

## Requirements for a universally applicable evaluation concept

© November 2011 IDEAS AidRating, Switzerland. Contact: [jan.stiefel@ideas-expert.ch](mailto:jan.stiefel@ideas-expert.ch)

### A Case for Open Data from a User's Perspective

One of the main obstacles to improve concepts for developments at all levels is that there is no common basis on which to compare and learn from experience. Such a basis would require a common standard of reporting, and a number as large and from diverse operational backgrounds as possible of cases to be studied.

This makes the case for a general standardized and widely used reporting standard, of which the best and most advanced today is the IATI standard. To express its potential fully, it should help boost evaluation concepts and performance. These, on the other hand, should be capable of absorbing the flow of data that can be expected. This is why the two concepts should be developed alongside and in view of each other.

Experts, planners and researchers have been discussing for many years how evaluation of development projects could be structured so that different projects and types of intervention and their characteristics could be compared. Up until today, the development community is waiting for such a standard.

### Open Data and standardized evaluation are linked

But what requirements would this standard have to meet in the first place? And for what purpose? The present discussion about open data gives new relevance to these questions for two important reasons:

On one side, comparison of evaluation results in a standardized format makes most sense if a maximum number of cases can be compared. This *per se* calls for a broad-based standardized reporting system such as the IATI standard.

The other side is that knowing what data are really needed would greatly help to clarify what needs to be included in such a reporting system: The usefulness of comparability through a comprehensive evaluation standard would be so overwhelming that once it becomes clear what is needed for such an evaluation standard, these needs should automatically determine the priorities to use or further develop datasets like the IATI standard. This paper examines the terms which a universally applicable evaluation standard should fulfil.

### Universal Applicability and Comparability

There is no doubt that the core issue of any evaluation is the in- depth analysis of usually more or less relevant components of a project or programme, or of all of them, combined with an assessment of results, and followed by conclusions what their advantages and weaknesses are in view of reaching the results, desired and others.

Usually, the focus is defined by the project design and by specific job lists (questions) asked by the project designers. The scope of the exercise is almost always limited to the set-up already in place, which means: The project is usually looked at within the boundaries of what project designers saw in the planning stage. What they have not seen is typically not looked at in the evaluation exercise, either. Very frequently, evaluation puts much emphasis on "management aspects", a term that contains anything between looking at specific personnel questions to accounting. Add to this a wide range in evaluation methodology, again influenced by project design and background of evaluation team members. The outcome is a

range of reports that can wildly vary in focus, methodology, depth, and, no wonder, conclusion.

Not provided, but needed if lessons are to be learnt reaching beyond each individual case, are two things: Broad, if possible universal, applicability of the concept, and comparability of its results.

All conceptual elements of an innovative evaluation approach would have to serve these two purposes. Deducted from this, the habitual priority given to management aspects loses much of its glamour. The underlying reason is that management is not a purpose in itself, but must serve one: It is the planner's judgement on setting, planned outcome, budget, and problems to be expected which define what management options are to be chosen, and not the other way round.

The weakness of many evaluations has its roots in this narrowed concept focusing on little more than "good management": An agency body at a comparatively "high level" typically commissions an "evaluation" of management performance at a "lower level". This is sometimes used to find fault for project deficiencies or even failure at that lower level. Often enough, the true origin of problems- poor baseline assessment and project planning- which was under responsibility of the "higher level" who commissioned the "evaluation"- remains conveniently out of focus.

An additional problem with such "focused" evaluations is often enough that the reader will never know whether an aspect or side effect has not been mentioned because it was irrelevant, or simply because it was not looked at, not being "in the focus" of the job.

So here is our thesis: A universal evaluation concept must start out from the problem to be solved and its setting, and must answer questions closely related to these. The management set-up is nothing more than a function to best do the job. The purpose, in other words, defines the best suited management options. Other adaptations may be a result of pragmatic choice under given circumstances, but nothing more.

For those, however, who might see their beliefs in management as being the key to proper problem solving fade away, there may be a word of comfort: management aspects, when looked at, would have to be looked at in a way far more comprehensive than most are used to.

### Underlying qualitative criteria

Any intervention is intended to bring some kind of improvement in a given development setting. This setting and its future or at least well-based assumptions on it must be known as a baseline for any properly planned intervention. To be comparable, the system boundaries must be wide enough to take care of any conceivable setting, operational mode, and outcome. This includes naturally all criteria that describe the setting of the human population and of target groups that may have been singled out in view of their economic, social, and cultural situation and perspective. The geographic and environmental setting must also be looked at, as they are key features.

At the same time, the concept would have to describe this complexity in a way that is understandable, without producing huge data mountains which for their mere quantity prevent a clear view on the important facts. Therefore: Needed is a formula that describes all important development issues comprehensively, but is still simple enough to be manageable.

### Dimension

A basic element for all comparison is dimension, be it numbers or size. Unless we know how large an intervention is, we cannot appreciate it. Do we want to improve the lives of 100

families, or of a million? Is the area which should benefit from our project one village, or a whole district, region, or country? The quantity of money spent is of little help here. It may give an indirect indicator on the size of an intervention. But it does not allow, taken for itself alone, to know if the intervention is coherent, and if funds are used effectively. Without having a measurement, we cannot distinguish fleas from elephants.

### Time vector

The timeline and point at which a project is reviewed is essential to appreciate results. At the beginning, even the best plan is nothing but a plan. Later on, results of project activities should become visible, first perhaps only very plain things like the opening of a new office, the hiring of staff, the purchase of equipment. This may grow into more targeted results such as students being trained, seminars held, tree nurseries established, and so on: Typically termed as “output” and “outcome”, these results are not the long term ones possibly intended, but undeniably “results” are there in some way in any project, whether successful or not.

It takes yet more time to see intended results to become visible. They are less than the impact, but more than mere output. It is land reclaimed or irrigated, graduates leaving schools and colleges, extension systems providing useful services, etc.

It takes still more time to see true “impact”: It would be what is so readily given in project plans as intended “results”: Have the farmers become better off, the local food basket more varied, children healthier, and so on? There must be awareness of links between project activities and intended impact, as this is the developmental justification for any intervention. And it is a daunting task to measure or at least reasonably estimate these impacts.

No matter what intervention is looked at, the point in time it is looked at defines what results can be expected. Therefore, a project and its merits can not be fully understood without understanding the time span and past activities that have led to the results that will be found on the ground.

### Role of “management”

Form follows function. Management should follow the needs of given objectives in a given environment. This is the most important requirement, and it includes under “management” everything from the first fact finding mission and planning onward. What good would be a well and efficiently managed intervention that has adverse results because it was ill-targeted from the beginning? Management must ensure that resources are used effectively, deadlines met, results are firm, and that negative side effects are avoided. It may also have to set the base for continuity, either by introducing a locally rooted management, or by ensuring that results remain once the project is gone.

Management needs first to be looked at as a process that oversees planning till hand-over, and that serves specific goals. One of the best parameters is then to relate the achieved results (output, outcome, and impact all together) to the cost at which they were produced.

### Data quality and traceability

Understanding a complex and large project in a difficult environment often suffers from lack of information on the many questions that need answers. Sooner or later there is a practical limit to the amount of data necessary to make an informed judgement.

Does this mean that no analysis is possible except where data are abundant? No. In everyday field work, much information that is not available first hand is substituted by more or less well founded assumptions. Most projects would not be possible without such

assumptions. This is always a weak point, and where good practice is applied, care is taken to distinguish where firm knowledge is the basis, and where assumptions were taken, and be able to verify these later.

Why should it be less permissible to work with assumptions when doing an evaluation? Most evaluations if not all are also based upon assumptions. The problem is not so much whether we have worked with assumptions, but rather: how well founded these assumptions are, and what care we take to improve on them. To be realistically applicable, not even the most comprehensive evaluation concept can avoid working with assumptions. They may, just as in the planning phase, substitute true knowledge.

Can we then never come to useful conclusions and good projects, since we never have enough of the right data? Yes we can. The key in both cases is: keep track of where assumptions have been made, and learn to improve their basis as you go. Good evaluation does this, among other things. The values of such a concept and of Open data would greatly enhance each other.

### Continuous improvement

A last requirement must be noted for this “universal evaluation standard”. It is the in-built capability of allowing to adapt to future improvements and developments while keeping older findings as a valuable source of learning. The term of the day is “forward compatibility”. In specific, all relevant components of the standard must allow for new knowledge to be included, for more accurate measurements (if necessary) to be applied, for finer distinction of differences to be made, and so on.

We of AidRating think that an evaluation standard complying with all this, combined with a standardized database like IATI as widely used as possible, the potential to make a huge leap in development learning and knowledge would be immense. A large number of cases to compare and test brings more and better analysis options; a good evaluation standard based on these data shows their usefulness.

This thing “would fly”. Because of this, we have looked for an evaluation concept which is capable to fulfil all these requirements. We took a close look at these requirements and have found one. It is now developed and refined to a stage where it is ready to be presented to the wide public. An overview and example of use can be downloaded from the AidRating homepage under

<http://www.aidrating.org/downloads/articleimpactratingwithexamplesv22.pdf>.

Information used from the database has a strong field orientation and results focus which should make them particularly interesting for CSOs in partner countries. The information needed according to the AidRating concept to understand a given project can be downloaded here: <http://www.aidrating.org/downloads/10keyquestions.pdf>.

More details on the AidRating Project Impact Rating concept will be presented at the Civil Society Forum in Busan on 26<sup>th</sup> and 27<sup>th</sup> November 2011. For questions and comments feel free to contact [jan.stiefel@ideas-expert.ch](mailto:jan.stiefel@ideas-expert.ch).

Keywords: AidRating, accountability, aid effectiveness, aid transparency, CSO involvement, evaluation, standardized reporting, Jan Stiefel, transparency

© IDEAS AidRating 05 Nov 2011