



Canadian International  
Development Agency

Agence canadienne de  
développement international

## **CODING AND COUNTING LESSONS**



## **FROM CIDA's PSD REVIEW**

Canadian International Development Agency  
200 Promenade du Portage  
Gatineau, Quebec  
K1A 0G4

Tel: (819) 997-5006

Toll free: 1-800-230-6349

Fax: (819) 953-6088

(For the hearing and speech impaired only (TDD/TTY): (819) 953-5023

Toll free for the hearing and speech impaired only: 1-800-331-5018)

E-mail: [info@acdi-cida.gc.ca](mailto:info@acdi-cida.gc.ca)

**Canada** 

## **Summary**

### **CODING AND COUNTING LESSONS FROM CIDA's PSD REVIEW**

#### **Background**

The experience gained with data analysis in previous performance reviews (Women in Development and Gender Equity, Basic Human Needs and Food Aid) plus a cursory assessment of existing private sector development (PSD) data suggested that the PSD review would suffer because of the quality of information available in the area. Before commencing the review, therefore, Performance Review Branch (PRB) took a number of steps to try to improve the situation. It had the Corporate Memory unit develop a preliminary database of PSD investments for the last ten fiscal years (1989/90 through 1998/99). This was done by using a predefined set of conventions and applying them to the data contained in CIDA's legacy systems (AIDIS and Corporate Memory) as well as obtaining additional information from the program branches where necessary (for example, PSD data related to core funding initiatives; PSD initiatives that may have been carried out within Canada Fund programming, lesson learnt, etc.). The new Agency coding structure was then applied to the projects identified to test the veracity and usefulness of the new structure and to see if a richer set of data could be generated. The typology of PSD investments was also extended so that more meaningful portrayals of PSD programming could be developed than would have been the case given the existing priority coding structure. The information generated was then provided to individual evaluation teams who validated the information in conjunction with the program branches.

This paper provides the results of this exercise along with the lessons that have been learned. The findings are instructive especially in terms of identifying the barriers to improved data, the strengths and weaknesses of the new coding structure, and the types of information that CIDA is likely to be able to generate as a result of its informatics renewal efforts. It also suggests a number of means and ways that the quality of information within the Agency can be improved. In this regard, most, if not all of the findings can also be applied to CIDA's other priority areas and data collection efforts.

#### **Problems Encountered**

Although there appears to be a good understanding of what constitutes PSD programming throughout the Agency, the capture and coding of this information into meaningful categories that can be aggregated and used for program analysis and reporting purposes presents a number of problems. Most of these have very little to do with either existing information systems or the technological infrastructure within which the Agency operates, but rather the information management principles and procedures used to define PSD and to collect and organize the information.

### *Lack of Understanding Between Coding Tables Used*

Generally, there is a lack of understanding between the tables used (priorities, sectors, themes, delivery mechanisms and beneficiaries) as well as how inputs, outputs and outcomes relate to these and how they should be coded to provide an accurate and complete set of data on CIDA's PSD investments. There also appears to be different understandings and perceptions of the contributions of projects to individual sectors, themes and priorities and how these contributions should be counted. This has resulted in data, that may for any given project be accurate from an individual project point of view, but provides little consistency when data is rolled-up across projects to respond to essentially simple questions about the extent and nature of PSD programming within the Agency.

### *The Way Documents and Data are Captured and Managed*

The way in which data is captured and stored within CIDA was found to be perhaps the single largest factor contributing to such quality issues as completeness, accuracy, and timeliness. Different types and levels of data are captured depending upon the type of investment (e.g., between Canada Funds, SPPEs, and whether the project is directive or responsive); the investment's size (under/over \$500,000); and, the Branch within which it is being carried out. This means that when data is aggregated, without considerable reconstruction, it can only provide a partial view of the programming within any given area. These differences are also reflected in the way in which project files are generated and saved. There is no one standardized covering page between different types of projects and programming that has consistent coding. Often it is not known what documents are on file for any specific project or which documents should be on file and different branches manage their documents differently.

### *Differences Between Branch Conventions*

Different desks and branches have different conventions and rules regarding how these documents are collected and stored for retrieval purposes. This problem is further compounded because different branches define PSD programming differently or have developed conventions that skew the data provided. CEE, for example, has a tendency to code its PSD investments in terms of economic integration, INC in terms of private sector linkages; and within Multilateral programming priority data is available for only the last three years and only at an aggregate level which means it can not be disaggregated by country, organization, sector, or beneficiary.

### *The Complexity of CIDA's Programming*

The complexity of CIDA programming also made the identification and categorization of PSD investments difficult. CIDA currently undertakes a wide range of projects in support of private sector development and these are not always easy to distinguish. Different types of PSD funding are carried out within the same project or in conjunction with another type of programming (much of micro-credit financing, for example, is not coded under PSD because such programming is seen as supporting other

priorities (e.g., Basic Human Needs). PSD programming can also be hidden within other programming areas because some desks are less reluctant to code private sector programming than other types of programming<sup>1</sup>.

### *Differences Between Data Collection Tools*

The definitions, coding and conventions used within the various data collection tools of the Agency (project approval documents, logical frameworks, project summaries, closing reports, Priority Coding and Counting, and Annual Project Performance Reports) also created problems. Most of these tools do not apply Agency coding in the same way. There is no requirement to re-code projects within closing reports, for example, even though the project may have changed dramatically from how it was planned. Likewise, within the APPRs, only one priority area was identified in terms of results achieved<sup>2</sup> whereas within the priority reports and approval documents any number of priorities can be identified for a single project. This meant that results from different databases could not be easily reconciled or married to provide a more complete picture of the area.

### *The Application of the New Coding Structure*

Although better than the previous structure, the application of the new coding structure and the conventions surrounding individual tables was not without difficulty. The priority table does not adequately reflect PSD investments or the results the Agency is trying to achieve in the area. Categories between and within tables also overlap and given the absence of clear definitions this can and does force officers to choose between competing priorities (e.g., where private sector development is targeted at women) which reduces the amount of funding that can be attributed within each priority area.

The rules of formation in creating separate tables have also not been adequately adhered to so that the flexibility that is gained by having multiple tables is significantly reduced. This is especially true between the sector and priority table where PSD programming can be found in each table without a clear description of what the differences between the two are or what the separate coding contributes to, except additional work.

The application of the new coding to past projects was also problematic. In creating the new coding structure an effort was made to map past coding to the new coding and to convert the old coding through a conversion exercise. While partially successful, (when there was either a one to one correlation between terms or a specific term was mapped to a more general term) the approach failed when a more general term was mapped to a more specific term. As a result of this, a decision was made not to use the new sector coding adopted by the Agency even though the table represents a qualitative improvement over the previous sector table.

---

<sup>1</sup> Some officers within CIDA believe PSD is not a legitimate activity for CIDA to undertake.

<sup>2</sup> This inconsistency will be fixed with the introduction of the new RBM module within SAP which is being introduced in the summer of 2000.

### *Aggregation Conventions*

Aggregation conventions surrounding individual tables also created problems. Different tables use different counting conventions. The priority table, for example, uses proportionate counting whereby expenditures are attributed based upon the proportion of the project's budget each priority area identified represents, so that when all priority areas coded are summed they equal 100% of the project's value. Other tables, however such as beneficiaries uses a marking mechanism. When beneficiary data is aggregated, therefore, the entire project budget is assigned to each beneficiary even if there is more than one beneficiary subscribed to the project. This means that when PSD investments are profiled (e.g., the beneficiaries of different types of PSD priority programming) an inaccurate picture is provided. This severely limits the usefulness of the findings and can lead to misunderstandings related to what is being identified.

### *Incomplete Coding Structure*

Finally, the lack of an adequate results table (that classifies the objectives of individual projects) and an activity table (that describes how intended results are to be achieved) meant that it was difficult to analyze the different types of PSD programming and to assess their strengths and weaknesses. As a consequence, only anecdotal information could be provided for the typologies produced.

### **Areas of Convergence**

Despite these problems, a considerable amount of consensus related to coding issues amongst the individual evaluation teams did emerge. First, the use of multiple tables (e.g., sectors, themes, priorities, intended results and delivery mechanisms) to portray PSD was seen as an improvement over a single integrated coding table that had been advocated by some specialists within the Agency. This appears to validate the overall approach that was taken by Phoenix in developing the new coding structure.

At the same time, evaluators believed that individual tables could be improved. In particular, they found the existing priority table was not as useful as it could have been in portraying PSD investments. As a result, most made recommendations about how the table could be improved. These ranged from collapsing the number of categories from five to four by integrating skills development and micro-enterprise development to making the table more results oriented.<sup>3</sup>

They also believed the level of granularity associated with other tables (themes, sectors, beneficiaries, organization types) could be improved so that more "meaningful portrayals" of PSD investments could be developed than is the case with the new tables adopted. There was little agreement between them, however, on what attributes each table should provide.

---

<sup>3</sup> It was finally determined that without examining all the tables used by CIDA, it is impossible to provide any rational input into how the table should be revised.

They also believed that in the absence of an activity table (that identifies training, scholarships, equipment, conferences, technology transfer, construction, and the like) it is extremely difficult to identify and assess the different "means and ways" by which the Agency tries to achieve intended results within individual priority areas. In fact, they argued that most of the lessons the Agency can learn have more to do with "how" results are trying to be achieved than anything else. Yet it is difficult to even try to identify these different means and ways except for time consuming file reviews.

Specialists also believed that a distinction between the recipient of the assistance (which is often an institution or organization) and the intended beneficiary was useful in helping to understand and depict the nature of CIDA's programming. This varies with the Agency's adoption of a direct and indirect beneficiary table which is not as "tight" a depiction and open to more interpretation.

Evaluators were also of the same mind in believing that the types of data collected on CIDA's programming needed to be the same regardless of the size, or type of programming or delivery mechanism. The lack of relevant information from the Partnership and Multilateral Branches and to a lesser degree Central and Eastern Europe caused significant problems as did the lack of information on the inputs of all of CIDA's projects. This was particularly true for the technology transfer and MBA components of the review where data had to be generated from project files thereby increasing the costs of the reviews.

Finally, all performance review team members felt that the Agency could no longer afford to maintain the current approach to coding whereby officers/partners categorized their projects without consequences and/or any quality assurance program or support. The hidden costs associated with such a governance structure in terms of having to carry out retrospective analyses of data to obtain more valid information was seen as simply too high. They also pointed out that such analyses could be potentially embarrassing to the Agency in that they almost always provide an alternate set of data than "official" Agency data that has been generated from officer/partner coding.

## **Lessons Learned and Main Recommendations**

Many of these findings and observations, of course, are not new and have been or are being addressed by the Agency through its information renewal initiatives. With the new Agency Information System (AIS), Branches are expected to use the same tables and definitions for their coding. Both a beneficiary and results table are also being defined and the beneficiary table will soon become part of the coding within SAP. The transparency of the AIS will also improve the coding as more and more officers see aggregate data.

Unfortunately, much of the coding being requested is optional, there are few corporate definitions of the coding to be used, there is considerable overlap between and within tables, and no mechanisms are in place to ensure completeness, consistency, timeliness or accuracy. As a result, it is likely that significant gaps in the Agency's data and hence knowledge will continue unless significant improvements are made. In this regard the experience of the PSD review is informative.

### ***The Need for a Uniform Set of Data***

For both reporting and performance review requirements, the Agency needs a uniform set of data and coding to be collected on all of its investments regardless of type, size or delivery mechanism. In the past, many reasons have been provided why this can not be done. The most often given is that the nature of programming within the Agency is so different that the data collected also needs to be different. Both the Canadian Partnership and Multilateral Branches, for example, will argue that because of the nature of their programming the types of information they require is different than "typical" bilateral projects. Both point to the responsive nature of their programming as well as the speed by which certain investments need to be delivered (emergency food aid) to justify data collection differences. The Bilateral Branches also use this argument for not coding Canada Funds and SPPEs. The Canadian Partnership Branch can also argue that the sheer number of the projects they deliver annually, and the "unique" nature of "core funding" is an impediment to data collection and coding. This is especially true since much of the coding needed, has to be done by partners who may not have the capacity or resources to carry out such coding. All branches will also point out that smaller projects need not be coded since they represent such a small proportion of the funding provided.

While all of these arguments are true, the similarities between the programming (from a results perspective) are still greater than the differences. Means and ways, therefore, must be developed to ensure that all investments are coded so that there is a complete set of data related to intended beneficiaries and priority areas. Such a convention will also allow the Agency to learn by comparing its intentions to what it actually does.

1. It is recommended, therefore, that *a predefined set of data elements as well as coding on all investments regardless of size, type or Branch be identified and made mandatory.*

*In conjunction with this, the new CIO Branch should be mandated with identifying the means and ways by which this common set of data could be collected, coded and processed across all Branches. This could be as simple as creating a single covering page on all projects that is mandatory before a project is approved regardless of the Branch or program. The CIO should also ensure that all data collection and reporting tools be redesigned so that they use the same definitions and coding and counting conventions.*

### ***The Need for Better Project Management Systems***

The PSD review highlighted the fact that many of the project management systems of the Agency have deteriorated over the years. In particular, the ways in which documents are collected, categorized and stored is problematic. The existing Agency classification system needs to be reviewed and applied more consistently from one desk to another. The quality of project files also needs to be examined and conventions adopted associated with what is put on the official record and how these are to be stored after a project is closed. Since more documents are being kept electronically procedures also have to be created on how and where these need to be stored for easy access.

While the Agency's Corporate Memory system collects key bilateral documents (approvals, management plans, monitoring and termination reports, evaluations and reviews) and categorizes them by type, author, sector, country, priority so that analysts and officers can be made aware of who did what, where and when this process needs to be extended to CEE, CPB and Multilateral Programming. Procedures also need to be put in place to ensure that all key documents are forwarded to the system.

2. *It is recommended, therefore, that the CIO Branch is mandated with reviewing the existing document management systems and developing standards surrounding the management of project documents produced by and for the Agency so that they can be made available to all officers and managers.*

*Such standards should include how documents are produced, collected, stored and classified both electronically and in paper form.*

### ***The New Coding Structure***

Although the new coding structure represents a significant improvement over the past coding structure, the PSD review suggests there is room for improvement. In particular, tables need to be streamlined to reflect different views of the information being captured and to decrease the amount of overlap that exists between them, they need to be recast on consistent and well understood definitions and rules of formation, and where required, coding conventions should be developed and new tables created to allow the Agency to better portray and analyze its programming.

More specifically, the priority coding structure needs to be reviewed in terms of actual priorities and/or modified to reflect intended results in each of the priority areas. The DAC sector table adopted reflects a multitude of activities, economic sectors, social sectors and themes (which reflects a compromise between the different reporting requirements of various donor organizations) and as such is less useful to cross reference different aspects of programming than it could be. The area of good governance is a case in point. An appropriate set of tables would allow one to select improving "public sector financial management" in "health care" as opposed to choosing between them or selecting both sectors. This would significantly increase the flexibility of the coding supplied and provide more accurate descriptions of the nature of the programming.<sup>4</sup>

It should also be determined whether the Agency needs or does not need a results table and an activity table. The experience of the PSD review suggests that if the Agency is to choose between competing alternatives in delivering intended results it should have such tables.

Finally, the counting rules and conventions related to all tables need to be reviewed as does the level of granularity associated with them to ensure that performance review and learning requirements are met as well as existing reporting requirements.

---

<sup>4</sup> The coding and reference group in adopting the DAC table had already decided that it should be modified to better meet CIDA's needs. This initiative (with the exception of the elimination of the gender section) has not yet been carried out.

3. *It is recommended, therefore, that the coding reference group be reconvened and in conjunction with PRB ensure that:*

*1) coding tables adopted meet performance review as well as corporate reporting needs;  
2) are constructed in such a way to maximize data generation (e.g., that the sector and priority tables provide completely different views of CIDA's programming); and,  
3) all tables needed (e.g., activity and results tables) are available.*

### ***Improving the Quality of PSD Coding and Counting***

Past performance reviews have suggested that the quality of information available to the Agency can be improved by introducing mechanisms that will assist officers and partners to do a better job. These have included:

- w** producing a better set of definitions and conventions to be used (including appropriate examples);
- w** providing better means and ways to access such definitions and conventions<sup>5</sup>;
- w** introducing training programs, courses and workshops that explain coding and the interrelationships between individual tables and codes; and,
- w** feeding back the results of periodic reviews of the Agency's data and changing the data so that it corresponds to the review's findings.

They have also suggested that the reward system within CIDA may need to be changed so that officers are held accountable for the appropriate coding of their projects.

The experience of the PSD review, however, suggests that while such mechanisms would help, their introduction may not improve the overall quality of information significantly. Even with a small team of coders it was found that clear definitions and conventions did not always result in consistent coding since individuals brought their own set of views to the exercise. Easy access to such definitions is also not a panacea if coders bring different interpretations and interest to the exercise. Even PSD specialists within the review disagreed about what should and should not be included in a definition and even had different interpretations of what coding should be used for the same project. Past training programs (e.g., on the LFA) have also been ineffective in improving deficiencies. While feeding the results of periodic reviews back into the system may be effective the timeliness of such efforts suggests that ongoing reporting would suffer.

---

<sup>5</sup> Definitions, for example, could be added to SAP so that when coding options are provided the definition associated with each code is also available.

Other organizations have also found that the more coders there are the less consistency there is. Staff turnover, individual officer interest, the politics surrounding coding, and more urgent project management demands also mitigate against officer/partner coding.

4. *If the Agency wants to improve the overall quality of its information available, therefore, it is recommended that a two pronged approach to coding be implemented whereby officers/partners are responsible for the coding of their projects at the highest level only (the six priority areas, the main sectors, the main organization types) and another group is responsible for any detailed coding. Such a group could consist of either policy analysts or data management specialists. The World Bank, for example, has implemented an approach whereby officers and specialists are jointly responsible for planning, reviewing and coding projects. This model works better for Bilateral programming than Partnership programming, however, and it is not clear whether a more consistent set of data would be generated based upon the different views of specialists/officers. It is suggested, therefore, that the introduction of a data administration unit might be a better alternative.*

Such a unit would appear to offer many benefits. First, it would reduce the amount of work required by both officers and managers while still obliging them to examine their projects in light of corporate reporting requirements. Second and perhaps more importantly, since the more extensive coding would be carried out by coding specialists, the data generated would be more consistent, complete, accurate and timely. The costs of the unit would also be offset by the costs saved in not having to do so many retrospective analyses of data which currently constitutes millions of hidden information management costs.

Such a unit could also provide value added services by maintaining the existing coding structure and adapting it to emerging needs as new information requirements are identified. It could also address completeness, accuracy and reliability issues as well as being mandated with the annual collection of development information from other government departments; coding information at an even more relevant level if Agency information were required (e.g., aids; micro-nutrients; children at risk; etc.); and, generating trends and special analyses as and when needed. The unit could also promote good data management practices within the Agency and assist officers and specialists in meeting their information requirements rather than having them hire independent consultants who develop their own definitions and conventions in collecting and generating relevant information.

### ***Improved Linkages Between Coding and Qualitative Data***

Regardless of the mechanisms implemented to improve coding, the PSD review suggests that better linkages between the coding supplied and the qualitative data collected needs to be made. This needs to be done in two ways. First, the qualitative data collected has to support the coding provided. All too often in the review, for example, women would be identified as a beneficiary of the PSD project (through the coding provided) but no qualitative description of how women would benefit was available. Second, the outputs and outcomes provided within an investment's description should be reasonable

given the resources of the investment. It makes little sense, for example, to contend that CIDA is going to improve the competitiveness of a particular group (entrepreneurs, women, foresters, etc.) by bringing civil servants from particular countries to Canada and teaching them "how Canada does it".

5. *For every priority area identified, therefore, it is recommended that a convention is adopted whereby there has to be a corresponding expected outcome<sup>6</sup> ideally defined in terms of the expected change in the condition of the beneficiary/recipient rather than some ultimate goal.*

### ***Bridge to the Past***

The PSD review also found that the bridge between new coding and past coding was not rigorous enough to support automatic roll-ups of data that are reliable and can be used to generate historical trends. This was especially true for sector coding but it was also true for priority coding and will be true for beneficiaries and thematic coding as well. This will mean that historical trends will be impossible without special independent studies. Such studies, however, have been found to be unreliable because they tend to use their own definitions and conventions as well as almost always generating a positive bias towards the subject area under review which overestimates its importance within CIDA programming. This was found to be true even within the PSD review, whereby individual evaluation teams saw certain investments as either "theirs" or "someone else's" based upon their view of the subject matter being reviewed.

6. *Given the adoption of the new coding tables, the International Development Information Centre (IDIC) of CIDA should undertake a study to determine the costs and benefits of creating a better link between current and past coding than simple conversion tables can provide.*

### **Conclusion**

CIDA's information systems should be able to provide simple answers to such simple questions as what is in the pipeline for any type of programming; what countries and executing agencies receive the most funding; what types of investments do partners do well and not so well and why? Currently, this is not the case nor is it likely to improve significantly with the implementation of the new Agency Information System. For as the PSD review shows, the quality of information available is more of a function of the way in which data is collected, coded and aggregated and the value individual officers and managers place on generating such information than the technology used. If CIDA really wants to improve the quality of its information, therefore, it is these things that have to be addressed.

---

<sup>6</sup> A corollary to this, of course, would be that outcomes would be supported by the outputs of the investment so that the relationship between data elements is also obvious.