ICAI Independent
Commission
for Aid Impact

Assessing UK Aid's Results in Water, Sanitation and Hygiene (WASH)

Approach Paper

October 2015

Contents

1. Purpose and rationale

This review assesses the contribution of UK Aid to achieving sustained access to clean drinking water, improved sanitation and hygiene (WASH). The review is expected to generate useful insights into the impact of DFID's WASH portfolio and into its systems for measuring and reporting results. This will support both accountability and learning.

2015 marks the end date for the Millennium Development Goals (MDGs) and the adoption of the global goals for sustainable development. There is widespread public interest in global progress towards the MDGs and in the UK's contribution. Success in WASH is not only considered important in its own right but is pivotal in achieving other development goals. These include child mortality, maternal health, nutrition, gender equality and poverty reduction.

DFID set itself an ambitious target of reaching 60 million people with WASH interventions in the period 2011 to 2015. In its Annual Report 2014-15, DFID reported that it had exceeded this target by 2.9 million.[1] With the wider reckoning on MDG achievement now taking place, this is a good moment to assess the validity of DFID's results claim. It is also a good time to influence how DFID responds to the proposed WASH target in the forthcoming Sustainable Development Goals.

2. Type of review

This is an **impact review**. ICAI impact reviews examine results achieved by a particular aid programme or portfolio and whether this represents a good return on UK aid investment. In this review we will examine the results in WASH contributed by UK aid and the quality of the evidence underlying those results. This includes a close look at the aggregate results set out in *DFID's Results Framework 2011-2015*[2] and at specific claims to impact made as part of selected WASH programmes.

3. Scope

The review will cover DFID programmes and expenditure feeding into the aggregate results for WASH set out in *DFID's Results Framework*. According to DFID this includes 71 bilateral projects (including those implemented by multilateral agencies) in 27 countries providing WASH access for 54.1 million people. UK core funding of multilateral organisations accounts for the remaining 8.86 million people provided with WASH access during the reporting period (with World Bank programmes contributing the largest share at 7.9 million).[3]

4. Background

DFID's 2012 commitment to providing 60 million people with sustainable WASH in the 2011-2015 period more than doubled a previous commitment (made in 2011).[4]

---

[1]   *Department for International Development Annual Report and Account 2014-15*, DFID, July 2015.
[2]   *DFID's Results Framework: Managing and reporting DFID results*, DFID, 2011.
[3]   Multilateral results are determined by imputing a proportion of the WASH results reported by each multilateral agency to DFID, based on the UK's share of total funding for that agency.
[4]   *Water and Sanitation: UK to double its support*, DFID, April 2012

At least 79 different project interventions have reported results that make up the aggregate figure.[5] Most are bilateral projects working in a single country, managed from the DFID country office. There are also several 'vertical' initiatives run by central DFID departments.[6] These vertical programmes are of considerable interest to the review, as the International Development Committee (IDC) has criticised poor coordination between central and country programmes[7] in the past.

DFID uses a range of delivery channels for its WASH programmes, including multilateral agencies, contractors, NGOs and partner governments. DFID informs us that 60% of its programmes are delivered by UNICEF. In some countries, including Ethiopia, Mozambique and Tanzania, the national government is the primary implementer, using budget support or other financial aid. In some humanitarian contexts, such as South Sudan, WASH results are delivered through multi-donor trust funds. These may be implemented by the United Nations or the World Bank.

The majority of projects incorporate water, sanitation and hygiene interventions. A small number are focused only on water supply (e.g. in Zimbabwe and Syria).

The UN reports that the MDG target for safe drinking water was met in 2010, well ahead of the target date of 2015. Over 90% of the world's population now has access to improved sources of drinking water. The 2015 MDG target for sanitation has not been met, however. Although 2.1 billion people have gained access to improved sanitation since 1990, the rate of progress has barely kept up with population growth. In absolute terms, the number of people lacking access to improved sanitation facilities today (2.4 billion) has not changed substantially over the MDG period. This raises critical questions about what needs to be done to accelerate and scale-up sanitation provision. [8]

5. Theory of change

There is no single strategy or policy document summarising DFID's theory of change (or intervention logic) in WASH. Individual DFID projects should have their own theory of change as part of the Business Case process. These should set out, with key assumptions explained, how DFID activities (inputs) are linked to their intended direct results (outputs), and to wider benefits to people's lives (outcomes and impacts).

To understand the theory of change we need to understand what direct and indirect benefits are expected to follow from spending on WASH. DFID documents point to a number of first-order benefits including time saved fetching water (with possible

---

[5]   One project, 'Hygiene and Water for All in Off-Track Countries', is split into 9 sub-projects, each working in a separate country.

[6]   Three are managed by Human Development Department, and one each by Research and Evidence Division, Africa Regional Department and the Civil Society Department.

[7]   *Recovery and Development in Sierra Leone and Liberia*, International Development Committee, Sixth Report of Session 2014-15, September 2014, pp. 24-27.

[8]   *Progress on Sanitation and Drinking Water Update 2015*, WHO and UNICEF, 2015.

economic implications), and a reduced burden of disease. These in turn lead to reduced mortality, morbidity and disability,[9] particularly for women[10] and children.

As well as direct benefits, support for WASH is associated with a number of second-order benefits such as:

- increased food production and reductions in hunger and malnutrition;
- improvements in girls' education through the provision of sanitation facilities at schools;
- improvements in security for girls and women through reduced vulnerability during open defecation and water collection.

While the evidence behind the primary health benefits is strong, that behind some of the secondary benefits is limited.[11]

To achieve these benefits, DFID WASH interventions fund the delivery of specific outputs and outcomes, including:

- constructing and maintaining water points;
- improving hygiene practices;
- reducing open defecation through the provision of sanitation facilities and behaviour changes;
- increasing the consumption of safe water;
- reducing the time spent collecting water, particularly for women and girls.

Sustainability is a key issue for WASH investments. Ensuring long-term benefits raises institutional, financial and technical issues. The literature suggests that sustainability can be enhanced through:

- improving the enabling environment, for instance implementing polices to encourage and promote sustainable WASH improvement;
- improving leadership, technical competence and the number of trained staff;
- strengthening community structures to ensure systems are operated and maintained;
- encouraging new financing mechanisms for the installation and maintenance of water infrastructure;
- communication and social mobilization around new behaviours (norm change);
- changing sanitation and hygiene practices in public facilities such as schools and clinics.

How DFID programming seeks to support the sustainability of WASH impacts is a key focus of this review.

6. Existing evidence

In March 2012, DFID undertook a review of its WASH portfolio.[12] The review assessed the evidence base for wider development impact and recommended

---

[9] *Water Sanitation and Health Evidence Paper,* DFID, May 2013.

[10] Aamirah Patel, 'Safe water and basic sanitation would slash maternal deaths, report says', *The Guardian*, 16 December 2014, link.

[11] *Water, Sanitation and Hygiene Portfolio Review*, DFID, March 2012, page 3.

tailored programming and financing models. Their aim was to address different drivers of demand for water supply. It also recommended a number of actions to maximise value for money, such as more systematic monitoring and reporting and improved integration between health and WASH programmes. The portfolio review provides useful baseline data on the state of the portfolio at the beginning of the results period. An ICAI review in 2016 will also provide a timely opportunity to assess how well DFID has responded to these recommendations.

Most DFID WASH programmes have embedded monitoring and evaluation frameworks. As well as measuring other results, they contain a standard indicator from *DFID's Results Framework*[13] on the number of unique people who have received access to WASH through DFID interventions, using a standard methodology.[14] This methodology makes certain assumptions (such as on the numbers of beneficiaries accessing water points) that will need to be examined.

DFID informs us that, of the 71 projects in its WASH portfolio, only six have been subject to independent evaluation. These include three in Vietnam and one each in the Democratic Republic of Congo (DRC), Zimbabwe and Bangladesh. Evaluations of a further ten projects are planned. We have found thus far that only the DRC example has a published exploration of impact.[15] A further eight evaluations are planned for projects identified as contributing to DFID's WASH results. DFID is also funding an assessment of the quality of results in WaterAid activities in 2016 (WaterAid receives core DFID funding through a Programme Partnership Arrangement). DFID has also commissioned operational research and monitoring, verification and evaluation services for three centrally funded projects: the Water and Sanitation for the Urban Poor Programme, the WASH Results Programme and the Accelerating Sanitation & Water for All in Off-track Countries. While work is in progress, these evidence gathering exercises provide an important point of reference for this review.

The wider evaluation literature provides a guide to the kinds of impact that might be expected from DFID's WASH portfolio.[16] The main causal mechanisms are health gains, including reduction in water and sanitation related diseases, particularly diarrhoea, which represents 90% of the avoidable disease burden. There have been a number of systematic reviews on WASH impact that may be of relevance. 3iE have summarised many of these in a 'Gap Map' that categorises WASH interventions as well as identifying areas where evaluative knowledge is weak.[17] It identifies a number of systematic reviews assessing the effects of WASH interventions on diarrhoea. However, there are only a few systematic reviews of other results such as time use, safety and musculoskeletal injury (particularly for women and girls), and

---

[12] *Water, Sanitation and Hygiene Portfolio Review*, DFID, March 2012.

[13] *DFID's Results Framework: Managing and reporting DFID results*, DFID, 2011.

[14] *Number of unique people reached with one or more water, sanitation or hygiene promotion intervention*, DFID, undated.

[15] *Evaluation of the School Sanitation and Village Sanitation Programme (EVA Programme)*, Eau et Assainissement pour l'Afrique, undated: Executive Summary.

[16] In particular, *Water, Sanitation and Hygiene Evidence Paper*, DFID, May 2013; Michael Loevinsohn et al., *The cost of a knowledge silo: a systematic re-review of water-sanitation and hygiene interventions*, Health Policy and Planning, No. 30(5), 2015, pp. 660-674; Sandy Cairncross et al., *Water, sanitation and hygiene for the prevention of diarrhoea*, International Journal of Epidemiology, No. 39, 2010, pp. 193-205.

[17] See Link. This refers to 23 systematic and 139 impact evaluation studies of WASH interventions.

economic outcomes. It also highlights several gaps in the primary evidence base, in particular a lack of prospective impact evidence from sub-Saharan Africa and studies assessing sanitation and hygiene programmes at scale. The extent to which DFID is aware of and working to fill some of these evidence gaps will be of particular interest to this review.

7. Review criteria and questions

The review uses three of the standard OECD-DAC evaluation criteria: effectiveness, impact, and sustainability. These are elaborated into the following questions and sub-questions.

1. **Impact:** what has been the impact of DFID's WASH programming in the 2011-2015 DFID's Results Framework period?

   1.1. What level of results has been achieved through DFID's WASH portfolio?
   1.2. Have DFID's WASH investments been targeted effectively on the poorest and most marginalised communities and individuals?
   1.3. What has been the variation in impact across programmes, countries and delivery channels? What does this tell us about value for money and return on investment?

2. **Effectiveness (results measurement):** are the aggregate WASH results reported under DFID's Results Framework based on credible evidence?
   2.1. How credible are the processes that DFID uses to collate, review and validate evidence on results?
   2.2. To what extent have WASH interventions been strengthened in response to lessons learned on delivering results?

3. **Sustainability:** has the impact from DFID's WASH programming proved sustainable?
   3.1. Have DFID's WASH interventions been designed and implemented with a view to maximising the sustainability of results?
   3.2. Are appropriate arrangements in place to monitor the sustainability of results after the end of the funding period?

8. ICAI Themes and Core Issues

This review primarily adheres to the 'leaving no-one behind' theme. It relates to a core MDG commitment that is also one of the seventeen Sustainable Development Goals.[18] It addresses the equitable provision of basic services, to overcome poverty and marginalisation. It has particular pertinence to women and girls, given their disproportionate share of the burden imposed by poor access to WASH.

In terms of core issues, the review focuses primarily on results management, including managing results through a variety of delivery channels and DFID's ability to capture and report on aggregate results. It also raises important questions of

---

[18]   United Nations General Assembly resolution on the SDGs report, 10 September 2015

value for money, and DFID's ability to learn lessons on how to maximise sustainable results and apply them across a complex portfolio.

Gender equality is an important sub-theme. The review will examine the extent to which WASH results data is gender disaggregated and whether development results specific to women and girls (such as reduction in time poverty) are being measured and maximised.

9. Methodology

Fig 1. Summary of methodology components
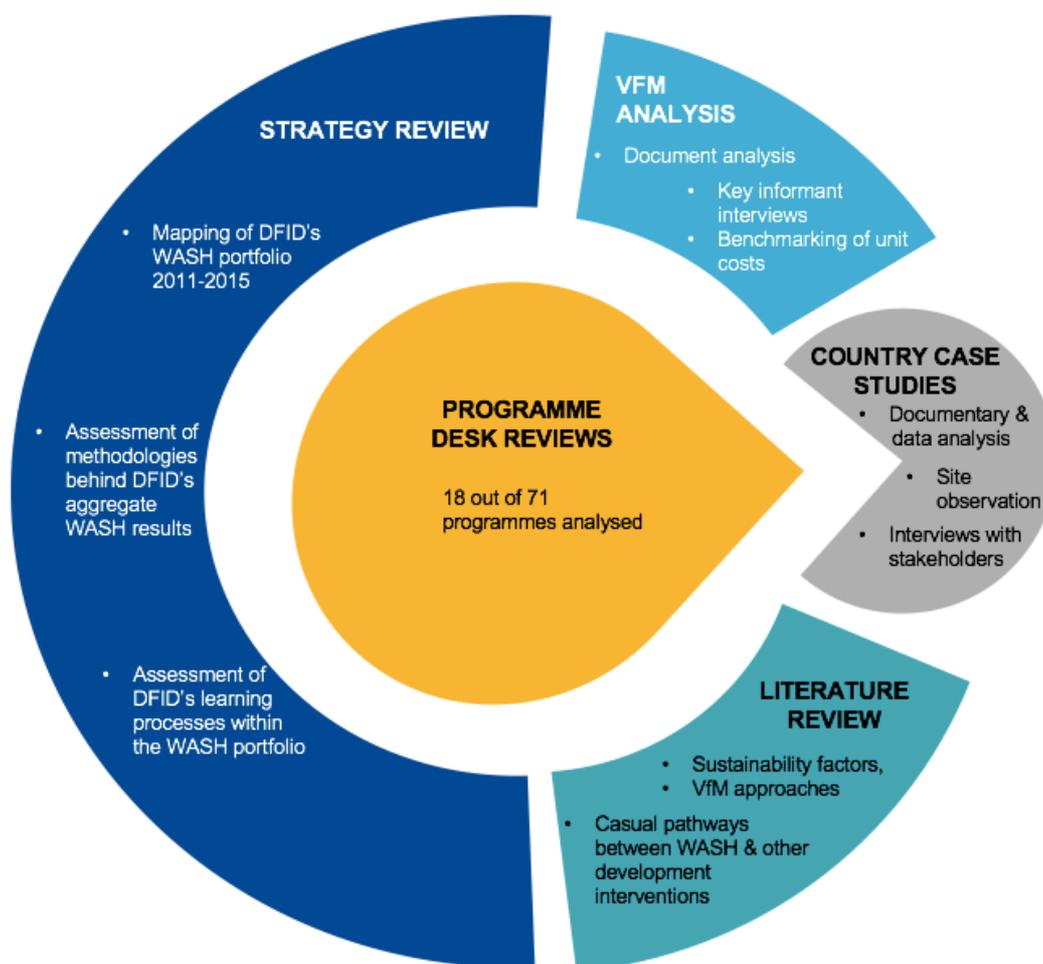
The methodology for this review will (i) assemble and analyse available information on impact from across DFID's WASH portfolio; and (ii) assess the robustness of DFID's results claims in WASH. No new impact data will be generated. Instead the focus is on examining data already available and subjecting it to various tests of validity. In particular, it will probe whether DFID is able to substantiate its claims to have expanded WASH access, leading to sustainable development impact. It will also use internal comparators to assess whether DFID is making sufficient effort to maximise the overall return on its investment and achieve value for money. Details of how evidence will be gathered in support of the review questions are described in

the review framework in Annex 1.

The methodology has five linked components.

i) Literature review

It will draw substantially on existing reviews and syntheses to identify:

- the causal pathways between WASH and other development interventions and the availability and strength of evidence in support of each;
- factors affecting the sustainability of WASH interventions;
- approaches and techniques for assessing value for money in WASH.

The literature itself will be identified through database searches and expert judgement, drawing on a limited consultation with key experts to identify gaps and provide triangulation.

ii) Strategy review

- A mapping of DFID's WASH portfolio 2011-2015 involving examining the main patterns of expenditure and activity. It also assesses how they have evolved in response to evidence and lessons learned.
- An assessment of the methodology behind DFID's aggregate WASH results. This will include a review of the DRF WASH indicators and associated guidance. In particular we will assess whether the guidelines for monitoring and evaluating WASH programmes enable DFID to capture accurately the range of outputs, outcomes and impacts it is attempting to achieve.
- An assessment of DFID's learning processes within the WASH portfolio, including DFID's effort to assemble evidence about what works, the level of investment in impact evaluation and how lessons are shared across the portfolio. We will assess the technical adequacy of DFID's internal guidance on WASH programming and whether it is kept updated to reflect current evidence and learning.

The methods we will use while undertaking the portfolio review may therefore include:

- document review
- stakeholder analysis (of DFID and relevant UK actors)
- structured focus groups/interviews (e.g. of major UK WASH policy influencers and implementers)
- key informant interviews.

iii) Programme desk reviews

In depth analysis will be undertaken on a 25% sample of the 71 DFID programmes that contribute to DFID's WASH results. The sample will include 4 global

programmes, 13 country-based programmes and one multilateral engagement (see section 12 below on sampling).

A Desk Review Analytical Framework will be developed, following completion of the literature review and elements of the strategy review. This will capture standardised information and allow comparable assessments to be made against common assessment criteria for each of the sample programmes.

The framework will capture the following data: (i) types of WASH intervention (typology to be developed following the literature review, reflecting current evidence on how to maximise sustainable impact); (ii) the types of results (beyond WASH access rates) that the programmes seek to deliver; (iii) any theory of change linking inputs to outputs, outcomes and impacts; (iv) delivery channels; (v) monitoring and evaluation arrangements, including the existence of baseline data, monitoring strategies, use and timing of evaluations, and total expenditure on monitoring and evaluation relative to project budget; (vi) approach to maximising value for money (from Business Case and Annual Reviews), together with any available information on unit costs; (vii) evidence of course corrections following Annual Reviews; and (viii) available evidence on results and their sustainability.

iv) Value for money assessment

The review will aim to identify how DFID seeks to maximise value for money. A bespoke value for money assessment methodology will be developed during the review. This will be undertaken once we have a better understanding of the financial and results data available at programme level and the feasibility of different possible comparators from within and beyond DFID's WASH portfolio.

During the period of the review, a DFID-funded research project on value for money in WASH is due to report.[19] We will make use of the conclusions, by interviewing participants in the study and examining their source data.

The methods we will use when undertaking the value for money analysis are likely to include:
- document analysis
- key informant interviews (using semi-structured interviews)
- benchmarking of units costs, to the extent possible and a review of value for money approaches in WASH across development partners.

v) **Case studies** provide a means of examining DFID's approach to result measurement and management 'on the ground'. The aim is to test the quality of evidence underlying the results claims made by individual programmes. We

---

[19]   DFID has commissioned a consortium led by Oxford Policy Management and comprising London School of Hygiene and Tropical Medicine (LSHTM), the University of Leeds, Trémolet Consulting and Oxfam to identify how value for money and sustainability can be improved in DFID-funded WASH programmes, and to assess the sustainability of rural WASH services in Africa and South Asia. This project will report towards the end of 2015.

will compare how WASH results are calculated and aggregated across different contexts.

We will also assess whether the case study programmes are achieving, or are set up to achieve, sustainable results. We will test the programmes against evidence and good practices on sustainability identified in the literature and highlighted in the desk reviews.

Case studies will gather evidence against a <u>Case Study Analytical Framework</u>, a modified version of the overall review framework, covering three areas: how DFID seeks to maximise impact; measure impact; and ensure sustainability. The Framework will address the same set of issues as for the programme desk reviews, but in greater depth with more opportunity to test the underlying reasons for key findings. The Case Study Framework covers:

i) The types of impact (beyond WASH access rates) that the programmes seek to deliver.
ii) Analysis of the impacts in relation to the different types of WASH interventions.
iii) Analysis of the impacts in relation to different delivery channels; the theory of change that links inputs to outputs, outcomes and impacts.
iv) Monitoring and evaluation arrangements, including the existence of baseline data, monitoring and evaluation strategies, the use and timing of evaluations.
v) Programme approaches to value for money together with any available information on unit costs.
vi) Evidence of learning and, if necessary, course correction following Annual Reviews.

The case studies will also gather data (where feasible) on:

vii) The combined effects of different types of WASH intervention.
viii) Different approaches to achieving sustainability of results.
ix) Whether evidence exists on the life-time of programme benefits.
x) What results the programmes hoped to achieve over and above WASH access rates.
xi) Whether programme designs are optimised to achieve those results.
xii) Whether there are credible processes in place for measuring those results.
xiii) The extent to which external counterparts (e.g. Government and key local stakeholders) are able to verify programme related results and sustainability.

We will visit two countries - Mozambique and Zimbabwe - for 10 days each to undertake detailed case studies. These will cover all DFID WASH interventions in each country.

The final case study approach will be completed three weeks before the country visits take place.

The methods used are likely to include:

a. **Documentary and data analysis,** including gathering information from DFID's country programme and project management documents, from counterparts (e.g. government and other local actors) and third parties (e.g. local academic sources).

b. **Site observation** of at least three per project in each country to:

- Identify how results data is collected in practice.
- Triangulate data sources where feasible, including comparing reported results with third party data, direct observation and stakeholder feedback.
- Assess whether the institutional, technical and financial arrangements are in place to achieve sustainability (e.g. local governance and maintenance arrangements, payment schemes and technical soundness).

Semi structured interviews with key stakeholders to examine how DFID works to achieve impact in WASH. Interviews will also study the fit between the broader evidence on sustainability, DFID's investment in evidence on what works and its practice on the ground.

Annex 1 sets out the review framework, showing how these methodological elements relate to the review questions.

Sampling strategy

There are three samples to be selected for this methodology:

i) Programme desk studies
ii) Country case studies, and within these,
iii) Selection of project sites.

The sampling strategy and resulting samples are set out in detail in Annex 2.

Programmes for the Desk Review

The population of DFID WASH programmes is heavily dominated by a small number of large programmes: over 50% of total reported results are generated from just three countries. To be significant our sample must include the largest programmes. This may introduce a bias however, if larger programmes tend to perform better than the average on results reporting and management. We have therefore added a random component to the sample to correct for this.

To ensure a sample of material significance, we set a threshold of 50% of the results claimed and 25% of the total programmes. To ensure the sample is representative, we used the following additional selection criteria:

- A combination of country-level and central programmes.
- A spread of delivery channels.
- A spread of water, sanitation and hygiene projects.
- Maturity of results (to provide a basis for judging sustainability).

- Rural/urban delivery environments.

Following the criteria above, programmes for the desk review were selected as follows:

- We selected the largest individual programmes by reported results, until we reached 50% of the total results, yielding six projects in three countries (Bangladesh, Zambia and Nigeria).
- We selected three of the five largest centrally managed projects.
- We added multilateral expenditure through the World Bank, which accounts for 89% of WASH results reported from the multilateral budget.
- We added a random selection of eight projects, limited to one per country, yielding 18 projects (25% of the total).

We then checked the resulting sample against the additional selection criteria, verifying that it was representative. The sample is set out in full in Annex 2.

Case study selection. The criteria used to select our two case study countries were wholly purposive and were based on:

- Programming in DFID priority countries (there are 28).
- Programme maturity.
- A sufficient density of programming across water, sanitation and hygiene over the two case studies.
- A diversity of delivery channels and programme types (including vertical and bilateral).
- Programming across rural and urban areas where possible.

We also took into account certain practical considerations. We excluded insecure environments where field research is heavily constrained (as humanitarian programming is not a particular focus of this review). We also took into account the need to minimise the burden on individual DFID country offices across the ICAI review programme.

We considered materiality of spend as a criterion, but decided against it on the basis that there was no reason to assume that the quality of results management or sustainability of results would vary with the size of the programme budget. We also considered availability of existing evaluations as a criterion, but found too few cases for it to be reliable.

Using these criteria, we generated a shortlist of suitable countries, and from that list selected Mozambique and Zimbabwe as demonstrating the highest density of programming of interest. The combination of these two countries provides a good opportunity to get insights from various angles: longevity of programming, cumulative results over time, a variety of projects in country including centrally and country-managed, and variety of delivery channels and partners.

Site visits within case study countries

The sample of project sites for the team to visit in each of the two case study countries will be chosen during the preparation phase for each visit, when more

detailed information is available on the programmes in question. Each case study country has a number of WASH programmes. We will select a minimum of two and ideally three sites per programme to visit. These will be geographically disbursed and demonstrate the range of project interventions (e.g. urban, rural) undertaken by the project. We will prioritise project sites where the intervention has been fully operational for at least two years. Location data for the selection will be obtained from DFID and the selection will be made by the review team.

10. Limitations to the methodology

There are limitations in the accuracy with which we can map DFID WASH expenditure. DFID's systems do not always separately identify expenditure on WASH within multi-sectoral programmes (including budget support), which comprises a significant proportion of the total. The methodology does not attempt to investigate multi-sectoral programmes to calculate the WASH expenditure. This would be a lengthy process, the accuracy of which would be too dependent on data provided by DFID's implementing partners that could not be verified.

This review is not a full evaluation. The methodology is not able to specify the level of sustainable results actually achieved across the whole of DFID's WASH portfolio. Instead it will assess the validity of DFID's claim to have provided sustainable WASH access to 62.9 million people based on a review of methodology and underlying data sources. Our ability to identify wider development results from DFID's WASH programming is dependent on the quality of impact data collected by DFID.

We note that it may prove impossible to make precise value for money comparisons across DFID programmes. WASH delivery costs necessarily vary by country, location and numerous local factors. Our ability to make comparisons between the value for money achieved by DFID and other development partners will depend on the existence of published data. If it proves impossible to make robust comparisons of unit costs and rates of return, we will focus instead on making a qualitative assessment of whether DFID has a credible strategy for maximising value for money across its WASH portfolio.

Our case study approach is limited to considering DFID's performance in two countries. Data generated during the case study field visits will not be representative of the programmes in question or the portfolio as a whole. There will be limitations to the extent to which we can generalise from those findings. However, the findings can be used to explore hypotheses formulated in other components of the review. For example, they can be used to triangulate data obtained from DFID and to generate single-point data that can be used to assess DFID on its programming choices and approach.

11. Ethical considerations

ICAI reviews are undertaken with integrity and transparency. When commissioning and conducting reviews we are guided by ethical guidelines and codes of conduct of professional bodies such as DAC OECD[20]. When engaging with stakeholders we respect the principles of confidentiality and informed consent. When consulting

---

[20] DAC OECD, Quality Standards for Development Evaluation 2010

participants or beneficiaries ICAI review teams ensure the research does not constitute harm. It is also sensitive of aspects such as cultural environment, gender roles, age and disability.

12. Quality assurance and peer review

This review will be carried out under the guidance of ICAI lead Commissioner, Richard Gledhill. ICAI Commissioner Tina Fahm is peer Commissioner. The review oversight manager in ICAI is Cvetina Yocheva.

The review will be peer reviewed externally at three points: methodology, emerging findings, and draft report. The thematic peer reviewer will be Roger Calow of the Overseas Development Institute, with Howard White providing peer review of the methodology. The Service Provider will provide copies of the methodology and draft report to the peer reviewers. The peer reviewers will then prepare a short review (2-3 pages) summarising the robustness and technical quality of the Approach Paper/draft review. The thematic peer reviewer will be invited to attend the emerging findings presentation and offer comments on the emerging findings and analysis. He will also be available to the team and to ICAI for periodic consultations over the life of the review.

13. Risk management

Risk	Mitigation and management actions
Data not available to answer the review questions	If so, this will become a key finding. DFID central document systems do not hold sufficient project level detail. Detailed results data will therefore be collected from each country office. We will engage early on with the DFID staff response for each programme in the sample to facilitate access to data. Close management of documents and data will be undertaken to ensure coverage.
Country visit not possible due to unexpected events or deteriorating security conditions	The country visits are integral to the methodology, enabling us to perform detailed case studies and generate a richer set of findings. In the event that a case study visit has to be cancelled at short notice, we would discuss with ICAI whether another country visit could be substituted (this would have implications for publication date) or to do a more limited case study based on telephone interviews. In the event that a major component of the methodology proves infeasible, the resulting variation will be indicated in the final report. An assessment would be made to determine how to make up for any resulting weakness or bias in the methodology.
Security risk to Commissioners, Secretariat staff or review teams	The level of risk depends upon the countries selected. Mozambique and Zimbabwe are both low-to-moderate risks. Prior to any review, the Programme Director of the Security Provider will carry out a risk assessment and provide advice and guidance to the team on where to travel and by what means. If Nigeria is chosen, a more formal duty of care process will be triggered in accordance with the ICAI Handbook. This will include arrangements to monitor for changes in risk level, additional communication protocols and procedures for evacuation. If Commissioners or Secretariat staff are involved in the visit, duty of care responsibilities for the visit would be shared between the Service Provider and DFID.

14. Timeline and deliverables

Phase	Process and deliverables	Timing
Inception	Approval of Approach Paper	Early September
Data collection	Data collection and field work Evidence Pack Emerging Findings	Late August – week of mid November Mid December Mid January
Reporting	Draft to DFID for Fact Check Publication	Mid April

Annexes

Annex 1: Review framework

Review criteria, question/sub-question	Evidence required	Applicable component and activities	Analytical approach
Impact			
1. What has been the impact of DFID's WASH programming in the 2011-2015 DFID's Results Framework period?			
1.1 What level of results has been achieved through DFID's WASH portfolio?	▪ The results reported at corporate, country and project level in the period 2011-15 ▪ The results imputed from DFID's multilateral funding in the period 2011-15 ▪ To what extent do DFID's results data track the impact of WASH on wider development outcomes (e.g. child health, gender equity, education, other) ▪ If and how the DFID results have been questioned by third party sources.	Literature Review ▪ External commentary on DFID results claims Strategy review ▪ Capture of overall results from DFID systems ▪ Identification of country level results Programme Desk Review ▪ Extraction and collation of individual project results ▪ Capture of impact/ assumptions of WASH interventions on wider development outcomes Case Studies ▪ Cross-check of reported results derived from project documentation captured as corporate results	Mapping of results ▪ Clean DFID data if required ▪ Synthesise reported WASH data into excel datasets at corporate, country/institutional and project level (within sample only) ▪ Disaggregate DFID data by key factors including year, by Water, Sanitation, Hygiene, by channel and by urban/rural etc. ▪ Capture gender related reporting if available ▪ Create higher level development outcome dataset against WASH data if feasible (within sample only) ▪ Capture of country level development data ▪ Capture of UN JMP data as required Visualisation will be used as the preferred method of data presentation where feasible

Review criteria, question/sub-question	Evidence required	Applicable component and activities	Analytical approach
1.2 Have DFID's WASH investments been targeted effectively on the poorest and most marginalised communities and individuals?	▪ To what extent are DFID funded interventions designed to ensure the poorest and most marginalised communities and individuals are prioritised ▪ Whether DFID reporting captures outcomes and impacts on the poorest and most marginalised ▪ How DFID adapts programming to ensure benefits are optimised for the poorest and most marginalised	Literature Review ▪ Wider evidence about how to target programming on the needs of the poorest and most marginalised Strategy review ▪ Review corporate documents and guidance, policy and how to notes etc. Programme Desk Review ▪ Identify whether and how programme designs prioritise the needs of the most marginalised ▪ Identify if project reporting is sufficiently disaggregated for targeting purposes Case Studies ▪ Detail on the practices for the case studies ▪ Evidence of adaptation	Capture project level data ▪ Collate information captured from programme results frameworks ▪ Analyse trends and patterns Case study summaries

Review criteria, question/sub-question	Evidence required	Applicable component and activities	Analytical approach
1.3 What has been the variation in impact across programmes, countries and delivery channels? What does this tell us about value for money and return on investment?	• Findings from academic and grey literature on how inputs, costs, results, value for money and returns on investment for WASH interventions vary across different delivery channels, types of programme and contexts • Variation in costs and results across channels, countries and programmes across the DFID portfolio • The extent to which DFID tracks differences in costs across the portfolio • Whether there are gaps in data and knowledge about costs, value for money and return on investment	Literature Review • Synthesis of lessons learnt on costings from available literature (including generated from DFID's operational research) Value for Money Assessment • Identification of patterns relating to delivery channel, type or location of programme, using comparators to be determined • Key informant interviews during both Strategy review and case study activities	Identify and classify evidence for cost and value for money causality from literature. Identify and quantify whether cost and value for money data is available from projects in the sample. Analyse data to identify patterns of unit cost against programme variables. Qualitative assessment of DFID's approach to maximising value for money. Semi-structured interviews of DFID value for money study participants.
2. Effectiveness of Results Measurement: are the aggregate WASH results reported under the DFID's Results Framework based on credible evidence?			
2.1 How credible are the processes that DFID uses to collate, review and validate evidence on results?	• Technical adequacy of guidance and protocols that DFID and its partners use to oversee the capture, collation, review and validation of results • Extent to which DFID departments and partners abide by the guidance, protocols and processes for reporting and collating results • Extent to which the reported results	Strategy review • Analyse DFID guidance Programme Desk Review • Assess how well guidance is reflected in programme documentation at design, implementation and reporting stages • Assessment of how DFID staff make	Expert assessment Comparison of guidance with best practice from the literature. Analysis of data to identify which data has been subject to validation, of what type and by whom.

Review criteria, question/sub-question	Evidence required	Applicable component and activities	Analytical approach
	for 2011-15 have been subject to evaluation or validation ▪ The quality of the data sources that DFID uses to construct its results at project, country and institutional (for multilateral level) are comprehensive and able to provide the necessary data ▪ How DFID staff make judgements *in practice* on how to include data into reported reports for project, country, institutional and corporate level ▪ Extent to which these reported results have been subject to evaluation or validation, and extent to which they are estimates ▪ Assessments from third parties of the reliability and validity of the results claims	judgements of reported results Case Studies ▪ Cross-check of reported results derived from project documentation captured as corporate results Key informant interviews at portfolio, programme desk review and case study stages.	
2.2 To what extent have WASH interventions been strengthened in response to lessons learned on delivering results?	▪ The extent to which individual project design, implementation, reporting and evaluation reflects and incorporates learning from across the portfolio ▪ Quality of DFID learning processes within this portfolio	Strategy review ▪ Review of portfolio level documentation (such as DFID's Portfolio Review, policy guidance) to identify how DFID seeks to support adaptation of projects based on lesson-learning (a. corporately, and b. across the WASH portfolio)	Documentary analysis of policies, evaluations and guidance Documentary analysis from project documents ▪ Frequency analysis of adaptation to lessons learned appearing in sample (design, implementation and reporting documentation)

Review criteria, question/sub-question	Evidence required	Applicable component and activities	Analytical approach
		Programme Desk Review ▪ Capture whether and how project level documentation reflects lessons learned in design, implementation, reporting and evaluation Case Studies ▪ Detail on the coverage and depth of lessons learned as reported in the documentation Key informant interviews during the strategy review, programme desk review and case study assessment ▪ How DFID has adapted its WASH projects in the light of learning	Case study and desk review frameworks ▪ Identification of adaptation and process of change from key informant interviews
Sustainability			
3. Has the impact from DFID's WASH programming proved sustainable?			
3.1 Have DFID's WASH interventions been designed and implemented with a view to maximising the sustainability of results?	▪ Extent to which water and sanitation interventions include capital and operational maintenance after project completion ▪ Extent to which long-term governance and oversight is addressed in programme design ▪ Extent to which hygiene interventions	Strategy review ▪ Identify lesson learning and guidance at corporate level that seeks to ensure sustainability of results Programme Desk Review ▪ Capture of information from project level documentation (at design,	Documentary analysis from project documents ▪ Analysis of trends and patterns across desk reviews ▪ Qualitative assessment of the attention paid to sustainability in sample against best practice hypothesis (at design,

Review criteria, question/sub-question	Evidence required	Applicable component and activities	Analytical approach
	are planned and implemented to ensure that behaviour change continues to take place after project life ▪ Extent to which interventions include beneficiaries in planning, implementation, monitoring and oversight ▪ Extent to which water facilities continue to function ▪ Extent to which sanitation facilities are (i) maintained and (ii) used as planned ▪ Extent to which behaviour change in target communities on hygiene has continued after programme completion	annual review and completion report stages) to identify evidence of sustainability of results being included throughout Case studies of WASH interventions in one/two countries ▪ Site observations of water and sanitation facilities, against project specifications for projects that have been completed ▪ whether reported results can still be observed ▪ whether results have been verified post completion ▪ Whether institutional, financial and technical sustainability factors are in place	implementation and reporting stage)
3.2 Are there arrangements in place to monitor the sustainability of results after the end of the funding period?	▪ The point in the programme life at which DFID captures its DRF results ▪ Whether DFID has implemented evaluations of project and portfolio results covering sustainability ▪ Extent to which has DFID put in place processes to review sustainability after project completion at project level going forward ▪ Extent to which water facilities	Strategy review ▪ Identification of processes at corporate level to monitor sustainability of results after end of funding period Programme Desk Review ▪ Capture of information from project level (at design, annual review and completion report stages) to identify	Documentary analysis from project documents ▪ Pattern analysis ▪ Qualitative assessment of the attention paid to long term M&E (theory based assessment) ▪ Qualitative identification of process and gaps

Review criteria, question/sub-question	Evidence required	Applicable component and activities	Analytical approach
	continue to function ▪ Extent to which sanitation facilities are (i) maintained and (ii) used as planned ▪ Extent to which behaviour change in target communities on hygiene has continued after programme completion	evidence of systems being in place for long term monitoring. ▪ Capture of evidence from evaluations Case studies of WASH interventions in one/two countries ▪ Identification of long-term monitoring processes (documents, interviews, observation) by DFID or other partners	

Annex 2. Sampling strategy

This annex addresses:

1. The sampling of projects to be desk reviewed.
2. The selection of country case studies.

1. Sampling strategy for desk reviews

Sample size: in choosing the sample size, statistical validity is not a relevant consideration. We will not be using statistical methods to extrapolate from the sample to the population as a whole. Rather, we will be identifying recurrent patterns (strengths and weaknesses) in results and results monitoring across the portfolio. The sample should include enough projects to be able to identify those recurrent patterns, without being forced to rely on single instances. However, the portfolio is also dominated by a number of very large projects. The sample should include a sufficient proportion of the total results, so that the sample is material.

In determining materiality, both size of expenditure and size of reported results are relevant to materiality. The project-level expenditure data we have, however, is unreliable, as around 40% of the programme expenditure is outside WASH. The results data are specific to WASH, but are reported at country, rather than project level. We can make reasonable inferences as to results by project by comparing project-level expenditure data with country-level results data. We therefore propose to use results data, rather than expenditure data, to determine materiality.

While there is no single correct answer for the sample size, we would suggest that it should cover (i) at least 25% of projects (18 out of 71), and (ii) at least 50% of the reported results.

Sampling criteria: our sample will be selected against the following criteria:

a. **Materiality:** the sample should cover 50% of reported WASH results and 25% of WASH programmes.
b. **Country and central programmes:** DFID has five central or vertical programmes, which together are expected to contribute to overall WASH results. It is therefore important that the vertical programmes are included in the sample.
c. **Delivery channel:** we are asked to assess variations in the return on investment across different delivery channels (review question 1.1). Our analysis of the portfolio reveals the following delivery channels: partner government (budget support or other financial aid); multilateral agency; international non-governmental organization (INGO); national non-governmental organization (NGO); and contractor. According to DFID, 60% of WASH programming by spend is delivered by UNICEF, which should therefore be well represented in the sample. We also wish to cover projects managed by DFID country offices, DFID central departments and multilateral programmes.

In addition, the sample should include consideration of the following:

d. ***Water, sanitation and hygiene projects:*** the WASH results aggregate water, sanitation and hygiene interventions. We are interested in splitting these out and determining if there are patterns in results or results measurement specific to each.

e. ***Maturity of results:*** we are interested in assessing the sustainability of results. This requires looking at results that have been reported in the earlier years of the DRF period (2011-2015) and determining whether they are still observable. Our sample therefore should include projects that reported results two to four years ago. We do not have a detailed breakdown of DFID's results data according to the year in which results were reported by different projects. We can, however, use project completion status as a marker to indicate results at the appropriate level of maturity. Our sample should therefore include a number of completed projects.

f. ***Rural/urban:*** our preliminary review of the literature suggests that there are distinct challenges in achieving sustainable WASH results in rural and urban contexts. The sample should therefore include a mixture of both.

We have used a combined purposive and random sampling approach, as described below.

For the purposive section of the sample, to ensure materiality, we began by selecting the largest individual programmes, by claimed result, until we reach 50% of the total. This yielded six projects, located in Bangladesh, Zambia and Nigeria.

We then added three of the five vertical programmes. We chose the three most substantial, in terms of the number of countries they cover and their contribution to overall results. This yielded the following three programmes:

- WaterAid Programme Partnership Arrangement (PPA), which provides £21 million over five years, from which WaterAid is expected to provide 9.2 million with access to WASH across 23 countries.
- The Water and Sanitation for the Poor (WSUP), run by the Human Development Department and implemented through a public-private partnership with a consortium of firms and universities. This has a budget of £15 million, which is providing 16.5 million people with improved WASH access across nine countries.
- Accelerating Sanitation, Hygiene and Water for All in Off-Track Countries, run by the Human Development Department and implemented by UNICEF, which provides £37 million for improving WASH access in nine countries that are lagging on the MDG target.

In addition, we have included a single multilateral partner, the World Bank, which alone represents 89% of the WASH results reported from the multilateral budget.

To guard against biases introduced by a purposive sampling that prioritises large programmes, we added a further eight projects selected on a random basis, bringing the total number to 18 (25% of the total). This section of the sample was generated by giving each project a number then generating eight random integers using an online tool at Random.Org. The sample was limited to one project period country; when duplicates were generated, an alternative was selected. One randomly selected project – the Sudan Common Humanitarian Fund – was rejected on the basis that it had already been reviewed by ICAI and an alternative was chosen. The Bangladesh Chars Livelihood programme has also been reviewed before (four years ago), but has been included because it is material in terms of reported results.

The full sample thus generated is set out in Table A1.

We checked the resulting sample against criteria c. to f. to determine if it is representative. The results are as follows:

- *Delivery channel:* National NGO (3); Contractor (3); Multilateral (9, including 6 UNICEF projects); INGO (2); Recipient government (3).
- *Water, sanitation and hygiene interventions:* all three (11); water and sanitation (3); sanitation and hygiene (1); unknown (3).
- *Maturity of results:* no. of completed projects (4); no. of projects with start dates in 2012 or earlier (13).
- *Rural/urban:* both (7); rural (6); urban (3); unknown (2).

This provides a sufficient coverage of the remaining criteria, so no further sampling is required.

Country Case Studies

Case study selection was informed by the following guidance:

- Country visits provide a means of stress testing results claims and challenging DFID's processes.

- They are not expected to generate new impact data, but provide insights, stories and colour that complement the broader analysis.

- For this review, rather than validating impact data from impact evaluations, we will test the validity and significance of results claims under DFID's Results Framework.

- Particular care should be taken when generalising from the findings of case studies.

- Case study country selection would be made independently of the selection of the 18 projects for desk review.

- Two country visits should be made.

- The final selection would be judgement based on a balance of criteria (set out below).

Selection approach

The starting point is the list of 27 countries that DFID informs could contribute WASH results under DFID's Results Framework that have live WASH projects. Within this group, 20 countries are reported to have contributed results in the period 2011-15.

The selection criteria include necessary requirements, exclusions and desirable attributes:

Necessary requirements and exclusions:

- Must be DFID priority countries (there are 28).

- Must demonstrate longevity of programme, with at least one programme that has contributed results at least two years ago.

- Should not include insecure environments where field research is heavily constrained (as humanitarian programming is not a strong focus of the review).

- Should not include countries that have been subject to an ICAI visit in the past year, or more than three ICAI visits in total.

Selection criteria:

- Should have a sufficient density of programming across Water, Sanitation and Hygiene (across the two case studies).

- Should demonstrate a diversity of delivery channels (across the two case studies).

- Should have both vertical and bilateral programmes (across both).

- Should cover both rural and urban programming (across both).

We considered materiality of spend as a criterion, but decided against it on the basis that there was no reason to assume that quality of results management or sustainability of results would vary according to programme budget. We also considered availability of existing evaluations as a criterion, but decided against it since the case studies will not validate impact data and there was insufficient detail on the coverage and utility of the evaluations.

Selection of shortlist

From the list of 20 countries contributing results for the 2011-15 period, the following were excluded:

- Vietnam was removed as not a DFID priority country.

- Yemen, Syria, South Sudan were removed due to FCO advice against all travel.

- Sierra Leone was also excluded due to the emergency programming put in place in response to the Ebola epidemic, making country analysis problematic.

- Pakistan was removed because of restriction on field work in much of the country, because its WASH programming is purely humanitarian and because it was visited by ICAI in January 2015.

- Uganda was removed because its sole project (Uganda Budget Support) was suspended and only 3,397 results were claimed.

- The following were excluded because they have been visited more than three times in ICAI phase 1: Bangladesh (6), Ethiopia (5), Malawi (4) and India (4).

- Sudan was removed because it was the subject of a dedicated ICAI review: *DFID's Water, Sanitation and Hygiene Programming in Sudan* (February 2013).

- DRC was removed because one of its two WASH programmes (VEA – Village Ecoles Assainis) had previously been a case study project for an ICAI review: *DFID's work Through UNICEF* (March 2013).

This left a shortlist of five countries: Mozambique, Nepal, Nigeria, Tanzania and Zimbabwe.

We were advised by DFID that:

- Access to some of the field sites in Northern Nigeria may be constrained for security reasons.

- Nepal is still in recovery mode from the 2014 earthquake and that we may have limited access to government officials and field sites.

We note that both Nigeria and Nepal are covered in the desk review sample. Mozambique and Zimbabwe demonstrated the highest density of programming of interest and have therefore been chosen as the case study countries.