THE CHALLENGE OF SHARING DATA IN COOPERATION PROJECTS: CAUSE FOR REFLECTION

M. Domini*, S. Gibellini*, F. Villa*, A. Vidoni*, L. Di Francesco*, S. Sorlini*, M. Vaccari*

* CeTAmb LAB, Research Laboratory on Appropriate Technologies for Environmental Management in resource-limited Countries, DICATAM, University of Brescia, via Branze 43, 25123 Brescia, Italy, cetamb@unibs.it

Sommario

La raccolta e gestione dei dati continua a rappresentare un problema importante per i progetti nei paesi a basso reddito. La raccolta dei dati richiede grandi investimenti, e nonostante gli sforzi profusi nella creazione di strumenti per la condivisione, resta ancora molto da fare a livello locale. I dati hanno un grande valore al fine di sviluppare strategie che portino ad uno sviluppo locale effettivo. Il CeTAmb LAB, sulla base della propria esperienza, propone una riflessione sul ruolo dell'accademia, basata sui principi della costruzione di competenze e condivisione della conoscenza, nel promuovere l'importanza di una corretta gestione dei dati.

Abstract

The collection of data and their management remain a never-ending challenge within projects carried out in low-income countries. Data collection requires high investments, and even if great efforts have been done in building tools for data sharing, these processes need to be improved at local level in developing countries. Data are valuable items for developing strategies leading to more effective local development. The CeTAmb LAB, on the basis of its experience, proposes a reflection on the role of the academia, which is based on principles of capacity building and sharing knowledge, in promoting the importance of data collection, conservation and sharing.

Keywords

Data management, sharing, reliability, access, cooperation

Introduction

The collection of data and the assessment of their reliability remain a never-ending challenge within projects carried out in low and middle-income countries. Developing countries are not only faced with the challenge of insufficient and unreliable environmental monitoring but also with the challenge of sharing such data and information (Michener, 2015). Data and information management is poor and disseminating data is still a challenge (Ndzabandzaba, 2015). The

collection of data, mostly technical and environmental but also economic, social and institutional, is of primary importance in order to design and develop sustainable technologies and systems. Data are valuable items, important for developing strategies and solutions within projects leading to more effective local development.

The collection of new data, and site-specific data, requires high investments in terms of time, human resources and money. In low-income contexts, even the collection of existent data pose significant problems linked to many reasons, among those: the lack of databases or of a reference institution; the haphazard contacts among non-governmental organizations (Ngos) or institutions working in the same area; gaps in communication and organization within local institutions; low importance given to data and their validation; competition and data property.

The paper would propose a reflection about these issues based on experiences of PhD students and researchers of the CeTAmb LAB (Research laboratory on appropriate technologies for environmental management in resource-limited Countries), which deals with strategies and solutions for waste and water management in developing countries. Based on our infield experience we are reflecting on our role in overcoming difficulties related to the collection and management of data in academic research in the context of international cooperation.

Difficulties in data collection and sharing

Projects carried out within the CeTAmb LAB are related to environmental management and appropriate technologies for waste and water treatment in developing countries. Depending on the nature of the research, the collection of different data is required. Data could be quantitative (e.g. chemical, physical and microbiological parameters of water or solid waste samples, water or waste flows, temperature, area, population, number of waste bins or latrines, etc.) or qualitative (hygienic behaviors, risk perception, cleanliness of streets, etc.). Data could be collected directly through on-site measures, analysis, observations, interviews, focus groups or questionnaires, or indirectly from publications, reports or grey literature.

New data collection

Many challenges are related to the collection of data in the field. When data are collected by the researcher, a problem we encountered is the language and the direct communication with the data

owner, in particular during surveys or interviews. Even with the support of local partners or translators, many nuances could be lost. As an example, during focus groups conducted with the communities of peri-urban wards in Tanzania, concerning household latrine realization, groups were facilitated by local partners speaking Kiswahili, with the support of a local researcher. Even if main points were summarized and translated into English for the researcher, information were partially lost in particular concerning technical details and it was difficult to answer to some of issues raised by the community.

Socio-cultural aspects, and local customs and traditions have to be taken into consideration, to avoid the risk of collecting incorrect data. There are information that could be considered as sensitive: as an example, in certain cultures to ask about the number of livestock units is equivalent to ask about the bank account in Italy. In others, women are not allowed to express an opinion different from that of men. In order to obtain congruent information on, as an example, water harvesting or collection, it would be necessary to dialogue with women in absence of men. Moreover, it is necessary to consider that it is possible to collect false positive information, as in some cultures the expression of a negative opinion is intended as an offence to the interlocutor.

Another issue emerged in our experiences concerns the quality and accuracy of data collected by others on the behalf of the researcher: even if a protocol is shared, it is difficult to assess the reliability of data when there are inconsistencies. Moreover, the lack of proper tools brings to further difficulties and risks of inaccuracy. For example in some projects in African countries (Burkina Faso, Tanzania and Mozambique) we faced the problem of the unavailability of laboratory sterilized containers for water sampling for microbiological analysis. We decided to use sealed mineral water bottles, to be emptied just before sampling, being careful not to contaminate any part of the bottle during this operation. If the operator does not respect such precautions, the results of microbiological analysis could be misrepresentative. Another difficulty is represented by power cuts, which are common in some areas, affecting the use of electrical equipment for samples storage and analysis (e.g. refrigerator, incubator).

Finally, there are areas which are inaccessible because of conflict or emergency situations where even local institutions and associations have not access. Projects carried out in those areas, as an example in zones controlled by Fuerzas Armadas Revolucionarias de Columbia-Ejército de Pueblo (Farc-Ep) and Ejército de Liberación National (Eln) in Colombia, would result in a gap of environmental data.

Reliability of existing data

Other sources of data are peer-reviewed articles, whilst an important source is the grey literature, which includes documents, reports, handbooks, thesis etc. produced from different actors and not peer-reviewed (Schöpfel and Farace, 2010) and even newspapers' articles. An issue is related to aggregated data: usually published or grey literature contains already elaborated data, and it is difficult if even not possible to access raw data. Information on how data have been elaborated or analyzed is sometime absent as well, making impossible their validation and with the risk of making inaccessible data and details useful for other projects.

Each of these sources has a different level of reliability, which should be assessed by double-checking information when possible, even if it is necessary to keep in mind the difference between levels of reliability.

In some low-income contexts, particularly in rural areas, the oral tradition is still important and recognized at community level even if, for us, it is not 'scientific'. Some information could be only obtained in this way, and could not be verified or assessed. Moreover, in some rural areas even official data from local authorities could be outdated and do not reflect the actual situation, especially when services as Civil or Land offices do not exist. The capacity and ability to cross check information and data in order to assess those that cannot be validated becomes fundamental.

Access to existing data

Another big challenge we encountered in data collection in our experience in low-income contexts, is the accessibility to existent data.

If difficulties in collecting reliable data are a recognized challenge in the sector, to obtain already existing data, previously collected by others, represents an additional difficulty, and often depends on the willingness of local partners (Strande at al. 2014). Based on our experiences in African and Latin American countries, more than once we struggled with the problem of finding or obtaining already existent data, which were collected by other Ngo or local authorities within previous or concurrent projects.

These problems come from different causes or a combination of them. Often data are not well managed and stored by the owner or responsible, and they are not consistent and well organized, but dispersed in different offices, or remain with consultants hired to do some research. Causes are to be

found in the change of reference person, the lack of structured systems for data storage, the lack of communication between different offices and actors of data collection.

As an example, during a project in Tanzania it was necessary to sample and analyze water from wells in the project area. After lot of time we could obtain the list of existing wells from the Municipality, but it was incomplete, due to the fact, as reported by Municipal officers, that usually people do not declare to have or to dig a well. The team therefore decided to search for wells in field, asking to the population and local leaders. It emerged that many of identified existent wells were realized within a previous cooperation project, but data about the location and characteristics of wells were not available at level of local and municipal authorities, while the Ngo did not work anymore in the town.

In another water and sanitation project in Mozambique, after discovering, during a survey in households, that another Ngo already performed a similar investigation some years before, we tried to contact the Ngo in order to have some information about it. Data were not shared because they did not have any more the project file folders in the local office and the reference person did not longer work for that organization, since the conclusion of that project.

In many low-income context, data are still paper-based and not stored in electronic form and this can affect data accessibility. Even if digital archives make easier to disseminate data, they do not solve automatically problems related to this issue, such as the physical location of the storage or the risk of loss. Moreover, in some contexts, hardcopy archives are more usable, as they do not need electricity or electronic equipment to be accessed.

The question of language, mentioned above in the paper, emerges again as an important discriminatory factor also for existing data. In our experience, we rely upon literature in English, French or Spanish. Even if English is recognized as common international language, it is not necessarily the only language in which we could find information in particular in low-income countries: we can think about emerging areas as China, South and Central America, and Arabic speaking countries. Discussing the language of scientific communication, Van Weijen (2012) shows as an indicator the ratio between the number of journal articles published by researchers in English to those in the official language of studied countries. Although only eight countries are shown, it is interesting to point out how this ratio varied from 10:1 in 1996-1999 to 2.5:1 in 2008-2011 in Brazil (the only country with a descending trend) and remained stable around 2:1 in China. As European

countries shown values ranging from 40:1 to 5:1 (reference period 2008-2011), the perception of a European researcher could be influenced by the abundance of paper in English, leading to the neglect of a huge amount of information. Still speaking about scientific communication, a reflection on cultural hegemony lies under the choice of the communication language.

At the local level, most data collected by authorities are available in local languages (e.g. tables with data about waste or wastewater treatment plants) and need to be translated. In Tulkarem (West Bank) a whole database on waste collected by the local waste management company was available, but all supporting information were in Arabic language. In such a case, a tight collaboration between a trusted local officer and the researcher is needed, in order to translate properly technical terms and definitions.

Data ownership

In our experiences we reflected about some risks and competition related to data property. Who has the ownership of data? The collector or the beneficiary? Who can use them and how? Does it exist a competition based on data accessibility, which could permit the owner to access to knowledge, results or funds that could influence their management? Ngos can have a role in promoting the importance of data collection and in enhancing transparency, as in the case of Tulkarem (West Bank). The local authority for solid waste management started to collect data at the beginning of a development project promoted by an Italian Ngo, as it was necessary for monitoring obtained results. In the final phase of the project, the access to the database was possible thanks to the strong relationship between local authorities and the Ngo.

Data from failures

Final results of a project are more likely to be shared if the project was successful, but what does it happen when a project or a single action of a project fails? Often, data collected under project actions that do not meet their target are not published nor shared. To our experience, negative results are scarcely reported in literature. This can be due both to a cultural approach in which errors and failures are not admitted, and to funding mechanisms of development cooperation, which reward successful projects. As a consequence, some of these data are likely not to be shared: managing data requires time and efforts, and it could seem meaningless if they are related to a project which has not reached expected results. Anyway, such raw information, even if partial or

"negative", could be useful for other practitioners or researches, as a starting point or just as a mistake not to be repeated.

Discussion

Our reflection, as young researchers at CeTAmb LAB, leads to recognize that scarce importance is given to data collection, accessibility, management and sharing in many low-income contexts.

One of key points of reflection concerns the role of academics and researchers working within international cooperation projects in increasing the awareness of projects stakeholders about the value of data and consequently data management. At the beginning of each new cooperation project in which we will be involved as researchers, it should be important to promote the creation of a policy of data management and sharing, in which all partners and project beneficiaries have a role.

At a local level, institutional partners' needs to be involved and trained in this process, to make sure that data are understood, safely and systematically collected, stored, shared and updated. Storage methods have to be developed together with local partners and adapted to their skills. Another important action we are already undertaking, is to establish contacts, when possible, with local universities in order to create partnerships and data sharing between universities in Italy and in low-income countries.

As observed, the willingness and low awareness are not the only reasons impeding data acquisition and sharing. There are situations in which is not possible to exactly follow procedures, or respect monitoring plans or guarantee correct sampling and measures, due to economic constraints or unavailability of adequate tools. We would like to stimulate within the academia and practitioners of the sector a reflection on methods and strategies to adopt in these situations. On one side, it would allow to anyway get environmental information; on the other, it would build skills for 'data collectors' in low-resources contexts. Once data are collected, it should be necessary to build and support strategies, tools and methodologies for their management. These should be encouraged and implemented at local level, involving and sensitizing local officers and data owners, but also at research and academic level, developing strategies for data sharing, accessibility, validation, preservation and transparency. Data should be collected and stored in a clear form to be accessible and comprehensible by others, with particular attention to raw data which could be used for other elaborations.

International NGOs and aid agencies are doing great efforts in building tools and platforms for data sharing and to increase transparency.

We can cite as an example the site of World Health Organization (Who, http://www.who.int), containing a huge amount of data and statistics on health from all over the world; the site Aquastat (http://www.fao.org/nr/water/aquastat/main/index.stm), Food and Agriculture Organization's (Fao) global water information system, source on global water statistics, or the Waste Atlas, which represents a crowdsourcing free access map that visualizes municipal solid waste management data across the world (http://www.atlas.d-waste.com/). At academic level we can mention the DaBaCU, an online platform promoted by the Coordination of Academic Cooperation of the Dgcs-Maeci with Italian universities, aimed at creating a community and an opportunity of sharing and exchange knowledge and experiences. Networks are important tools for accessing and sharing information about initiatives already carried out in similar areas or concerning similar themes. Such an approach could support also the acquisition of site-specific data.

While from one side data sharing and open access data should be at the base of advancements in knowledge and development for low income countries, on the other hands some risks could occur. Data can be used improperly, as pointed out also in Michener (2015): an example could be the use of geographical information for both social control or crime in areas with social conflicts (e.g. favelas), as well as the use of geological data for identifying the presence of precious metals, in those cases in which exploitation has a negative impact.

Reflecting on free sharing of data, soon our thoughts turn to the intellectual property concept and on the risks in the dissemination of data. The introduction of a specific formation of PhD students about these issues would be promoted within the CeTAmb LAB.

It would be also interesting to valorize data collected within projects which did not reach their goal, or even the experience with collection procedures showing weakness. In fact, in development cooperation projects Trial & Error is a common process, which demonstrates the capability to recognize errors and modify project choices, but it is often understated. Research on development cooperation, and development cooperation itself, will benefit from a proper confrontation on failures.

The academia is based on principles of capacity building and sharing knowledge and could play an important role in data sharing. Universities should reflect and invest part of their efforts in

promoting the importance of data collection, conservation and sharing, actively developing solutions and tools for it. At an academic level, CeTAmb LAB ongoing reflection is leading to different proposals for stimulating the accessibility and sharing of data and information for researches in international cooperation. Communication channels would be a periodic newsletter, the promotion and support of conferences and training in countries where we are involved, the organization of conferences and meetings to strengthen networks at a local level in Italy, the use of existing tools and databases to disseminate its activities and researches.

Conclusions and way forward

The acquisition, validation, storage and sharing of environmental data in low income countries pose a series of difficulties at research level. Young researchers at CeTAmb LAB started a reflection about this theme based on their experiences in low-income contexts, recognizing the important role the academia could play in highlighting the importance of data correct management, accessibility and sharing, and proposing solutions which could be implemented within CeTAmb LAB projects.

References

Michener W. K. (2015), "Ecological data sharing", Ecol. Inform., 29, pp. 33-44. Available at: https://doi.org/10.1016/j.ecoinf.2015.06.010 (last accessed: 28/04/2017)

Ndzabandzaba C. (2015), "Brief for GSDR 2015, Data sharing for sustainable development in less developed and developing countries", Institute for Water Research, Rhodes University, South Africa. Available at: https://sustainabledevelopment.un.org/index.php?page=view&type=111&nr=6158&menu=35 (last accessed: 28/04/2017)

Schöpfel J. and Farace D. J. (2010), "Grey literature", in (2010) Encyclopedia of Library and Information Sciences, CRC Press, pp. 2029-2039

Strande L., Ronteltap M., Brdjanovic D. (2014), "Faecal sludge management. Systems approach for implementation and operation". IWA Publishing

Van Weijen D. (2012) "The Language of (Future) Scientific Communication. Research Trends" . Available at: https://www.researchtrends.com/issue-31-november-2012/the-language-of-future-scientific-communication/ (last accessed: 28/06/2017)

Acronyms

CeTAmb LAB	Laboratorio di Ricerca sulle Tecnologie Appropriate per la Gestione dell'Ambiente nei Paesi a Risorse Limitate
DaBaCU	DataBase della Cooperazione Universitaria
Dgcs	Direzione Generale per la Cooperazione allo Sviluppo
Eln	Ejército de Liberación National
Fao	Food and Agriculture Organization
Farc-Ep	Fuerzas Armadas Revolucionarias de Columbia - Ejército de Pueblo
Maeci	Ministero degli Affari Esteri e della Cooperazione Internazionale
Ngo	Non-governmental organization
Who	World Health Organization