**The Devel(opment) is in the Details: A Detailed Look into our Metrics**

In parts 1 and 2 of this series, we explored how the robust, timely, and valid collection of output data enhances transparency and collaboration, lead to better outcomes, and even support claims of causality. In this last part of the series, we invite the reader to take a deeper dive with us into our scorecard methodology. Throughout this discussion, we’ll be using the actual data we collected throughout our review of donor sites and their project documentation. Please note that although we will discuss ways in which donors struggle to meet the ideal system standards outlines in part 1, our intention throughout the entire process was not to name and shame; rather, we do this to highlight opportunities for learning and growth and to help donors identify concrete steps they can take to improve operations and bring about better development results.

**Operationalization**

Having settled on our criteria outlined in Part 2 of [the series](http://www.developmentgateway.org/blog/), we moved on to operationalizing a way to generate a score for each element in the most objective and transparent manner possible.

Maintaining scoring consistency as we evaluated 16 radically different M&E systems was a critical concern. To promote fair scoring, we attempted to keep scales for each criterion as objective as possible. Examples of our efforts to minimize subjectivity include:

1. Using quantitative measures like percentages and counts whenever possible
2. Focusing more on the existence of tools and policies rather than their quality
3. Avoiding ambiguous measures like “high/low,” or “good/bad”
4. Avoiding normative statements and value judgments, such as “DPs should” or “the most important”
5. Constructing a terminology reference for commonly used phrases such as “outcome,” “output,” “geospatial,” and “open.”

At continuation is a summary of the operationalization of our scoring mechanism.

Table 2: Doctoring our Scores- How we Operate

| **Category** | **Category Definition** | **Criterion** | **Classification** | **Scale** |
| --- | --- | --- | --- | --- |
| Monitoring Data Accessibility | The ability of the public to access and understand the content of project documentation that contains monitoring data (aka- “outputs” or “project-level data”) Monitoring data keep organizations accountable and aid the innovative reuse of project data. | Site Stability | Ordinal | [0-4] 0: Completely unstable 4: Completely stable |
| Site Navigability | Ordinal | [0-4] 0: Completely unnavigable 4: Completely navigable |
| Central Portal or Site | Binary | [0,4] 0: No 4: Yes |
| Ease of Automatic Monitoring Data Extraction | The ease of writing computerized scraping algorithms. Automatic extraction increases the robustness and scalability of the output tracking. More robust and plentiful output data help donors draw conclusions about project effectiveness and efficacy. | Use of Open Formats | Ordinal | [0-4] 0: Never used 4: Always used |
| Distribution of Monitoring Template | Binary | [0,4] 0: No 4: Yes |
| Use of Monitoring Template | Ordinal | [0-4] 0: Never used 4: Always used |
| Monitoring and Reporting Clarity | The ease with which the public can understand and analyze output information contained in project documents. The more detailed and precise the output reporting, the more transparent, accountable the donor it, and the greater the potential for innovative reuse. | Legibility of Results Documentation | Ordinal | [0-4] 0: Totally illegible 4: Totally legible |
| Completeness of Monitoring Reporting | Ordinal | [0-4] 0: Monitoring data never reported 4: Monitoring data always reported |
| Internal Consistency with Terminology | Ordinal | [0-4] 0: Monitoring data never follow DP definition 4: Monitoring data always follow DP definition |
| Precision of Monitoring Reporting | Ordinal | [0-4] 0: Monitoring data always reported using vague language 4: Monitoring data always reported using exact language |
| Monitoring Dataset Transparency | To what degree datasets containing output information are available (existing) and accessible (able to be accessed by the public). The greatest potential for transparency and reuse exists where both output datasets and corresponding spatial data are available for all projects. | Availability of Monitoring Datasets | Ordinal | [0-4] 0: No evidence exists that monitoring datasets exist 2: Some evidence that monitoring datasets exist, or full evidence that monitoring datasets exist for some projects only 4: Full evidence that monitoring datasets exist for all projects |
| Accessibility of Monitoring Datasets | Ordinal | [0-4] 0: No public access to datasets for any projects 2: Limited public access (some datasets public) 4: Full public access to datasets for all projects |
| Availability of Monitoring Spatial Data | Ordinal | [0-4] 0: No evidence exists that monitoring spatial data exist 2: Some evidence that monitoring spatial data exist, or full evidence that monitoring spatial data exist for some projects only 4: Full evidence that monitoring spatial data exist for all projects |
| Accessibility of Monitoring Spatial Data | Ordinal | [0-4] 0: No public access to monitoring spatial data for any projects 2: Limited public access (some monitoring spatial data public) 4: Full public access to monitoring spatial data for all projects |
| Monitoring and Reporting System | The policies, frameworks, and toolkits for the monitoring and reporting of project-level information. These systems promote cultures of rigorous monitoring and learning. The existence of output reporting system frameworks and plans is typically a precursor to the development of more specific tools project managers can use to report outputs. | Standardized Terminology | Binary | [0,4] 0: Publicly available dictionary exists 4: Publicly available dictionary doesn't exist |
| Monitoring and Reporting Policy | Ordinal | [0-4] 0: No publicly available policy/framework exists 2: Publicly available policy/framework exists, but missing key information 4: Framework publicly available, and is robust |
| Open Access to Monitoring Data Policy | Ordinal | [0-4] 0: No public access, or no policy found 2: Limited public access (some materials not public), or policy found but ambiguous 4: Full public access |

**Weighing In: End User Purpose Informs our Weights**

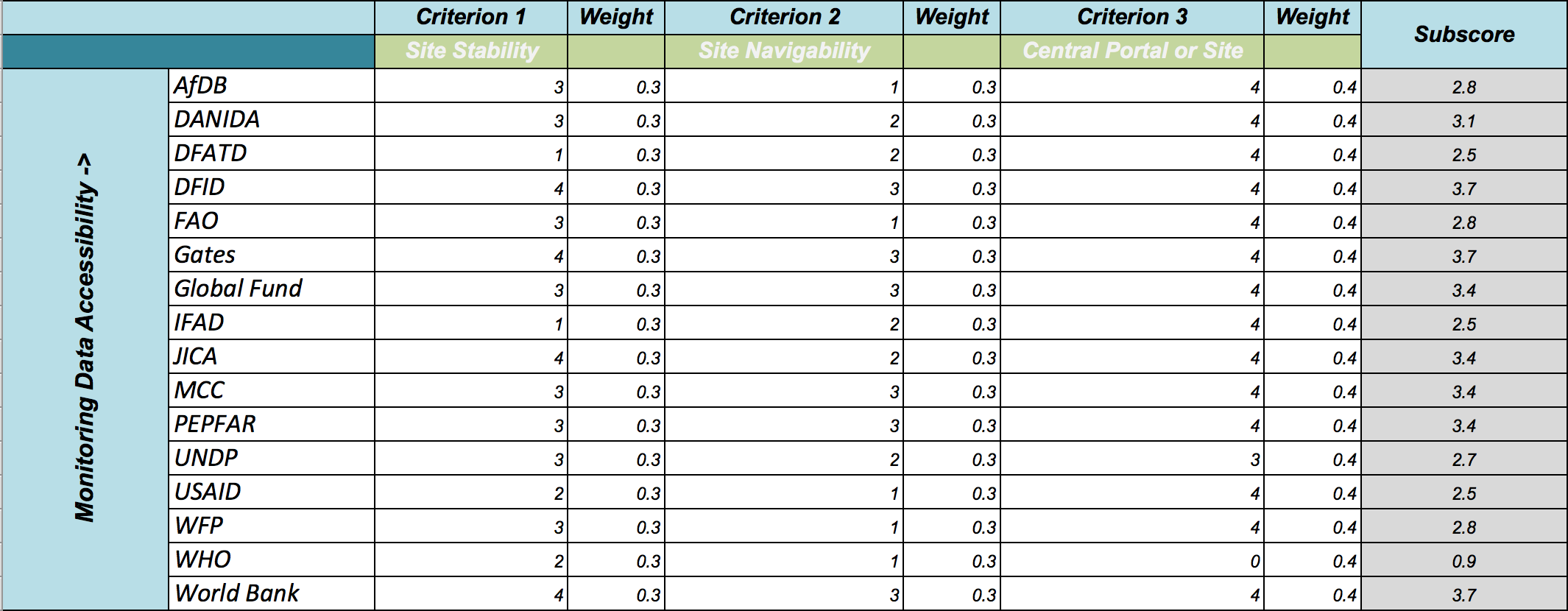
With our literature-informed operationalization constructed, we then settled on a weighting methodology for each of our 5 categories. In choosing our weights, we kept the end user in mind at all times. A key tenant of the [Data Revolution](http://www.undatarevolution.org/data-revolution/) is that development professionals must ask of themselves [“data for what, data for whom?”](http://www.developmentgateway.org/2015/09/18/designing-human-centered-data-revolution/)  At DG, we believe that data collection efforts must center on the needs of the users. Understand how data will be used and by whom can help ensure data efforts are targeted and efficient. Similarly, in assigning weights, we wanted to ensure that the elements of the monitoring systems that will most affect the end user held the most weight in the final DP score. For this reason, the weight assigned to each criterion is anchored to how essential that element is to the experience of the end user; in the “Ease of Automatic Monitoring Data Extraction” category, for example, the DP use of [open formats](https://www.data.gov/developers/blog/primer-machine-readability-online-documents-and-data) is more essential for the end user to make use of DP data than is than is DP use of a standardized monitoring template, so it receives a higher weight.

Table 3: Weight for it- Prioritizing End User Experience

|  |  |  |
| --- | --- | --- |
| **Category** | **Criterion** | **Weight** |
| Monitoring Data Accessibility | Site Stability | 0.3 |
| Site Navigability | 0.3 |
| Central Portal or Site | 0.4 |
| Ease of Automatic Monitoring Data Extraction | Use of Open Formats | 0.5 |
| Distribution of Monitoring Template | 0.3 |
| Use of Monitoring Template | 0.2 |
| Monitoring and Reporting Clarity | Legibility of Results Documentation | 0.3 |
| Completeness of Monitoring Reporting | 0.2 |
| Internal Consistency with Terminology | 0.2 |
| Precision of Monitoring Reporting | 0.2 |
| Monitoring Dataset Transparency | Availability of Monitoring Datasets | 0.3 |
| Accessibility of Monitoring Datasets | 0.3 |
| Availability of Monitoring Spatial Data | 0.2 |
| Accessibility of Monitoring Spatial Data | 0.2 |
| Monitoring and Reporting System | Standardized Terminology | 0.4 |
| Monitoring and Reporting Policy | 0.3 |
| Open Access to Monitoring Data Policy | 0.3 |

**The (Final) Product of Our Environment**

With all indicators scored for all DPs, weighted category subscores calculated, and a final average of subscores run to generate a final DP score, our scorecard work was finally complete. In the spirit of transparency, we will release the entire scorecard dataset soon. In the meantime, we would like to provide a visual overview of the scoring process. Below is a snippet from the final cross-DP scorecard set, with scores from the “Monitoring Data Accessibility” category.  
  
Figure 1: Picture perfect- A completed scorecard subsection



**(Data) Framing the Conversation with DPs**

With the scoring behind us, our priority now is to consult with DPs to discuss the results of the Scorecard exercise and hear directly about the challenges to data access and use they have encountered. The aim is to find ways to make project data available, open, and actionable. After all, if the data are not [useful](http://www.developmentgateway.org/2016/04/04/who-benefits-development-data/), [why collect them in the first place?](http://www.developmentgateway.org/2014/12/02/the-demand-side-of-the-data-revolution-lessons-from-the-government-of-nepal/)

Framing our conversations with DPs around the practical implications of improving their results tracking and reporting systems will highlight the pragmatic importance of the RDI project as a whole. We have seen in initial consultations that the “sexiest” hook of the Scorecard is the ability to “rank” DPs based on their composite score. As evidenced by this entirely too detailed post, we also spent a lot of time thinking about these numbers. We want to be clear that our interest in quantifying performance indicators does not come from a desire to compare DP performance or play a game of “Gotcha!” Rather, we developed this detailed numerical methodology to ensure we exercised internal consistency in exploring the many valid responses across DPs to the same set of [common challenges to results data reporting](http://www.developmentgateway.org/2016/03/03/results-indicators-costs-benefits/). In line with this, we have no interest in a cross-DP analysis other than to highlight extremely promising practices (See Part 3 of the series). A wise man (or, you know, at least a man) once said, “The points don’t matter.” We couldn’t agree more. We aren’t looking to debate the difference between a 1.4 and a 1.5; we are far more interested in exploring how each DP can leverage its existing tools and resources to tackle challenges to results reporting to meet their internally-defined goals.