentation. At the local level, indicator data is not always seen as an important resource, which results in inconsistent and incorrect data handling practices and may result in data errors when data is aggregated. The solutions to these problems lie mainly in better and more regular training for data collectors in use of data collection tools, and preparation of simple and understandable tools. The dual database structure adopted by the project is useful in addressing possible inaccuracies. The informant also noted that COWASH would benefit from having its own M&E staff at the local level to minimize the problems with data quality and delays, and to have better data storage and analysis systems at the *woreda* level.

The findings from the interview with the national COWASH M&E specialist of COWASH (informant 2) seem to mirror the regional state-level observations by Informant 1. Informant 2 regarded the systems currently used by COWASH (PMDB

and Facility Database) are well-designed and quite simple to use. This dual database structure is also interesting for purposes of analysis since it allows for more elaborate monitoring conclusions to be drawn, such as inspecting distribution of facilities vis-à-vis indicator data reporting.

In terms of challenges, Informant 2 noted (Fig. 2 A4) that there have been challenges in convincing regional staff to adopt the new databases used by the project, and some would have preferred to use other types of tools (commercial tools such as Excel) instead of the proprietary software currently in use, as these other tools seen by some as more convenient and simpler to use. However, this increases the risk of data loss considerably (especially in situations of personnel changes), and integration of systems across different administrative levels becomes very difficult if there are different systems at use on different levels. Informant 2 also raised the problems of data delays and incorrect encoding of data at the local levels of administration (ie. *kebele*, *woreda*) where beneficiary data is collected. According to informant 2, there is also problems in terms of understanding the importance of data custodianship at the local level – data collection is seen as a burden and an additional activity, which means that there is sometimes little incentive to invest sufficient efforts into good quality data collection practices. Informants on the national stratum (Informants 3 and 4) also noted that high staff turnover and lack of documentation at the *woreda* level can cause problems for data collection and noted that this causes problems at the local administrative level in terms of data reliability.

4.2.2 M&E system on the national stratum (national M&E system)

Observations on the functionality and effectiveness of the M&E system used at the national stratum were made mainly by the national M&E specialist (Informant 2) and the two MoWIE civil servants (Informant 3 and Informant 4). All of the informants noted that the OWNP Monitoring Information System (MIS) that has been under development under the last few years was still not functional at the time of the interviews.

Informant 3 noted that NWI Phase II has updated the baselines for WASH sector data, but otherwise the system is not currently functional (Fig. 2 B2). The same informant also noted that the planning and budgeting of the new MIS system is underway, but currently, there are still some actors using older paper-based (Excel, Word) reporting systems, and adoption of the new system is still ongoing.

Informant 4 noted that the NWI II used a system called COSMOS provided by the British consulting company Coffey Ltd for processing and storing data, and there is work underway to transform this system into being compatible with national Key Performance Indicators (KPIs) devised for the OWNP. This system allows for generation of various reports and maps and provides various other synergies with the data requirements of the national authorities. Requirements for the new system are now known, and the next step will be the procurement of a national server to host the data (since the data is currently located on a server abroad). After this, data transfer to the new system can take place, and the system can be better modified to fit national data usage needs.

Informant 2 noted that delays with the establishment of the national MIS system can be a problem down the line in terms of responsiveness since data is not distributed in a timely manner across the different levels of administration, and

with international partners. The informant also noted that the use of the nationally collected data is not very widespread at present, although the design specifications and structure of the prospective system (MIS) are compelling and well thought out. Informant 4 also noted that the system is not responsive at present, and there is a data backlog of approximately 2 years (since the data for NWI II was collected).

Interview participants had different opinions on the sources of errors in the data aggregated on the meso stratum, and the reasons for these errors (Fig. 2 B3). Informants 3 and 4 both noted that one major reason is the country's size and heterogeneity, and the fact that resources for collecting data are limited, which causes difficulties for collecting data in a uniform manner throughout the country. The administrative structure in place does not always allow for top-down interventions in terms of adopting a certain type of methodology or practice in data collection at the regional, zonal and *woreda* levels. Subsequently, levels of training and capacity differ considerably from place to place, despite the fact that efforts have been made to increase knowledge on effective M&E practices. Informant 3 also noted that documentation at the local level is often quite poor, requiring elaboration from the *woredas* and more work at the data aggregation phase.

However, both Informants 3 and 4 noted (Fig. 2 A5) that there are benefits to a high degree of local control in managing their own systems, and that local communities having control of the interventions should be encouraged. In this, mentorship and training can be used to alleviate the abovementioned challenges at the local level. This was also noted by Informant 7, who stated that federal governance models tend to be complicated in their execution, but there are also good reasons for devolution – there are necessary trade-off when it comes to building national systems.

Informant 2 noted (Fig. 2 B4) that the national indicators (and the respective data collection guidance) are currently not yet clear and should be elaborated, citing the example of water point functionality, for which data is collected during the rainy season in some *woredas* and the dry season in others. The data collection practices for these parameters are not uniform, resulting in inherited unclarities in data collected.

In terms of the adoption of more responsive, novel data collection practices (Fig. 2 B5), such as big data collection from households by smartphone use, Informant 3 noted that this could be a possibility in the future since mobile data coverage has been increasing considerably, and it is important to recognize on the national level that new technological opportunities in monitoring are emerging. Informant 4 stated that the new system is being developed in such a way that additional functionalities, such as big data integration, becomes possible, but this is not planned once the system becomes functional. Informant 2 noted that this type of data collection may become viable, but at present the national M&E system is not equipped to handle this type of data (and neither are the systems used by COWASH and other similar projects), and the system must be designed deliberately to allow this type of operation, otherwise there is risk of data loss or poor data reliability. As such, informant 2 did not see this as a feasible solution at the present moment.

4.2.3 M&E systems on the macro stratum (JMP and MoFA)

Information on the functionality and effectiveness of the donor and UN M&E systems were provided by Informant 6 (Ministry of Foreign Affairs reporting) and Informant 7 (JMP reporting). Since the MoFA and JMP systems are separate and serve different functions, they will be discussed separately.

Informant 6 (Fig. 2 C2) said in the interview that the MoFA results management and M&E system has a lot of inputs from different sources (country programmes, multilateral initiatives, NGOs, and the private sector) that are integrated into the data aggregated for MoFA reporting. As such, it is necessary to abstract results into higher-level impact goals (outputs and outcomes) that are less concrete than those at the level of the individual interventions. While results are "filtered" to higher levels of aggregation, project data is utilized when results are aggregated, and the types of indicators used in partnerships have been analyzed when the systems have been designed. In practice, annual reports from different sources are read by a civil servant at MoFA, after which the data is aggregated into the IT system used by the ministry. There have been efforts to ensure that while the figures are inputted, the person inputting the data has the opportunity to analyze the background to the results in as much detail as necessary, but the system also requires the data to be crystallized when it is abstracted. Although data is inputted into the system constantly, reporting with this data chiefly occurs every four years, or once in a Finnish parliamentary term.

The data collection and use requirements by the donor (MoFA), according to Informant 6, are mostly concerned with assessing the overall effectiveness of Finland's aid to the government and public, as well as to ensure international cohesion in terms of effective results-based management. This requires inspection at both the level of individual interventions, as well as that of country programmes and priority areas. This is then used to also evaluate the strengths of Finnish aid, as well as the operational areas which are most successful and most challenging. Methodological harmonization is conducted by MoFA, where indicators used by partners may be adopted into the MoFA system if they are seen as effective. The harmonization is being carried out both at the country programme level and sectorally, allowing for constant improvement of the system.

The JMP (Fig. 2 C3), according to Informant 7, also uses a variety of different data sources, but the JMP reporting relies heavily on population-level data in addition to administrative data sources. The datasets received from different sources may have different scopes and levels, there might be data outliers, miscoding or misclassifications in the survey data used. Furthermore, the units of analysis may be different between the different data sources, and administrative data sources tend to use households as the unit of analysis while the JMP is concerned more with the population (individuals). The JMP uses various statistical corrections to mitigate for differences in methodology, and the regression methodology used to produce estimates is effective at achieving this, and it allows for estimations and projections to be made from available data. Informant 7 noted that since there is a considerable amount of data inputs, the temporal lag in data access can be somewhat mitigated. There might be, however, a lag in the processing of data before it is usable by the JMP.

In some countries, according to Informant 7, there can be problems in terms of data quality, and that collection of reliable data can sometimes be challenging.

The informant underlined the importance of clear communication between the national data custodians (ministries and statistical agencies) and lower levels of administration, in order to ensure effective data flows.

Informant 7 noted that more robust SDG monitoring requires even closer coordination with the central statistical agencies and line ministries (like MoWIE) to ensure that all relevant data sources can be utilized where necessary. A fundamentally important feature that must be guarded, however, is that statistical agencies must be independent and have strong capacity in their operations. For NWI II, Informant 3 noted that the Central Statistics Agency of Ethiopia was involved with the process of data collection, and the data was useful in providing data for Ethiopia's country report, which is prepared by the statistical agency (the data custodian for JMP data).

In terms of the Ethiopian context specifically (Fig. 2 B6), Informant 7 perceived the country consultation process in terms of SDG monitoring as fruitful and stated that the national data custodians have engaged well with the data collection for the SDGs. The national stakeholder consultation process related to OWNP has been well thought out and enthusiastic, but it seems that the momentum has slowed recently. There was also a lot of enthusiasm around NWI II data collection and the data resolution (building an accurate baseline on the level of each household) as very enthusiastic, but it has been regrettable that there have been no updates to the data after the inventory was carried out.

Informant 7 noted (Fig. 2 A6) that some elements of SDG data (such as surface water quality) and some household data points can already be or may in the future be analyzed by remote sensing technologies or big data, respectively. With the latter, however, data reliability may become a concern: use of citizen-reported figures has to be very well planned and having large datasets with a lot of incomplete data points may cause problems in terms of data usability. For project or national data use purposes, however, the informant noted that this type of data might be very interesting and useful, even if they are not useable for the purposes of the JMP system.

4.3 Indicator correspondence and compatibility of M&E systems

4.3.1 Indicator correspondence analysis

In order to analyze the compatibility of indicators used at each stratum, a classification system for types of indicator correspondence was devised, illustrated in Figure 3. This provides a common framework for analyzing the relationships of the strata with each other, and to provide information on the ways in which indicator data is transformed between the three strata. Devising this type of classification was seen as necessary, since there was no existing methodology for objectively classifying indicator transformations in the literature.

In this classification system, five types of possible indicator transformations have been identified: 1) direct indicator correspondence between strata; 2) aggregation of several indicators into a composite indicator at a higher stratum; 3) disaggregation of a composite indicator to several individual indicators at a higher stratum; 4) combination with a supplementary data source / indicator at a higher stratum; 5) aggregation of data to a higher stratum without direct indicator correspondence. In terms of the functionality of the M&E system, it is assumed that there

ought to be some level of cross-strata coordination, and the analysis seeks to ascertain the degree to which the data is transformed when moved to a higher stratum. It should be noted that these five types are not exhaustive and may abstract from the nuance of data transformations undertaken when data is placed in a new repository or MIS, but it does capture the general compatibility of the indicators with each other.

The results of the indicator correspondence analysis have been illustrated in Figure 4. This figure summarizes the interdependencies between indicators that can be identified from project documentation, programme documentation at the national level and the level of the donor, as well as technical documentation on SDG 6 estimation by the JMP.

The analysis shows that the Micro stratum (the COWASH Phase III project) is, as noted in the results from interview data, in direct alignment with the national M&E system used by the Ethiopian federal government, as well as partially/indirectly aligned with the M&E system used by the donor (MoFA). The project collects data in its results framework using the Project Management Database (PMDB) that is directly compatible with the GTP II indicator used by the Government of Ethiopia. Efforts have, therefore, been made at the project level to harmonize the indicators with both systems, allowing for relatively smooth estimation of results between these actors' systems, and at present the systems are partially compliant. This was confirmed by Informant 2 in the interviews.

The type determined for the COWASH-GoE transformation was deemed as direct correspondence (Type 1), while the COWASH-MoFA transformation was deemed to represent mainly Type 1 with some elements of aggregation without correspondence (Type 5). The reason for this is that there is a very high degree of data aggregation from a variety of sources in the MoFA figure (including data from country programmes, individual projects, NGOs, and private-sector actors). The ultimate estimate includes certain elements that cannot be directly described using the top-level indicators, requiring certain data transformations by the donor. However, the data collected by COWASH is directly compatible with the indicators that MoFA uses, and this seems to be the norm in similar project interventions financed by MoFA. This was also confirmed by several informants in the interviews, including Informant 2, Informant 5 and Informant 6.

It is notable that, at the time of writing, the OWNP MIS, which will form the basis of the data system for M&E at the national level, is still not functional. As such, the data transformation between the COWASH project and the new national MIS cannot be reliably analyzed. However, since the background documentation of the OWNP lists the GTP II indicator already in use by the MoWIE as the prospective KPI for this system, it is assumed that this indicator transformation will also represent Type 1. However, since it was revealed in the interviews that there is debate on whether to revise the indicators used for this aspect of M&E monitoring in the near future, it is possible that this might change. Furthermore, since the new MIS allows for the integration of new types of monitoring tools and methods, it is possible that the new data inputs into the system may change the transformation when the system is operational.

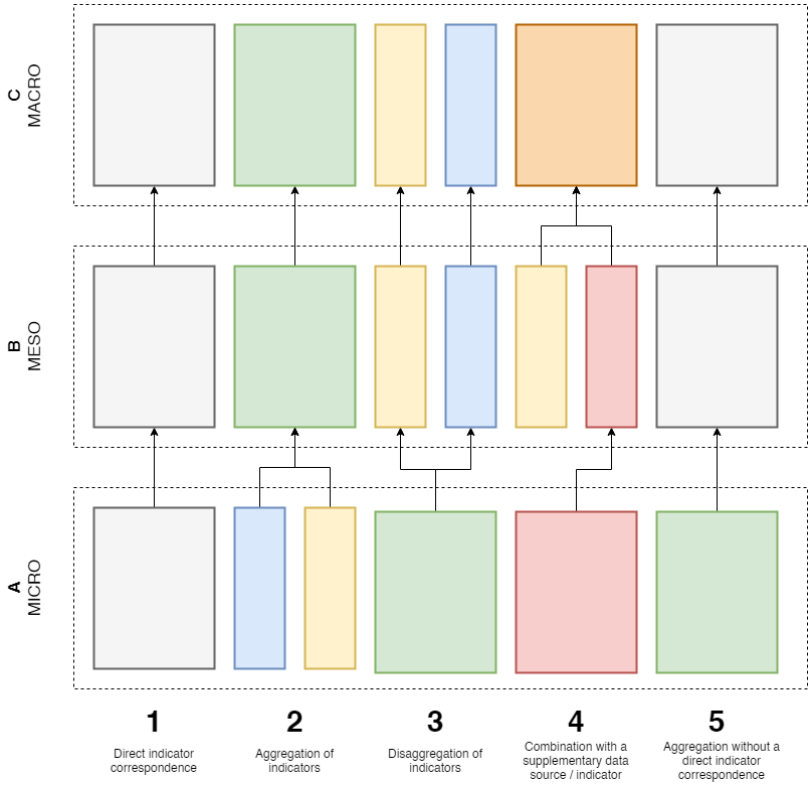

Figure 3: Classification of the types of indicator correspondence and possible data transformations, illustrated on three example levels of data aggregation

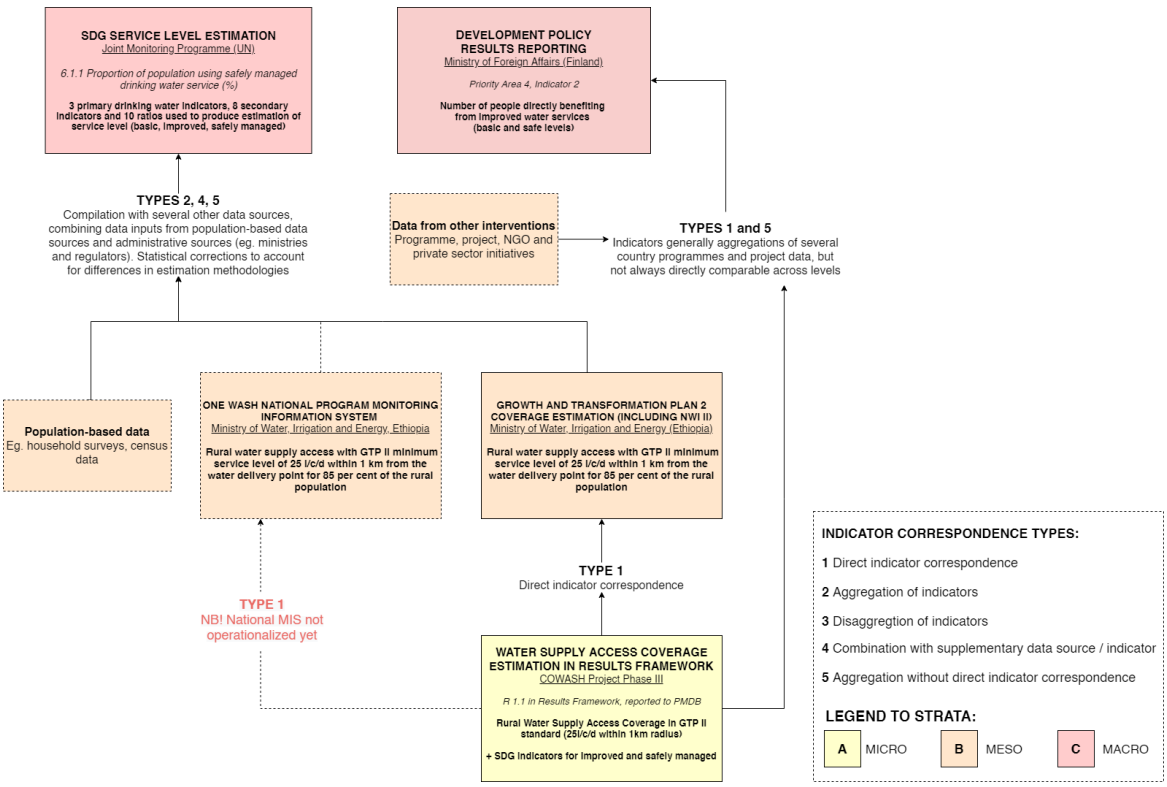

Figure 4: Results of the indicator correspondence analysis for indicators estimating rural household water access coverage across the three strata

On the Macro stratum, as stated earlier, the indicators used by MoFA are mostly the same as those used by JMP, although the JMP indicators use a slightly more refined estimation methodology with several more sub-indicators to construct estimates, as well as vastly different data sources. MoFA, in its overarching Development Policy results reporting, uses a KPI to measure household water access coverage that are directly compatible with SDG 6 reporting (Priority Area 4, Indicator 2), which measures the "Number of people directly benefiting from improved water services" on "basic and safe levels" of service delivery (Ministry for Foreign Affairs of Finland 2018). Data for these is collected by aggregating data from a variety of interventions that MoFA finances, and these results are aggregated together. However, MoFA reports the number of people who benefit from MoFA financed interventions of different types, while JMP reports these as a share of the population on a certain service delivery level at a country level.

In contrast to the figures calculated by the donor, both population-based data sources (eg. household surveys) and administrative data sources (eg. ministry and regulator data) are used to generate the JMP estimates. The JMP calculates the share of the population using improved water sources (as opposed to unimproved) and categorizes the safely managed service delivery levels into the main service level categories "limited drinking water service", "basic drinking water service" and "safely managed drinking water service". This is achieved through aggregating three primary indicators estimated directly from data inputs (improved drinking water sources, piped drinking water sources, and no drinking water facility / surface water), which are then used to generate eight secondary water indicators and four ratios in relation to the population using improved drinking water sources (WHO/UNICEF JMP 2018).

Therefore, the calculation of these service levels can be summarized using the following formulae 1-7 as follows:

$$W_u = 100\% - W_i - W_{sw} \quad (1)$$

$$W_l = W_i \times (100\% - \frac{W_{ct}}{W_i}) \quad (2)$$

$$W_b = W_i \times \frac{W_{ct}}{W_i} \quad (3)$$

$$Ws_1 = W_i \times (\frac{W_p}{W_i}) \quad (4)$$

$$Ws_2 = W_i \times (\frac{W_{wn}}{W_i}) \quad (5)$$

$$Ws_3 = W_i \times (\frac{W_{fc}}{W_i}) \quad (6)$$

$$W_s = (Ws_1 \geq Ws_{1\,min}) \wedge (Ws_2 \geq Ws_{2\,min}) \wedge (Ws_3 \geq Ws_{3\,min}) \quad (7)$$

where

W_u, W_l, W_b and W_s are the drinking water service levels unimproved, limited, basic and safely managed, respectively;

W_i is the proportion of the population using improved drinking water sources,

W_{sw} is the proportion of the population with no drinking water facility (surface water);

W_{ct} is the population using improved water sources not exceeding 30 minutes collection time;
W_p is the population using improved water sources accessible on premises,
W_{wn} is the population using improved water sources available when needed;
W_{fc} is the population using improved water sources free from contamination;
$Ws_{1\text{-}3}$ are the three secondary indicators used to calculate the W_s figure; and
$Ws_{1\text{-}3\ min}$ are the minimum threshold requirements for each of the three criteria.

In other words, the JMP estimation methodology calculates the ratio of improved and unimproved water sources used by the population; four ratios relating to the collection time, accessibility, availability, and lack of contamination of these improved water sources; which are then used to determine what share of the population has access to limited, basic or safely managed service levels. To satisfy the safely managed criterion, an improved water source must be on premises, be available when needed *and* be free from contamination, and the minimum of each three indicators for a given year must all be fulfilled *at the same time* to satisfy the components of this highest service level criterion as a whole.

To conclude, there are two types of fundamentally different estimation methodologies in use in the wider M&E frameworks relating to the COWASH project: nationally mandated GTP II based service level estimation with a minimum threshold, as well as the SDG 6-based service level estimation methodology that has been partially modified to suit the context of the project and the donor. While there are certain incompatible stages in the data aggregation workflow, certain statistical corrections and several data sources are used to account for the fact that indicators are not directly comparable in some stages of the aggregation workflow (Types 2, 4, 5). In addition, there are vastly different data collection methodologies in part because the data is collected from different sources and for different purposes.

5 Discussion

This section discusses the implications and meaning of the results obtained from the semi-structured interviews and indicator correspondence analysis.

The discussion of key findings in in Section 5.1 has been organized into four main observations that arise from the results. Section 5.1.1 focuses on the issues of vertical integration in the M&E system, and possibilities for increasing cohesion in M&E across strata (research question 3). Section 5.1.2 discusses the role of local communities and their empowerment for better M&E and data collection (research question 2). In Section 5.1.3, the future developments for SDG 6 and the possibility of increasing its applicability to local communities is considered (research question 1). Finally, Section 5.1.4 dives into the thematic of responsive and real-time monitoring through alternative data sources and acquisition methodologies (research question 2).

Section 5.2 discusses the limitations of the study, as well as the persisting research interest related to M&E, particularly in the rural Ethiopian water sector. Some of the wider implications of the results for development cooperation interventions are also briefly touched upon.

5.1 Discussion of key findings

5.1.1 Vertical integration and cross-strata coordination

After inspecting each of the three strata and the M&E systems in use at each, it becomes quite clear from that coordination between the system has been discussed in relation to the case study context, and there is clearly a desire amongst the various stakeholders related to the project to pursue more unified systems. This is demonstrated by the fact that interview participants, particularly on the macro stratum but also on the micro and meso strata, were aware of the interconnectedness between the different M&E systems that affect projects like COWASH.

One is therefore inclined to argue based on the results of this study that since data is being generated in project interventions, it only makes sense to use this data in a meaningful way where possible. This is further accentuated by the fact that the SDGs include a unified framework within which analyzing data can be carried out with some degree of cross-strata cohesion, at least to some degree.

Despite this shared momentum towards methodological cohesion and the rather convincing effort on the national stratum to include all stakeholders in the planning process for Ethiopia-wide reporting and monitoring (through the OWNP SWAp, in particular), there is still work to be done in terms of looking at the different components of the data chain holistically. It is also rather difficult to find studies where project data collection is used in analyzing the national and global context, or on how project data could be better utilized in alternative applications. While the literature calls for this on a theoretical level, there is still limited evidence that this is in fact occurring in practice, in the Ethiopian context or in Sub-Saharan Africa at large.

Since considerable effort is placed on developing M&E at the micro, meso and macro strata in the case study context analyzed, a natural follow-up question is the degree to which the adoption of practices from different strata (and different institutions therein) could be beneficial. Since it seems that the macro stratum actors

(JMP and MoFA) already pursue a rather good practice of integrating new indicators and/or data sources into their own systems, might it be useful to inspect *all* the strata together when systems are being reformed?

All that being said, it is of course to be recognized that analyzing all stages of a data collection workflow is extremely challenging. Indeed, this study was also forced to considerably abstract the stages of data aggregation, since not doing so would have taken a prohibitively large amount of time. However, it would likely be beneficial for data producers and data custodians in Ethiopia, especially the national government and project actors, to be more mindful of the various interlinkages between the data production, processing, and storage. Whether this comes in the form of better SDG integration, harmonization of databases for ease of data transfer, or closer and more routine coordination between different stakeholders, closer attention should be placed on ensuring that each of the parties that collects similar data in parallel are aligned wherever this is feasible or possible. This is also the case in instances where aggregate data that is used indirectly at higher stratum. There are certainly mutual interests and synergies from sharing data, and the systems used by different parties should be linked together where this can be realistically achieved.

One reason why there has not been a lot of consistent analysis of all aspects of the different M&E systems together is likely the fact that the number of data inputs also increases as data aggregation occurs. As such, the data tends to assume more a form of a *web of aggregation* in reality, rather than a hierarchy of aggregation as is implied by the discussion in this study that is based on one individual case.

When looking at the vast, interconnected, and massive system, it is clear that complete harmonization is likely impossible to achieve in the COWASH project or Ethiopia as a whole. This mirrors the conclusion reached by Butterworth et al in terms of improving M&E system performance and cohesion between different scales of M&E discussed in Section 2.2.1 (Butterworth et al. 2013). For this reason, the OWNP with its nationwide and stakeholder-wide consultation process is an interesting case to consider. Any efforts to approach methodological harmonization that are expected to be successful and widespread should probably be coordinated on the meso stratum (national level) in Ethiopia in the future, not least because of the fact that the SDGs place so much emphasis on national actors in monitoring and reporting of results. In terms of developing the national M&E system, the advice provided by the IRC framework (IRC 2016) provides a useful reference of how this should take place, and there needs to be holistic and concurrent development of all aspects of the system in unison.

5.1.2 Empowerment of local communities in M&E work

The theme of increased community involvement in matters of administering local water resources arises frequently in the literature, and bottom-up development cooperation approaches are championed by national governments and the international donor community alike (da Silva Wells, van Lieshout, and Uytewaal 2013). This is no different in Ethiopia and in the case of the COWASH project, where the CMP approach to project management and service delivery in WASH interventions has proved effective and popular in increasing local agency and management of rural water resources (Suominen and Rautiainen 2016; Behailu, Suominen, and Katko 2015).

This finding in the Ethiopian context is in alignment with Giné Garriga et al, who found that policy-relevant data collection and analysis needs to be complemented by capacity-building for local government entities in decentralized systems (Giné Garriga, Jiménez Fdez. de Palencia, and Pérez Foguet 2015). Furthermore, it also has a connection with the WASHCO legalization efforts and capacity development closest to the beneficiary level, mentioned in the project document for COWASH Phase IV as a key area of organizational development (Impact Consulting Oy Ltd 2019). Other areas of organizational capacity-building are intimately linked with efforts to improve data quality and appropriate methods of data collection, and they must not be thought of in isolation from one another.

There is still work to do in achieving operational effectiveness at the most local level, and in ensuring that the resources produced in interventions like COWASH achieve their end goal. In a recent study, Marvin explored the sustainability of operations and maintenance in the Ethiopian water sector, also in cooperation with the COWASH project. There is a clear linkage in the policy environment between effective O&M practices and effective M&E practices in achieving the safely managed service delivery levels for household water called upon by SDG 6.1 (Marvin 2021). This is all the more relevant in a federal country like Ethiopia, where regional and local levels of administration enjoy considerable autonomy and constitutional protection (Yilmaz and Venugopal 2008).

The empowerment of beneficiary communities (and the WASHCO, *kebele* and *woreda* levels of administration closest to those communities) should be seen in holistic terms when it comes to water sector development cooperation, and good quality, reliable data should be seen as a prerequisite for increasing local capacity in effective management of water resources. As such, it is not sufficient to assume that good planning and well-designed systems will be sufficient in increasing the capacity of the local level to assume an increasing operational role, but training and awareness-raising play an important part (Person et al. 2017).

It would be ideal would be to have an even more concerted effort in the context of the CMP project management approach to increase the awareness of the importance of well-documented results that are backed up by good quality data, backed by well-designed and easily comprehensible data collection methodologies. Although there is evidence based on interview results that this type of mentorship is currently already occurring on the ground, there should be more attention paid to the transferral of solid organizational practices and awareness when it comes to data. If ambitious results are expected on the ground, there should also be a correspondingly ambitious push towards realizing the ambitious principles of results-based management. It can be inferred from the findings of this study that better data goes hand in hand with better professional expertise on the ground.

5.1.3 Use and application of SDG 6 across different levels – way forward?

Based on the results, there is certainly some merit to the notion that the Sustainable Development Goals must be harmonized with existing M&E systems. It is also desirable that the M&E practices in use across different spheres of analysis share a common reference point when it comes to the terms at which progress is measured. However, it should also be noted that all systems need not categorically shift to using SDG 6 indicators directly and across the board and at all levels.

There are certainly instances where the use of other types of indicators for WASH interventions is merited and may even be more desirable than adopting the SDG 6 indicators directly, since the data use requirements vary considerably across the different strata analyzed. For example, if another domestic indicator is deemed more fit for purpose for national policy coordination or provides better decision-making support in some case, it may not be relevant to use SDG-compliant indicators blindly and for all cases.

It was mentioned in the interviews (with micro and meso-level informants in particular) the Ethiopian government uses data for quite different purposes than the JMP or donor partners. Indeed, it is an entirely different exercise to collect water sector data for national-level development planning and coordination of government resources than for country level SDG 6 progress evaluation. While the SDGs are a *raison d'être* of sorts for the JMP, they are not domestic authorities. This is further different from the accountability to the public and to the Finnish political establishment and faced by MoFA in its development policy-wide reporting. MoWIE has an entirely different use and purpose for data than the UN authorities or the donor government do. Since resources for M&E are limited, one should prioritize data collection based what institutional needs are and what is manageable.

All that being said, it is also an interesting and rather philosophical exercise to think about the ultimate *purpose* of the SDGs, and to what degree they may be seen as belonging to the beneficiaries that project interventions like COWASH are seeking to empower and assist. While the SDGs have been set by national governments at the level of the UN, the discourse on them seems to simultaneously imply that they are a universal system of goals and indicators. Furthermore, the Integrated Monitoring Guide for SDG 6 calls for extensive consultation of actors across different levels in implementing SDG 6 monitoring nationally, all while acknowledging their "global and aspirational nature" (UN Water 2017a).

If the SDGs are indeed universal, should they be understood – and utilized – by the lowest levels of local service delivery, coordination, and management? If this is the case, results from this study suggest that there is still work to be done before we achieve truly universal goals that are not merely for the use of the international community and donors, but for beneficiary communities as well.

This type of perspective strikes a rather post-developmental tone in its underpinnings, which is to say, understanding the need to transcend the traditional development cooperation's realm of rigidly imposed mechanisms and top-down imperatives by Western donors (Matthews 2010). However, this critique need not be about trivializing the existing mechanisms themselves, but rather critically analyzing the purpose and role of frameworks like the SDGs. In the case of SDG 6.1, for instance, pursuing a considerably more refined coverage estimation methodology at the international level is in contrast to persisting issues of M&E capacity and knowledge at the lowest levels of execution, as it seems based on interview data that organizational and capacity issues at the micro stratum in the case study context have not always been resolved. As such, when the SDGs are reviewed, integrating a more participatory and inclusive perspective to measuring them, in addition to high ambition and high-quality data collection, should be considered as a possibility.

5.1.4 Towards a more responsive and real-time M&E system

One clear conclusion from this study has been that the national OWNP-derived M&E system and the associated MIS system needs to be finalized before utilizing alternative data sources. Only then can the various novel data collection practices or new data sources such as big data from mobile phones, data collected from in-situ sensors embedded to water points, or remote sensing data from satellite imaging, can be seamlessly integrated with the national-level data that is being currently collected, as was discussed by Thomas et al (Thomas et al. 2018). But since it seems that efforts towards adopting this system will begin to bear fruit sometime in the near future, might it already be time to prepare for integrating these new, hopefully more real-time data sources with the more traditional monitoring practices currently in use, as the literature sources call for?

It is evident that these types of alternative data sources will provide a good supplementary source of information for more responsive and results-based decision making across the different levels. For project interventions, they may allow for a considerable degree of agility in allocating resources where they are most needed and allow for real-time planning related to possible corrective actions if problems are observed. At the national level, these data sources can act as a way to empower the local level and households at voicing their concerns and may provide insights on phenomena that may otherwise fly under the radar from traditional monitoring.

However, a clear underlying assumption in the adoption of these novel data collection practices is that there has to be a well-designed and robust data repository to store these types of data, which requires a lot of deliberate planning. This is discordant with the fact that there are a lot of calls for adopting these types of practices in the international donor community and in the literature. In the case of Ethiopia, while likely much more plausible in the future, widely adopting these types of practices does not seem very feasible over the next few years, as there is still work to do in the back-end development of the new M&E tools that are only beginning to be rolled out.

That being said, there should be ambition towards trialling and testing new approaches to collecting data, as it is difficult to know based on theoretical speculation what methods will in fact produce good alternative data sources. Once the new MIS system for WASH data is functional, and if there are resources at the level of projects such as COWASH to facilitate these types of tests, it is possible that they may prove to be useful sources of supplementary data.

There are examples of practical applications for alternative data sources in other similar sectors of rural development cooperation, such as agriculture. In a recent report on the twelfth replenishment of the International Fund for Agricultural Development (IFAD), it is recognized that novel ICT applications are key in improving efficiency and reducing vulnerability in agricultural development – there have already been cases where combining GIS modelling, earth observation data and social vulnerability assessments have be used to better target interventions and “tailor infrastructure adaptation plans to local risk levels and needs” (Hartman, Williams, and Grenra 2020). A 2019 report by the European Union-funded Technical Centre for Agricultural and Rural Cooperation ACP-EU (CTA), a joint institution between the African, Caribbean, and Pacific Group of States (ACP) and the European Union, has discussed in depth the various practical benefits of

digital applications in Africa's agricultural transformation, such as access to remote agricultural extension services and improving market access (Tsan et al. 2019). These types of cross-sectoral comparisons may not provide a directly applicable route to innovation in data collection and management for more results-based and real-time monitoring for WASH projects. However, combining several data sources across different levels holds promise for providing a better picture on the ground.

What is, however, clear is that these types of novel approaches will likely never completely replace conventional *woreda*-level data collection and processing by sub-national and national authorities in Ethiopia. The M&E methods used by all the actors in this case study context build upon decades of experience, and there are also many merits to empowering local communities and project staff at levels closest to service delivery in managing results. But time will tell whether the novel data sources will indeed help in bridging the long gaps currently observed in data collection, or whether it will be more behind-the-scenes system improvements and methodological harmonization that deliver this end goal instead. Indeed, terms like "big data" and "remote monitoring" are appealing and interesting because they seem so simple but implementing such a system in practice is a very difficult exercise that needs a plethora of resources and planning.

5.2 Limitations, future research interest and wider implications

Although there are several limitations to this study, it is clear that it also provides an opening for further inquiry, both into the context of development cooperation work in Ethiopia specifically and the methodology of assessing the performance of M&E systems more generally.

It is important to note that this study has, by necessity, provided a rather limited snapshot of the features of the M&E systems in relation to the COWASH Phase III case study context. While it provides a comprehensive general overview into the M&E systems that are directly related to COWASH, it is but one project intervention in a country with several initiatives of various sizes and scopes. It is fair to approach this study as an abstraction of the reality of the M&E practices, although one might note that M&E systems themselves also always abstract a complex reality into a more manageable form for purposes of analysis. Nevertheless, it does not provide the conclusive and exhaustive picture of all aspects of the M&E systems, but only begins to scratch the surface on underlying phenomena.

This study has been confined to remote interviews without a visit to the field, due mostly to the travel restrictions placed on travel due to the COVID-19 pandemic that has been ongoing during the time of writing. As such, the interviews are concentrated on levels above the regional state level, without direct interview data from zone, *woreda*, *kebele* and WASHCO levels. It is possible that the results on capacity of local actors, for instance, might have been slightly different, had this study included an interview from these local levels. However, the informants interviewed have been very helpful in providing enough insight into these so that certain conclusions could be reached.

The number of interview participants is still only seven informants, although participants from various different levels have been included in the analysis. This was mostly due to the time constraints placed on thesis work, although including evidence from more informants would have likely provided a more complete view of the complicated phenomena related to the case study context's M&E systems.

Furthermore, only one set of indicators was used in the indicator correspondence analysis (rural household water access coverage), and it is possible that indicators used for other purposes (such as sanitation indicators or institutional WASH indicators) do not follow the same pattern as the ones analyzed do.

There is not a plethora of existing studies with a similar holistic system-wide analysis of M&E practices or data aggregation processes. For this reason, no widely established research methodology could be identified directly from prior literature, and one had to be constructed for analyzing the thematic of M&E across different scales by modifying existing approaches to stakeholder analysis. It is likely that this methodology could be developed and modified for further research on the topic, or for other similar research applications in the future.

However, the lack of research is also one indication that more research with holistic and scale-sensitive perspectives on M&E processes should be pursued, and there is likely considerable research interest in analyzing practices around development cooperation M&E systems and practices across different strata in a real-world case study context. In Ethiopia's case, there are many developments underway, and the situation is likely to evolve over the upcoming years when it comes to nationwide monitoring and evaluation of rural WASH sector results. The COWASH project's Phase IV also begun in spring 2021, and it is very interesting to see how the new phase of the project (with a new results framework and project plan) will change the situation with regards to M&E. As such, there is persisting interest to continue research with regards to the case of Ethiopia and the COWASH project, and it is safe to assume that the evolution of the case study context will continue.

Finally, it is appropriate to discuss the generalizability and scalability of the results obtained from this study. Although the study was quite heavily constrained to water sector development cooperation, the findings are applicable to development cooperation interventions in general, beyond WASH development cooperation work alone. The study provides a steppingstone of sorts for further analyzing the practical implementation of M&E systems, for which there is likely great demand by actors across scales, and the systems devised for M&E in the WASH sector likely have potential applicability for other sectors as well. As such, cross-sectoral cooperation will be important in sharing the most functional M&E practices, along with new innovations.

Further developing the analytical framework and indicator correspondence analysis that have been used in this study are also rather encouraging avenues. Finding practical applications for organizational development and planning in practical development cooperation scenarios is a possible approach to their utilization. The rather novel perspective of analyzing the M&E systems and indicators on adjacent scales could prove useful for various organizations engaged in development cooperation work, and it could prove fruitful to develop these into an easily approachable tool or method for understanding the effectiveness of M&E systems.

6 Conclusion

The purpose of this study has been to analyse the monitoring and evaluation systems in the case study context of the COWASH Phase III project in Ethiopia. Since the effectiveness of monitoring and evaluation of results form a sizeable contribution towards building timely, targeted, and successful project interventions in the rural water sector, ensuring that results are well captured by systems is an important consideration. While there is clear recognition of the importance of access to sustainable and secure drinking water resources in building resilience and tackling poverty, there must be clear and reliable evidence that the interventions are best targeted towards tackling these important considerations This is increasingly the case as the SDGs call for an increased level of ambition in monitoring and evaluation, and new frontiers in more real-time monitoring are beginning to open upon various technical advancements. In plain speech, good intentions must be accompanied by good tracking of progress, and ambition must be met with ambition.

However, a review of the literature reveals that while the increase of ambition expected from M&E systems has been researched theoretically, there are many challenges to building a unified picture of M&E in the case study context across the different levels of analysis, and there are relatively dispersed trends and priorities affecting the system. There seems to be relatively little previous academic inquiry into cross-strata integration of M&E methodologies in practice.

This study has employed a novel type of vertical analysis of three different strata (micro, meso and macro) in the case study context to examine how increased M&E ambition stemming from the SDGs can be harmonized with the difficulty of implementing and operationalizing M&E systems in the development cooperation project intervention context. The methodology used has been semi-structured interviews with seven key informants, along with a cross-strata analysis of indicators for rural drinking water access coverage indicators. An analytical framework and indicator correspondence analysis (informed by the SNA and knowledge mapping methods to stakeholder analysis) have been devised to situate the findings.

Firstly, the study has looked at the incorporation of SDG 6 across the different levels of M&E in this project. Second, the degree to which M&E mechanisms and practices encourage effective data collection, responsiveness, and improvement of project interventions. Thirdly, the integration of indicators and M&E frameworks between strata has been analysed, along with relevant data aggregation challenges.

On the first research question, the results show that there are varying degrees of SDG 6 integration in the M&E mechanisms in use, but the micro and meso stratums generally use national indicators based on GTP II that are not directly SDG-compatible. The donor (MoFA) and the COWASH project, however, have made efforts to increase the SDG 6-compatibility of indicators and M&E systems. There needs to be further discussion of increasing the use of SDG 6 in Ethiopia, but also discussion on how the SDGs can be implemented and understood at levels close to the beneficiaries if they are indeed to be utilized more universally. In other words, since there is little knowledge (and, subsequently, implementation) of SDG 6 at the local level in the case study context, there should be active discussion on the degree to which the SDGs are relevant for the most local actors, and how their voice could be better included in the SDG discourse.

In terms of the second research question, there is evidence that actors on higher strata generally employ more elaborate data collection practices and have greater capacity in the field of M&E, such as accounting for errors in data from lower strata. Various procedures have been taken at all levels to increase data quality and effectiveness of monitoring, but there are still challenges particularly on the most local levels. As such, there is still work to do in effective empowerment of local actors in monitoring and evaluation work, and the conclusion that local-level actors need to be empowered in data collection and monitoring work is supported by the literature.

Further on the second research question, there is hope in broadening the sources of data used for monitoring and evaluation in the future in Ethiopia and the COWASH project's subsequent iterations. There exist examples of novel uses for citizen-reported big data, or in-situ sensor or remote sensing data, but there needs to be considerable attention paid to how these data sources can be effectively integrated in data systems. Time will tell whether these novel methodologies will one day provide effective supplementary data for monitoring and evaluation of rural water sector interventions.

On the third research question, results of the study indicate that systems devised on each stratum function well in isolation, but rather little attention has been paid to vertical integration of systems across the different strata. As such, there is a continued need for methodological harmonization and coordination in M&E between the different strata and institutions thereon. Although full harmonization is likely not possible or even desirable to achieve, the national stakeholder consultation process in Ethiopia is encouraging for increasing coordination, and the meso stratum (national actors) are key in increasing coherence between stakeholders.

Despite some limitations, the results of this study are promising for the field of development cooperation, and this study provides one contribution and perspective to further exploring the role of M&E in ensuring sustainable and safe drinking water and sanitation for all, as is called for by SDG 6. Practical analysis of project interventions, alongside theoretical discussions, is key in understanding how M&E systems and practices can better serve the aims of development cooperation. There ought to be greater academic inquiry into how best to measure progress in WASH sector development cooperation interventions.

References

Ayenew, Mehret. 2002. “Decentralization in Ethiopia: Two Case Studies on Devolution of Power and Responsibilities to Local Authorities.” In *Ethiopia: The Challenge of Democracy from Below*, edited by Bahuru Zewde and Siegfried Pausewang, 215. Uppsala, Addis Ababa: Nordiska Afrikainstitutet and Forum for Social Studies.

Behailu, Beshah M., Arto Suominen, and Tapio S. Katko. 2015. “Evolution of Community-Managed Water Supply Projects From 1994 to the 2010s in Ethiopia.” *Public Works Management and Policy* 20 (4): 379–400. https://doi.org/10.1177/1087724X15593955.

Butterworth, John, Katharina Welle, Kristof Bostoen, and Florian Schaefer. 2013. “WASH Sector Monitoring.” In *Achieving Water Security - Lessons from Research in Water Supply, Sanitation and Hygiene in Ethiopia*, edited by Roger Carlow, Eva Ludi, and Josephine Tucker, 49–68. Bourton on Dunsmore, Rug: Practical Action Publishing Ltd.

Calow, Roger, Eva Ludi, and Josephine Tucker. 2013. *Achieving Water Security - Lessons from Research in Water Supply, Sanitation and Hygiene in Ethiopia*. Edited by Roger Calow, Eva Ludi, and Josephine Tucker. 1st ed. Bourton on Dunsmore, Rug: Practical Action Publishing Ltd.

Creswell, John F. 2013. *Qualitative Inquiry & Research Design - Choosing Among Five Approaches*. Thousand Oaks, CA, USA: Sage Publications, Inc.

Damtew, Yohannes Tefera, and Abraham Geremew. 2020. “Households with Unimproved Water Sources in Ethiopia: Spatial Variation and Point-of-Use Treatment Based on 2016 Demographic and Health Survey.” *Environmental Health and Preventive Medicine* 25 (1): 1–11. https://doi.org/10.1186/s12199-020-00921-1.

Feeny, Simon. 2020. “Transitioning from the MDGs to the SDGs: Lessons Learnt?” In *Moving from the Millennium to the Sustainable Development Goals: Lessons and Recommendations*, edited by Sefa Awaworyi Churchill, 343–51. Palgrave Macmillan. https://doi.org/10.1007/978-981-15-1556-9.

Gedamu, Yohannes. 2021. *The Politics of Contemporary Ethiopia*. 1st ed. New York, NY: Routledge.

Giné Garriga, Ricard, Alejandro Jiménez Fdez. de Palencia, and Agustí Pérez Foguet. 2015. “Improved Monitoring Framework for Local Planning in the Water, Sanitation and Hygiene Sector: From Data to Decision-Making.” *Science of The Total Environment* 526 (September). https://doi.org/10.1016/j.scitotenv.2015.04.078.

Government of Ethiopia. 2015. “Federal Democratic Republic of Ethiopia Growth and Transformation Plan II (GTP II).”

Hammond, A. L., A. Adriaanse, Eric Rodenburg, D. Bryant, and R. Woodward. 1995. *Environmental Indicators : A Systematic Approach to Measuring and Reporting on Environmental Policy Performance in the Context of Sustainable Development*. *Undefined*. 1st ed. Washington, DC: World Resources Institute.

Hartman, Ronald, Leon Williams, and Deirdre Mc Grenra. 2020. “Report of the Consultation on the Twelfth of IFAD’s Resources - Recovery, Rebuilding, Resilience.” https://webapps.ifad.org/members/repl/12/4/docs/IFAD12-4-R-2-Rev-1.pdf.

Impact Consulting Oy Ltd. 2019. “Finland’s Support to the WASH Sector in Ethiopia (COWASH IV and CWA II) Project Document Ministry for Foreign Affairs of Finland Unit for the Horn of Africa and Eastern Africa.”

IRC. 2016. “Organizing Framework for Functional National WASH Monitoring and Evaluation Systems.” *Evaluation*. https://www.ircwash.org/resources/organizing-framework-functional-national-wash-monitoring-and-evaluation-systems.

Janoušková, Svatava, Tomáš Hák, and Bedřich Moldan. 2018. “Global SDGs Assessments: Helping or Confusing Indicators?” *Sustainability (Switzerland)* 10 (5): 1–14. https://doi.org/10.3390/su10051540.

Kusek, Jody Zall, and Ray C Rist. 2004. *A Handbook for Development Practicioners - Ten Steps to a Results-Based Monitoring and Evaluation System*. Washington, DC: The International Bank for Reconstruction and Development / The World Bank.

Lockwood, Harold. 2015. “Transforming Accountability and Project Monitoring for Stronger National WASH Sectors.” In *From Infrastructure to Services*, 63–84. Practical Action Publishing Ltd. https://doi.org/10.3362/9781780448138.004.

Marvin, Esra. 2021. “Policy Analysis of Ethiopia’s Rural Water Operation and Maintenance Policies.” Espoo, Finland.

Matthews, Sally J. 2010. “Postdevelopment Theory.” In *Oxford Research Encyclopedia of International Studies*. Oxford University Press. https://doi.org/10.1093/acrefore/9780190846626.013.39.

Ministry for Foreign Affairs of Finland. 2016. “Project Document - Support to Community-Led Accelerated WASH (COWASH) Phase III.” https://www.cmpethiopia.org/content/download/2574/10792/file/COWASH%20Phase%20III%20Project%20Document.pdf.

———. 2018. “Finland’s Development Policy Results Report 2018 Summary.” Helsinki, Finland. https://um.fi/publications/-/asset_publisher/TVOLgBmLyZvu/content/kehityspolitiikan-tulosraportti-2018.

Person, Margaret T., Maryann G. Delea, Joshua v. Garn, Kelly Alexander, Bekele Abaire, and Matthew C. Freeman. 2017. “Assessing the Influence of Social Capital on Water Point Sustainability in Rural Ethiopia.” *Journal of Water Sanitation and Hygiene for Development* 7 (4): 611–22. https://doi.org/10.2166/washdev.2017.167.

Puusa, Anu. 2020. “Haastattelutyypit Ja Niiden Metodologiset Ominaisuudet.” In *Laadullisen Tutkimuksen Näkökulmat Ja Menetelmät*, edited by Anu Puusa and Pauli Juuti, 1st ed., 103–17. Helsinki: Gaudeamus Oy.

Reed, Mark S., Anil Graves, Norman Dandy, Helena Posthumus, Klaus Hubacek, Joe Morris, Christina Prell, Claire H. Quinn, and Lindsay C. Stringer. 2009. “Who’s in and Why? A Typology of Stakeholder Analysis Methods for Natural Resource Management.” *Journal of Environmental Management* 90 (5): 1933–49. https://doi.org/10.1016/j.jenvman.2009.01.001.

Rodarte, Mary, ed. 2003. *African Wisdom*. Kansas City, Missouri: Andrews McMeel Universal.

Sachs, Jeffrey D. 2012. “From Millennium Development Goals to Sustainable Development Goals.” *The Lancet*. Elsevier B.V. https://doi.org/10.1016/S0140-6736(12)60685-0.

Sadoff, Claudia W., Edoardo Borgomeo, and Stefan Uhlenbrook. 2020. “Rethinking Water for SDG 6.” *Nature Sustainability* 3 (5): 346–47. https://doi.org/10.1038/s41893-020-0530-9.

Schwemlein, Stefanie, Ryan Cronk, and Jamie Bartram. 2016. “Indicators for Monitoring Water, Sanitation, and Hygiene: A Systematic Review of Indicator Selection Methods.” *International Journal of Environmental Research and Public Health* 13 (3). https://doi.org/10.3390/ijerph13030333.

Sen, Amartya. 2002. *Devleopment as Freedom*. 1st ed. New York: Alfred A. Knopf, Inc.

Senbeta, Fekadu Megersa, and Yang Shu. 2019. “Project Implementation Management Modalities and Their Implications on Sustainability Ofwater Services in Rural Areas in Ethiopia: Are Community-Managed Projects More Effective?” *Sustainability (Switzerland)* 11 (6). https://doi.org/10.3390/su11061675.

Silva Wells, Carmen da, René van Lieshout, and Erma Uytewaal. 2013. “Monitoring for Learning and Developing Capacities in the WASH Sector.” *Water Policy* 15 (SUPPL.2): 206–25. https://doi.org/10.2166/wp.2013.120.

Suominen, Arto, and Oona Rautiainen. 2016. “Effective and Sustainable WASH Services : A Case Study of Community Managed Project (CMP) Approach in Ethiopia.” In *38th WEDC International Conference, Loughborough University, UK, 2015*, 1–6. https://repository.lboro.ac.uk/articles/conference_contribution/Effective_and_sustainable_WASH_services_a_case_study_of_Community_Managed_Project_CMP_approach_in_Ethiopia/9586658.

Thomas, Evan, Luis Alberto Andres, Christian Borja-Vega, and German Sturzenegger. 2018. *Innovations in WASH Impact Measures*. https://openknowledge.worldbank.org/bitstream/handle/10986/29099/9781464811975.pdf?sequence=4&isAllowed=y%0Ahttp://elibrary.worldbank.org/doi/book/10.1596/978-1-4648-1197-5.

Thomson, Patrick, and Johanna Koehler. 2016. “Performance-Oriented Monitoring for the Water SDG – Challenges, Tensions and Opportunities.” *Aquatic Procedia* 6 (0): 87–95. https://doi.org/10.1016/j.aqpro.2016.06.010.

Tsan, Michael, Swetha Totapally, Michael Hailu, and Benjamin K Addom. 2019. “The Digitalisation of Agrican Agriculture Report.” https://www.cta.int/en/digitalisation-agriculture-africa.

UN Water. 2017a. “Integrated Monitoring Guide for Sustainable Development Goal 6 on Water and Sanitation - G1 Good Practices for Country Monitoring Systems.” https://www.unwater.org/publications/good-practices-sdg-6-monitoring/.

———. 2017b. “Integrated Monitoring Guide Sustainable Development Goal 6 on Water and Sanitation - G2 Targets and Global Indicators.” https://www.unwater.org/publications/sdg-6-targets-indicators/.

United Nations. 2015. “The Millennium Development Goals Report 2015.” New York. https://www.un.org/millenniumgoals/2015_MDG_Report/pdf/MDG%202015%20rev%20(July%201).pdf.

United Nations DESA. 2019. “United Nations Department of Economic and Social Affairs World Population Prospects 2019 - Ethiopia.” World Population Prospects 2019. 2019.

USAID. 2003. “Knowledge Mapping 101.” In *Knowledge for Development Seminar*, 1–39. https://pdf.usaid.gov/pdf_docs/PNADK308.pdf.

Welle, Katharina. 2013. “Ethiopia National WASH Inventory: Lessons Learned and Maximising Value Summary of a Seminar Held on 8 April 2013, Addis Ababa, Ethiopia.” Addis Ababa, Ethiopia: IRC WASH. https://www.ircwash.org/sites/default/files/nwi_seminar_8_april_2013_final.pdf.

WHO/UNICEF JMP. 2018. “JMP Methodology, 2017 Update & SDG Baselines.” https://washdata.org/report/jmp-methodology-2017-update.

———. 2021. “WHO/UNICEF JMP Database - Households (Ethiopia, SSA, World).” JMP Database. 2021. https://washdata.org/data/household.

Wilson, Richard, Kebede Faris, Asrat Getaneh, and Ayenew Admasu. 2018. “Federal Democratic Republic of Ethiopia One WASH National Programme - Phase II Programme Document.” https://www.cmpethiopia.org/media/ownp_phase_ii_final_version_2019.

Yilmaz, Serdar, and Varsha Venugopal. 2008. “Local Government Discretion and Accountability in Ethiopia.” 08–38. International Studies Program Working Paper. Atlanta, Georgia, USA. http://isp-aysps.gsu.edu.

Yin, Robert K. 2009. *Case Study Research - Design and Methods*. 4th ed. Vol. 5. Thousand Oaks, CA, USA: Sage Publications, Inc.

Ziai, Aram. 2016. *Development Discourse and Global History - From Colonialism to the Sustainable Development Goals*. 1st ed. New York, NY 10017: Routledge.

Annex I: Interview guide and participant notice for the semi-structured interviews

Questions used in the interview guide

1. **Role and position of interviewee**

a. Please describe your position in the COWASH project/OWNP

or

b. Please describe your work in relation to the monitoring related to SDG 6 (JMP)

2. **Data collection and analysis in COWASH on interviewee's level of work**

For interviewees involved with OWNP or JMP reporting only, ask generally about lower levels or Ethiopia generally (in lieu of COWASH specifically) and on higher-level practices on data collection/analysis specifically.

a. Please describe, how is data typically collected in COWASH project work? Who collects the data at the level of activity implementation (WASHCO level)?
b. What type of data processing/analysis do you carry out in your position?
c. How is collected data stored and/or aggregated at your level of work?
d. What types of data processing is carried out, and with what tools?
e. In your opinion, how does the data change when it is processed in the project work and reporting?
f. What types of challenges have you observed in the M&E system and data collection/processing workflow from the point of view of your work?
g. (*if relevant*) How does the GIS data system for the project (COWASH Database) function? Have you observed challenges in using this system?

3. **Effectiveness and accuracy of the data collection and reporting system**

a. How well do you feel the data/results produced in the project administration locally (kebele/woreda) and regionally represent actual progress on the ground?
b. Do you think there are possible sources of error in data collected? How can these be mitigated in the project work?
c. What do you think could/should be done to improve results and data accuracy?
d. How responsive do you think monitoring and evaluation is in the COWASH/OWNP, ie. are the results taken into consideration at your level of work/on other levels to improve the effectiveness of project work?

e. In your opinion, how well does the communication on M&E across different levels function in this project? What could be improved?

4. **Data aggregation and indicator correspondence across levels**

Ask about these with relation to positionality:

- *If involved directly with COWASH, on how well the data and results transcend into national and global levels of reporting*
- *If involved with OWNP/JMP, on how the national and global levels of data and result management are presently harmonized*

a. What is your knowledge of how the project data is transformed when aggregated across different levels (local, national, global)?
b. Do you feel there is effective communication between the local project level actors and the national level authorities? What could be improved in cohesion between these different levels?

(if applicable)

c. Do you feel that there are presently effective connections between national level M&E systems/frameworks and those used for global level of reporting? How could/should these be improved in your opinion?

5. Effects of the SDGs on project data collection

a. What are the most important changes to M&E and data management that have occurred in COWASH/OWNP?
b. Are you aware of the new requirements of the Sustainable Development Goals on the project work? How would you summarize the most important changes to monitoring and evaluation since SDG adoption?
c. How have the new SDG indicators and targets affected the project's outlook? How do the SDGs affect your own work with data collection/analysis?
d. What types of challenges do you feel the new types of indicators have placed on effectively carrying out project data collection?

6. Other

a. Is there something I did not ask about that you feel would be important to discuss?
b. Is there something you think would be important to address in the thesis?

Notice relayed to interview participants regarding the interview methodology and protection of personal data

Invitation to Research Interview - Master's thesis project,
Aalto University School of Engineering, Finland
Monitoring and Evaluation in the Rural Water Sector in Ethiopia - case of a water project in sector-wide progress towards achieving SDG 6

This is an invitation to participate in a research interview on monitoring and evaluation (M&E) methods in the rural water sector development cooperation. The study focuses on the Ethiopian rural water sector in particular, using the Community-Led Accelerated WASH (COWASH) Project Phase III and the related Ethiopian One WASH National Program (OWNP) as case studies to illustrate the underlying phenomena. The study seeks to form a better understanding of the whole chain of data collection, processing, and reporting in the context of this case study, and to evaluate the performance of the project through existing criteria. Furthermore, the aim is to understand the whole data collection and aggregation workflow for SDG 6.1 from the micro scales to the macro scale, and how each level in the case study interacts with other M&E frameworks and indicators in use at other scales.

Interviewer and details of the interview:

The interview will be carried out by Eelis Hemberg, MSc (Tech) candidate at Water and Development Research Group, Aalto University School of Engineering. The thesis work is supervised by Associate Professor Marko Keskinen DSc (Tech), with MSc(Tech) Anni Juvakoski and MSc(Tech) Arto Suominen acting as academic advisors.

The interview will be carried out by Zoom or equivalent teleconferencing medium, unless otherwise specified. The interviews in this project will follow a semi-structured interview format, with questions tailored specifically to the interviewee's position.

Privacy Statement:

The purpose of this interview is to obtain data for research with the primary purpose of completing a master's thesis. In addition, anonymous interview data may be used for other research purposes.

Data collected in this interview will follow the Aalto University General instructions for secure processing of personal data, national legislation in Finland, and the EU General Data Protection Regulation, as outlined here: https://www.aalto.fi/en/services/general-instructions-for-secure-processing-of-personal-data

Personal data collected in the interview (raw notes containing personal information and possible audio recordings) will be retained by the interviewer until December 2021, after which they will be destroyed. Personal data will not be used for any other purposes than research carried out by the interviewer, and for necessary verification by the thesis supervisor and advisors. Personal data will be retained in a secure platform, and it will not be transferred to any unrelated purpose. Personal data collected will not be transferred outside the EEA region. Additional consent will be obtained from the interviewee for any additional processing outside the intended scope or purpose.