

[BACK TO BLOG LISTING](#)

---

# Scraping By: What It Takes To Find and Use Donor Results Data

Mar 23, 2016 • Katherine Wikrent

A key component of DG's [Results Data Initiative](#) (RDI) is finding innovative ways to compare results data across donors and geographies. In order to support DG's successes in [crosswalking data](#) and conducting [qualitative field assessments](#), has created new approaches to find, extract and relate results data from different development organizations. But finding – much less exporting and using – results data from across the development community is no easy task. For example, consider [Publish What You Fund's \(PWYF\)](#) rigorous cross-donor analysis of transparent funding, the [Aid Transparency Index \(ATI\)](#). In its most recent report, PWYF found that only [15 of 68 included donors](#) excel in transparent reporting, whereas over a half rank in the “Poor” and “Very Poor” categories. PWYF notes significant challenges in transparent reporting, with major barriers to overcome in consistent publication of [activity-level and results information](#); only 6 organizations scored above 80% for cohesive results data reporting.

Donors across the board, from [World Bank](#) to [USAID](#), recognize that M&E should be a vital tool for improving operations and staying faithful to mission statements. The ability of a donor to track progress and recalibrate efforts to be responsive to local needs starts with the collection and dissemination of robust and accurate project-level data. And many donors are working to make this results information more transparent and accessible to external users.

Still, there are major opportunities for donors to improve the

highlights the challenges that keep data users from easily obtaining and using information about donor results. Here, we will highlight the key challenges we have overcome in our efforts to extract and relate results data, as well as provide examples of promising practices that reporting organizations should adopt and improve on.

## **I: Key data are unavailable to the public**

To crosswalk results across organizations, we search through project documentation to find information about quantitative outputs. A primary challenge with this scraping process is that critical documents and data are often missing or unavailable publicly. Of our sample of 395 Health and Agriculture projects in Ghana, for example, 88 (22%) were missing project documentation of any sort, and 222 (56%) were missing output-oriented documentation. The availability of project and documents varies widely across donors, but a recurring trend we have seen is that within organizations, data openness and publication policies are unclear or inconsistent.

The World Bank stands out as a promising organization for its [documented commitment to data openness](#). In addition to explicitly laying out its [policy on access to information](#), the World Bank also maintains several publically accessible data portals, including the [World DataBank](#) and [Microdata Library](#). Opening up its enormous repertoire of data to the public is an excellent strategy to promote institution-wide accountability and transparency, and may also result in the innovative use of public data by stakeholders across sectors. As a consequence of its data openness policy, World Bank scores highest on output availability: 23% of all projects boast publicly accessible output data.

DFATD scores second highest, with 17% of all projects containing output information. Unlike World Bank, however, DFATD does not report its information in through a central portal of project documents. Instead, DFATD publishes basic project information like funds allocated and start and end dates and output and outcome information directly onto individual

Though the contained information can be downloaded in CSV or XML formats, this reporting scheme isn't as desirable as the structured, semi-standardized forms provided by World Bank because their heavy formatting and largely qualitative nature do not lend themselves to automated scraping. An example of a typical DFATD project page is below.

**Project profile: CARE Canada - Program 2007-2012** [Share this page](#)

Download this profile in alternate format: [CSV](#) [XML](#)

Use [basic search](#) or [advanced search](#) to find other projects.

**Sectors** **Results** **Related Information**

**Transparency**

---

**Basic Information**

**Project Number:**  
S063755-PRG

**Maximum CIDA Contribution:**  
\$17,650,166

**Executing Agency — Partner:**  
[CARE Canada](#)

**Status:**  
Closed

**Start — End:**  
2007-07-01 — 2013-01-31

**Country/Region**

[Rwanda: 24.6%](#)  
[Kenya: 10%](#)  
[Zambia: 10%](#)  
[Zimbabwe: 7.2%](#)  
[Lesotho: 5.9%](#)  
[Mali: 5.1%](#)  
[Cuba: 4.2%](#)  
[Ghana: 4.1%](#)  
[Ethiopia: 3.6%](#)  
[Malawi: 3.3%](#)  
[Senegal: 3.3%](#)  
[Indonesia: 2.9%](#)  
[Nepal: 2.8%](#)  
[Haiti: 2.5%](#)  
[Cambodia: 2.3%](#)  
[Canada: 2.1%](#)  
[Peru: 2%](#)  
[Tanzania: 1.6%](#)  
[Bolivia: .8%](#)  
[Sudan: .7%](#)  
[Cameroon: .3%](#)  
[Honduras: .3%](#)  
[Pakistan: .3%](#)  
[Colombia: .1%](#)

**Description**

The aim of this program is to reach over 550,000 people in over 20 countries through projects that focus on health-HIV/AIDS, environment, natural resource management and economic development, while respecting the cross-cutting themes of gender equality, governance, capacity building and knowledge management. Women and girls and the rural poor are the main beneficiaries of the program. The overall goal is to improve the livelihood security of marginalized and vulnerable populations. Each of the sectors of health-HIV/AIDS, environment and natural resource management and economic development have a clear link to poverty and vulnerability.

**Sectors**

**Democratic governance**  
Human rights (015160): 4.28%  
Democratic participation and civil society (015150): .82%  
Public finance management (015111): .27%

**Improving health**  
Health education (012261): 9.25%  
Health personnel development (012281): 2.06%  
Reproductive health care (013020): 1.13%  
Malaria control (012262): .98%  
STD control including HIV/AIDS (013040): .52%  
Health policy and administrative management (012110): .39%  
Basic drinking water supply (014031): .11%  
Basic sanitation (014032): .11%

**Private sector development**  
Informal/semi-formal financial intermediaries (024040): 30.93%  
Food crop production (031161): 8.64%  
Agricultural development (031120): 8%  
Education/training in banking and financial services (024081): 5.95%  
Agricultural services (031191): 4.72%  
Agricultural inputs (031150): 2.8%  
Agricultural land resources (031130): 2.31%  
Agricultural policy and administrative management (031110): 1.42%  
Livestock (031163): 1.42%  
Vocational training (011330): 1.19%  
Agricultural extension (031166): .52%

**Strengthening basic education**  
Basic life skills for youth and adults (011230): 2.11%

Figure 1: DFATD Project Page

## II: Data are not available in open formats

Even organizations that do commit to publicly release performance data often do not offer these data in [open \(free, public, machine readable\) formats](#). Open data allow citizens to access and analyze data in an unrestricted fashion. The innovative use and reuse of open data can [spark innovation](#) in both the public and private sectors. The White House estimates that the value of innovative reuse of GPS data, for example,

information promotes transparency and accountability, and encourages the reuse of these data across the public, private, and nonprofit sectors.

DFID is a great example of an organization dedicated to open data formatting. In 2012, it officially declared its commitment to open data in an [Open Data Strategy](#), which includes a pledge to enter project data into the [IATI Registry](#), a global database of aid transparency information from nearly [400 publishers](#). DFID now offers [project documentation in open formats](#), including project descriptions in .odt.

### III: Donors don't use standardized report templates

Our analysis of both open and non-open project documentation is complicated by the absence of standardized report formats. Our data analysts are able to construct systematic and replicable data scrapes for donors that have an internally standardized system, avoiding the time and resource drain that comes from manually pulling data from documents not standardized at the organizational level. Complicating the matter further, as discussed in Section I, some organizations that do have standardized report forms don't always make the completed documents available to the public.

In tandem with our manual scraping efforts, the team is hard at work developing a series of [machine learning scraping algorithms](#). The potential of these algorithms is enormous: the ability to accurately scrape information from projects completed across the world in a fraction of the time it would take to do so manually. Project documentation released in standardized formats would enable these algorithms to function optimally.

UNDP does an excellent job of standardized reporting through offering project managers a set of [project documentation templates](#), including one for annual work plans and another for results frameworks. As shown below, the annual work plan for the [Community Resilience Through Early Warning \(Crew\)](#) project clearly and systematically reports activities and sub activities and their corresponding expenditures. Due to agency-wide use of these standardized formats, UNDP annual

data-mining algorithm.

Country	Project	Key Activities	Associated Data Party	Project Description	2015				Estimated Budget (USD)
					Q1	Q2	Q3	Q4	
Tanzania - Technical Education	Improving technical education and training for youth and women in the informal sector	Supporting technical education and training for youth and women in the informal sector	UNEP	Support technical education and training for youth and women in the informal sector					0.00
		Supporting technical education and training for youth and women in the informal sector	UNEP	Support technical education and training for youth and women in the informal sector					0.00
		Supporting technical education and training for youth and women in the informal sector	UNEP	Support technical education and training for youth and women in the informal sector					0.00
		Supporting technical education and training for youth and women in the informal sector	UNEP	Support technical education and training for youth and women in the informal sector					0.00
		Supporting technical education and training for youth and women in the informal sector	UNEP	Support technical education and training for youth and women in the informal sector					0.00
		Supporting technical education and training for youth and women in the informal sector	UNEP	Support technical education and training for youth and women in the informal sector					0.00
	Improving technical education and training for youth and women in the informal sector	Supporting technical education and training for youth and women in the informal sector	UNEP	Support technical education and training for youth and women in the informal sector					0.00
		Supporting technical education and training for youth and women in the informal sector	UNEP	Support technical education and training for youth and women in the informal sector					0.00
		Supporting technical education and training for youth and women in the informal sector	UNEP	Support technical education and training for youth and women in the informal sector					0.00
		Supporting technical education and training for youth and women in the informal sector	UNEP	Support technical education and training for youth and women in the informal sector					0.00
		Supporting technical education and training for youth and women in the informal sector	UNEP	Support technical education and training for youth and women in the informal sector					0.00
		Supporting technical education and training for youth and women in the informal sector	UNEP	Support technical education and training for youth and women in the informal sector					0.00

Figure 2: Documenting Excellence: UNDP Structured Documentation

While we applaud any organization that makes their project documentation available to the public, we find particularly challenging the methods certain organizations employ to release these data. Across organizations, we have seen instances in which even recent project documents are partially or completely illegible. Take, for instance, this standardized form employed to track outcome targets and projects. Though the form likely began as a “best practices” example—a standardized, open format spreadsheet that clearly linked achieved outputs to targets—over the course of the project cycle, it was printed, scanned, converted into an image file, and then size reduced. The result is a low-resolution, unscrapeable, and almost completely illegible file. As a result, we were unable to enter this project’s information into our database for further analysis, in spite of the fact that robust data for the project does exist.

Project Information									
Project Details					Financial Data				
Project Name	Project ID	Project Type	Project Status	Project Location	Budget	Actual	Variance	Comments	Notes
Project Name	Project ID	Project Type	Project Status	Project Location	Budget	Actual	Variance	Comments	Notes
Project Name	Project ID	Project Type	Project Status	Project Location	Budget	Actual	Variance	Comments	Notes
Project Name	Project ID	Project Type	Project Status	Project Location	Budget	Actual	Variance	Comments	Notes
Project Name	Project ID	Project Type	Project Status	Project Location	Budget	Actual	Variance	Comments	Notes
Project Name	Project ID	Project Type	Project Status	Project Location	Budget	Actual	Variance	Comments	Notes
Project Name	Project ID	Project Type	Project Status	Project Location	Budget	Actual	Variance	Comments	Notes
Project Name	Project ID	Project Type	Project Status	Project Location	Budget	Actual	Variance	Comments	Notes

Figure 3: When Good Data Go Bad

## Menu

sample size or quality documents remains low. Here, we define “quality” as public, free, machine-readable (or at least relatively easily convertible), standardized datasets and forms that clearly quantify numeric outputs. A summary of the project documentation we have collected for health and agriculture projects in Ghana is below. Providing quality data can only help organizations [track progress and increase accountability and transparency](#) in the short term, and generate deeper, most sustainable outcomes in the long term. [Specific benefits](#) of improving M&E systems include: uniting teams around a shared vision; creating an environment of knowledge sharing; guiding project and program strategy; efficiently allocating resources; engaging local stakeholders; increasing responsiveness to local needs; learning from project setbacks and successes; recognizing and fixing flawed systems; and meeting project targets and milestones on time. Given the massive internal and external benefits of opening data, we encourage all donors to take steps to reform their data management systems and release quality, robust output information for all their projects.

Organization	Projects	Projects with Output Documentation	Percent	Projects with Standardized Output Documentation	Percent
PEPFAR	8	8	100%	8	100%
WB	40	40	100%	39	98%
Global Fund	9	8	89%	8	89%
DFID	24	8	36%	5	21%
UNDP	19	3	18%	1	5%
AfDB	8	2	25%	0	0%
DANIDA	51	17	33%	0	0%
DFATD	131	28	22%	0	0%
FAO	23	0	0%	0	0%
IFAD	16	12	75%	0	0%
IFPRI	1	0	0%	0	0%
MCC	2	1	50%	0	0%
UN	11	10	91%	0	0%
USAID	24	8	50%	0	0%
WFP	5	3	60%	0	0%
WHO	10	2	20%	0	0%

*Figure 4: Cross Organizational Comparison of Quality Reporting*

## Next Steps: (Out)putting Your Money Where Your Mouth Is

Most donors recognize that the [benefits of democratizing data access](#) are colossal. In spite of this, no donor in our sample

[Menu](#)

challenges our team has faced in extracting output data from project documents highlights just how much opportunity there is for donors to improve the quality of their monitoring systems. Fortunately, the commitments many donors have made to open their data have illuminated concrete steps that organizations can take to improve the quality of their reporting schema. DG is hard at work synthesizing these “most promising practices” into a “0-to-data” toolkit. Though this toolkit will be the subject of a separate post, we conclude by offering a list of 5 steps that development organizations can take right now to improve their reporting.

1. Publicly release organization-wide [M&E standards and frameworks](#)
2. Adopt [standard definitions](#) of terms like “outputs” and “outcomes” and clearly designate tracking and reporting norms, including who monitors and reports data, how they do so, and when in the project cycle milestones must be met
3. Provide project management teams with reporting [toolkits](#) including templates, and checklists
4. Upload all project documents and datasets into a publicly searchable [portal](#)
5. Consider the ways project documents can begin to incorporate [quantifiable measures](#) and, ideally, [spatial data](#)

## Stay Informed



Menu

- innovation
- aid management
- open data
- results data

# Main Menu

ABOUT EXPERTISE REACH INNOVATION BLOG CONNECT

About | Contact Us | Privacy & Terms

© 2016 Development Gateway and/or its affiliated companies



Follow Us

