WASHCost

Data with a purpose

Database development, data management and dissemination in the WASH sector

Expert meeting on WASH Data

19th & 20th of January 2009, The Hague, The Netherlands

IRC International Water and Sanitation Centre

Table of Contents

Executive Summary

In January, the participants of the "Expert Meeting on WASH data" came together at IRC International Water and Sanitation Centre, in their new building in The Hague, The Netherlands. These experts represented organisations with experience in the domain of data collection extended to the use of evidence in policy-making and advocacy, from NGOs such as IRC and Connect International to larger international organisations such as OECD and GLAAS/JMP.

The first day consisted of a brainstorming exercise during which the participants mapped out and discussed the beneficiary to beneficiary (B2B) data cycle. The B2B data cycle extends through six steps and starts and ends with the beneficiary of WASH service delivery.

On the second day, this map was improved through presentations and discussions from each participant on a particular aspect of the cycle. Examples are quality and sustainability of data collection and the use of evidence in policy-making. Each individual contributed insights that have been brought together for the first time in this report.

The experiences shared during the expert meeting revealed amongst others the following:

First, collecting the right data, with an appropriate definition of that data, is the primary challenge. The human side to this is that data collectors and people surveyed on the ground can get bored with 'the next form to fill out' and data entrants can begin to 'play' with spreadsheets for data entry. Participants of the expert meeting agreed that early feedback with useful results to stakeholders is a way to keep momentum and avoid data collection fatigue. Contextual information 'behind' the data is crucial to collect for later analysis. Information such as the reliability of the collected data will set the boundaries for the data manipulation.

Second, a wealth of software, experience and examples are available for database manipulation. The main lesson here is less to do with databases but rather that data which has not been collected with the appropriate level of 'granularity' will result in less useful conclusions. In respect to GLAAS, for example, disaggregating aid for water and sanitation is cumbersome because donors had not previously seen the need to share or collect that level of disaggregated data.

Regarding, the use of information, the meeting pointed strongly towards 'getting to the client'. The interface to the collected data, modelling and background information has to appeal to the user and help them to wade through the maze of information and choices. The user interface may well be more important than any specific decision support tool.

Some highlights from participants:

- Connect International pays a lot of attention to and has a lot of experience in getting measurement incorporated into projects on the ground.
- OECD has made an overarching (second) interface to guide users through the data collection and make the right selection / sorting / export. They work with 'persona' (segmented 'typical users') to assist interface designers in recognizing users' information needs.
- UN Water / WHO are requesting an expanded set of cost and expense data from donors and recipient countries to support a global analysis of needs and trends in sector spending (GLAAS)." UN-WATER focuses on gathering UN data and information resulting in a number of reports. They start also from a user perspective.
- TR61, as a national sectoral unit cost database, has three decades of experience, with their service being paid for by the beneficiaries e.g. water companies.

The overriding conclusion is that the users (and, less clearly, the ultimate beneficiaries) of the system are key to the success or failure of the data cycle. There are stakeholders, users and beneficiaries at nearly every step of the cycle and as such, these are essential to the successful creation of any useful interface and decision support tool (DST).

The development of a DST requires careful consultation and participation of all relevant stakeholders. In the case of WASHCost, the key decision-makers are policymakers and budget-providers. Through Learning Alliances, WASHCost is attempting to ensure that these decision-makers, who should ultimately benefit from a DST, are involved in the development of the tool. At the same time, the people who answer the questionnaires and those who collect data, must also be kept in the loop in order to avoid data collection "fatigue", must ensure the data is useful (and likely to lead to action also for the level that is collected) and must ensure that the data collected is of appropriate quality to be useful.

1. Introduction: 'Setting the context'

The WASHCost Project researches the life-cycle costs of water, sanitation and hygiene (WASH) services in rural and peri-urban areas in four countries. The rationale is that WASH governance will improve at all levels, as decision-makers and stakeholders understand the costs of sustainable, equitable and efficient services and put their knowledge to use.

The WASHCost action research project started on 1 February 2008 and will finish in 2012. It will collect and collate information relating to the real disaggregated costs in the life-cycle (sustainable service cycle) of water, sanitation and hygiene service delivery to poor people in rural and peri-urban areas.

Within the WASHCost project the IRC International Water and Sanitation Centre (IRC) is in the process of developing a global decision support tool. In order to complement, and not duplicate, existing initiatives that collect data in the WASH sector and use web - based interfaces, the WASHCost project organised an Expert Meeting on WASH Data to learn from, and learn with, other people and organizations with similar interests.

The main objectives of the Expert Meeting were to:

- Map and discuss what tools are being used for data collection and management in the WASH sector, with a special focus on quality and sustainability
- Map and discuss the relevant use of (web) interfaces for decision support systems
- Introduce WASHCost and collect ideas for WASHCost to organize the development of its own data management processes

In addition, the meeting focussed on the needs of each organization (from a technical perspective), their experiences in working with different tools and on mapping what is attractive in terms of collaboration and joint development. However no commitments were expected from any of the participants. The story of the 'Blind men and an elephant' is an apt metaphor for the rationale of this meeting (see box 1).

Box 1. 'Blind men and an elephant'

In this ancient Indian story, a group of blind men touch an elephant to learn what it is like. Each one touches a different part of the elephant. When they compare notes on what they have felt, they learn they are in complete disagreement. In the story the blind man who feels a leg says the elephant is like a pillar, the one who feels the tail says the elephant is like a rope and so on. This story indicates that reality may be viewed differently depending upon one's perspective. The WASHCost project organized the Expert meeting on WASH Data in order to try to get the broadest possible picture of data base development and data management and sit together to paint one story.

The main objective of the first day was to map out the main issues with regard to database development and data management. The participants of the expert meeting represented various types of organisations and projects in the WASH sector. In order to share experiences the different areas of expertise were first mapped out in the Beneficiary 2 Beneficiary (B2B) mapping exercise. The B2B exercise visualises six main steps in the data cycle and starts and ends with the beneficiary of WASH service delivery. The full exercise is included as appendix 1.

These data cycle six steps are:

1. Beneficiary:
2. Data collection:
3. Input interface(s):
4. Database(s):
5. Output interface(s):
6. Beneficiary:

The main issues that were raised during this brainstorm session were used as input for the second and final day. To be able to go more into depth on the final day participants presented case studies they had been involved in around three topics:

i. Data collection
ii. Data management and
iii. Systems and users.

Discussions were then held around the framework of these three topics and the real life examples. This document brings together the shared experience and knowledge of the participants on the different stages in the development of decision support tools structured around the three main topics of the second day.

2. Data Collection: sustainability and quality control

The main questions with regard to data collection are how to collect data of the desired quality (when is it enough?) at an appropriate level of disaggregation or 'granularity' (how much detail is needed/useful?) and secondly how to ensure sustainability of the system of data collection, analysis and loops of learning. This section will discuss the shared knowledge and experience on these two topics.

2.1 Sustainability

Quick wins. The main issue to take into account when trying to ensure sustainability of data collection and its use, is to ensure that there is a direct benefit back to the institution which organises and/or supplies the data.

Strategic intend. The most suitable approach when it comes to data collection is to build on what already exists and try to fill in the gaps. The focus should not be on getting all participants to supply information on all indicators, but only on the indicators that they have information on or feel comfortable with. In the long term this will increase the quality and quantity of the data.

Champion / lesson learned. For example in IBNET there are 32 indicators, as all participants could not provide information on all indicators different levels were used. Over the years participants saw the benefit of supplying cost information (which they could not do initially) as they could use the overview from the database provided for their own business planning. As a result they started to supply more information on indicators they were previously not reporting on.

Buy inn. One of the first actions with regard to data collection is to determine who the keeper is of the data you are looking for. The process of data gathering can be long and difficult and is slowed down as governments or data owners need to sign off or validate the data. The most efficient way is to build on already existing data and focus the efforts on filling in the gaps. This way of working relieves the already existing research burden on the involved communities and avoids duplication of efforts.

Buy inn / quick wins. As in many countries in the South, communities and institutions have provided information in many questionnaires and surveys from which they have never seen any outputs. For example respondents in a research in Ghana told researchers; 'There are no consequences for being truthful in the questionnaire because nothing will be done with it anyway.' It is important in this regard therefore to avoid data collection fatigue which results from numerous repeated questionnaires without obvious benefit or feedback to those involved in the process.

Lesson learned. A success story that demonstrates how effective the provision of cost information can be if research outputs are returned back is a database[1] that was set up in Ethiopia that provided an overview of the costs of chlorine in different parts of Ethiopia. It turned out that there were high differences in chlorine prices in Ethiopia as in some parts of the country suppliers had a monopoly. The transparent overview increased the awareness of customers on the price differences and customers started to demand lower prices from suppliers as they realized they had been paying too much. Local networks were then set up to keep the exchange of cost information going. The WASHCost project also wants to do more than just supply cost information. The information, as in the Ethiopian example, is also to be used for advocacy purposes.

Member buy inn. One possible way to increase the sustainability of a database is through setting up a membership scheme in which members pay to access the data that is collected. In Indonesia for example municipalities pay for water utilities to be part of a benchmarking network. A business model that has been very successful, especially in terms of sustainability, is TR61 (see box below). TR61 started in 1977 and provides cost Information for Water Supply and Sewage Disposal. It now also incorporates context information and can be used for:

- Capital investment planning at company or divisional level
- Estimating outline capital costs of strategic policies to assist strategic investment decisions
- Asset valuation for current cost accounting procedures

Case study 1. TR6

TR61 is the United Kingdom water utilities joint costs database. The business model of TR61 is based around a membership structure where members, predominantly UK water companies, sign up for a period of three years at a time to be part of the network. This sytem has now been self-sustaining for over 30 years. Members supply information on an annual basis with the main source of information being data from capital investment and service contracts. In 2005, for example 2.200 contracts, of all sizes, were processed and anonymised data included in the dataset at an aggregated levels so that the source could not be determined. Anyone can be a member, for example a national government joined for a period to understand better the costs being claimed by their contractors. TR61 is now also being used by contractors to the water companies. Individual members pay 30.000 pounds/year to be a member and in return for supplying information and paying the fee, members can access the data.

The level of the membership fee is determined by the costs of managing and maintaining the database. The main reasons for the success of TR61 are:

- The information is treated on a confidential basis and companies submit the information on the basis of anonymity because the information is commercially sensitive
- Data is owned by the people who contribute to it
- The value of the data derives from its aggregation and comparability

[1] The reference to the prices of chlorine mentioned in the report only emerged from the running of a benchmarking and business planning workshop - and people talking and sharing information, rather than as a direct result of the collection of cost data.

- Use of appropriate software to build up the database, including a 'sense checker' in the software
- The use of a contributor steering group for management has created a dynamic user group process
- It is independent and externally audited.

Especially the fact that TR61 is independent and externally audited is important because if the information was, this would have serious implications. The auditing does not include examining individual data entries of each member. Auditing within TR61 consists of:

- Examining the data gathering process (but not the data from company contributors)
- Testing the functioning of the software

TR61 is being managed by a steering group of 19 participants who come together every six months to discuss the data. The total costs of keeping TR61 running are approximately one quarter of a million pounds (Euros) per year.

The breakdown of the main costs are:

- 2/3 on staff
- With software development taking up 50% of the time
- Engineering activities

The staff needed to maintain TR61:

- 1 full time senior person (though this is not seen as a particularly prestigious appointment)
- 4/5 staff for intensive periods of time (engineers / economists / software engineers)

Main fact database:

- Database is password protected
- Data entry forms are online
- System runs in / for the UK where Internet access is relatively easy

2.2 Quality control

One of the ways to enhance the quality of the data that is collected and to increase the response is to conduct data collection in a two-way traffic manner. If research outputs are returned back to the people who have provided the data, the quality of the data improves as there is a clear benefit for the beneficiary to participate in the data collection.

For example, Connect International uses a standard questionnaire with a 150 questions that covers 35 different topics. The questionnaire can be adapted to different demands and contexts needs to suit their clients or partner organisation. The information that is generated in the survey can be used as an input for the village development plans (see Figure 1). This means there is a direct benefit to the community for participating in the research and this increases participation and quality of the data. Local use of the data is therefore one way to increase the quality of data collection.

The example below illustrates the use of conditional formatting in the spreadsheet. On- or off track with planning is shown as green (on track or better), yellow (danger zone) and red (off track).

Figure 1. Example of information sheet produced by Connect International

	Refuse pits				Improved pit latrines			
Community	No. realized during quarter	Total no. in use	No. of households using the product	Plan next quarter	No. realized during quarter	Total no. in use	No. of households using the product	Plan next quarter
Sitouge (140 hhs)	54	124	124	21	23	38	37	35
Cestino (100 hhs)	22	31	33	0	17	71	65	0
Konkonte (109 hhs)	15	24	26	24	23	14	16	12

Doing market research to get an overview of which data is demanded can increase the quality of the data that is collected. Who your target audience is will drive your output. As the audience for data collection within the WASHCost project is very broad, this poses a challenge in defining and limiting data selection.

Benchmarking is important to define the data that might be collected. As an example, for WASHCost the structure of Basic Urban Services (BUS) was mentioned. The audience for TR61 was defined as those who want (comparable) cost information and user communities were asked how they wanted this information presented.

It is also is vital to build in quality checks in the database to cross-check the data that has been entered correctly. Software programs can make links between data and cross-check the data for accuracy along a formula of *'if data A equals X, data B cannot be Y'*. A system where the data is not only checked by the use of software but also by the human eye is most effective. For example in one programme in India data is entered twice and then cross-checked for differences by the use of software. Differences in entry are flagged by the software and a validation process is started by the human eye. The average error rate in data entry ranges between 2 – 5%. For example IBNet uses one person to undertake quality assurance in comparison to the OECD who use eight people for their, much larger, OECD.Stat database.

Some of the main lessons of quality control in data collection revolve around involving users in data collection as this increases the quality of the data. The more involved people are the more interested the get in using the data. Conducting user surveys every 3 to 4 months to get feedback from users is considered a good way to increase quality and sustainability (OECD and UN-Water are currently not conducting any surveys). The use of a data grading system was also mentioned as a tool to increase quality control of which IBNet is a good example (see case study 2).

Case study 2. IBNet

The initial concept of the International Benchmarking Network (IBNet) has been to set up a simple set of basic indicators to facilitate compilation and inter-utility sharing. The process of setting up IBNet took one and a half years to get agreement on appropriate benchmarks before delivering a functioning tool. IBNet has now been taken over by the Water and Sanitation Program (WSP) of the World Bank (WB).

The main characteristics of IBNet are:

- Use of "Off the shelf system"
- Benchmarking Start Up Kit, database and web shell for presentation on WB site
- The use of market research (focus groups etc) to determine what users wanted
- Wide ranging indicator categories – service, operational, billing, financial and processes
- Recognition of the limits of participation in establishing indicators: 'at some stage you have to make a decision'
- Emphasis on local "ownership" and encouraging start up of benchmarking as a *good management tool*
- To enhance and provide a complementary facility with other international and regional initiatives
- Data collection and sharing is ongoing
- Value lies in time series not in one time collection
- Part of the contract is to have a helpdesk to provide assistance and a newsletter
- Different grades of membership
- Data is available for free

One of the main ways to increase quality control in IBNet has been to introduce a quality grading system. This helps an external person to assess the quality and reliability of the data. This system consist of:

- A = Audited data
- B = Reliable data
- C = Uncertain data
- D = Estimate, no data
- or "X" to indicate data important and action needed to collect better information

With specific definitions of each of these terms.

When data is entered, individual data receives a tag (A to D) so users can easily assess the reliability of the data. Data is being checked at node level by the node manager. The node manager forms a critical person in the insurances of data quality. This person needs to be enthusiastic and computer literate and can be a critical person from water association/water utilities. Each node level can develop in their own way. The IBNet data transfer process can be seen in the figure below.

In total IBNet has:

- 72 data items
- 30 indicators
- 2 full time staff from WB

There is no requirement to supply each of the 72 data items. The concentration is on collection of good and valuable data, not collecting all the data. There is very detailed guidance as to what is meant by each data point: 'precise and agreed'. The focus is also long term with the general belief being that if members see the benefits of using the data they will voluntarily start supplying more and better quality data items.

The number of utilities involved in IBNet has risen from approximately 200 to over 1,900.

Lessons learned for sustainability from IBNet:

- The danger of being overly controlled from the centre, relative to providing a service back to imputers
- Colleague-partnership relations are important for peer-2-peer learning
- The importance of a country 'node manager' to take national responsibility
- To help kick-start such a process it is helpful to run a collaborative benchmarking workshop, bringing water associations and other interested groups together
- There is a need for basic seed core funding from a donor. However funding should not cover all expenses so there remains a need to be innovative and dynamic.
- The initial funder should not (be allowed to) take control of the product or the agenda
- Allowing for initial anonymity of data viewed, whilst users gained confidence
- IBNet could have been sustainable commercially if a suitable product had been developed on top of the core.
- Difference between the extent to which the donor might see the data as an end in itself as opposed to how managers can use it as a vehicle to their processes forward

2.3 Lessons learned for WASHCost

In order to increase sustainability and quality control of data collection the main critical issues are:

- Need to have a benefit back to the one who supplies (organises the supply of) the data
- Need to be a service.
- Anonymity encourages people to provide data and provide accurate data → voluntary & anonymity
- Value of database evolves over time
- Provide choices/flexibility so people feel comfortable to participate and hopefully over time (when they feel comfortable) they will provide more data
- Data collection can be on paper, if filled in by communities/ enumerator, or on software if there is the capacity and infrastructure. One of the advantages of paper questionnaires is that it does not require energy to function and it is not a complex system. However, as for using computers, literacy is required.
- Data entry to be duplicated by different people and then cross-checked
- Need for a human component in the system (data auditor or node manager) next to software
- Main issues are in data collection not in data base (this can be developed to suit needs)
- Return research outputs
- User surveys
- Don't 'straight-jacket' the data
- Involve users in data collection
- Data grading system
- In country data auditor or aggregator (ex: node manager in IBNET)
- The WASHCost project therefore uses a 'Learning Alliance' approach. WASHCost will also have the research methodology available after the second year of the project so partner countries or others interested can use it to collect their own cost information.

3. Data management: technological choices and standards

The creation of "the database" is a very sensitive exercise. Extensive training or capacity building is essential to ensure to quality of the database. One of the issues that affect the functioning of databases is the definition of each item of data. In order to have equality in the system it is very important to define the definition of each data very precisely.

Altering data that is not included in the database is not a good idea. Always keep the raw data in the database and never alter the files in the database permanently but make modifications only on one a copy of the file. Backups are essential.

In data management, the relatively simple spreadsheet Excel was the most used system. In most cases the aggregation of data comes from different sources with the majority of the databases using SQL or Excel. Excel is especially good for entering data at the level of the lowest common denominator. However the formulae in excel sheets can be very sensitive and problems can arise in uploading data as Excel can change things. The technological choices FAO is using are proprietary software and open source: they use Microsoft server and Oracle. However they do not recommended Oracle as it is very expensive to use. Open source is possible or 'off the shelf' software.

Technological advice from EU Joint Research Institute:

- Use systems developed with commercial software because then you have commercial support
- Data to be produced in standard format in order to be able to reprocess the data
- Use open source only for external use

Case study 3. GIMS (from the WHO Health Observatory)

The Global Information Management System (GIMS) is a comprehensive data management system for data collection and its functionality includes GIS mapping, data analysis and data dissemination. It is a dynamic and web-enabled (Web2.0) system.

GIMS follows the following principles and technological choices:

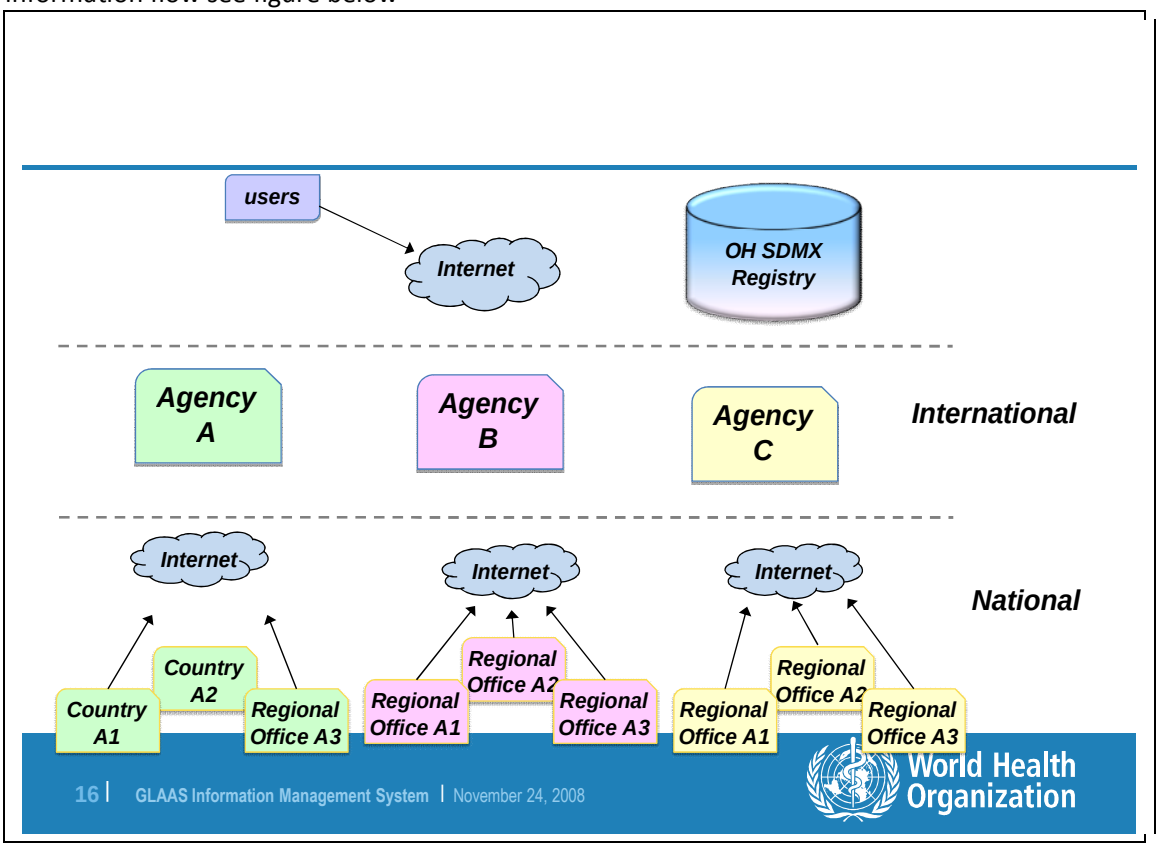
- Data is owned by the country
- Data is directly fed into the system (web-enabled) at country level
- Data is fed into the system through form feed (excel to XML)
- Data collection is participatory
- Country representatives feed data (designated officials with access)
- Online consultation (wiki environment)
- Dynamic data transfer using SDMX (Statistical Data Meta Exchange)
- Data collection through GIS mapping
- All data is freely available

GIMS is a real-time data analysis tool. The data is ready to be analyzed once fed and validated and a

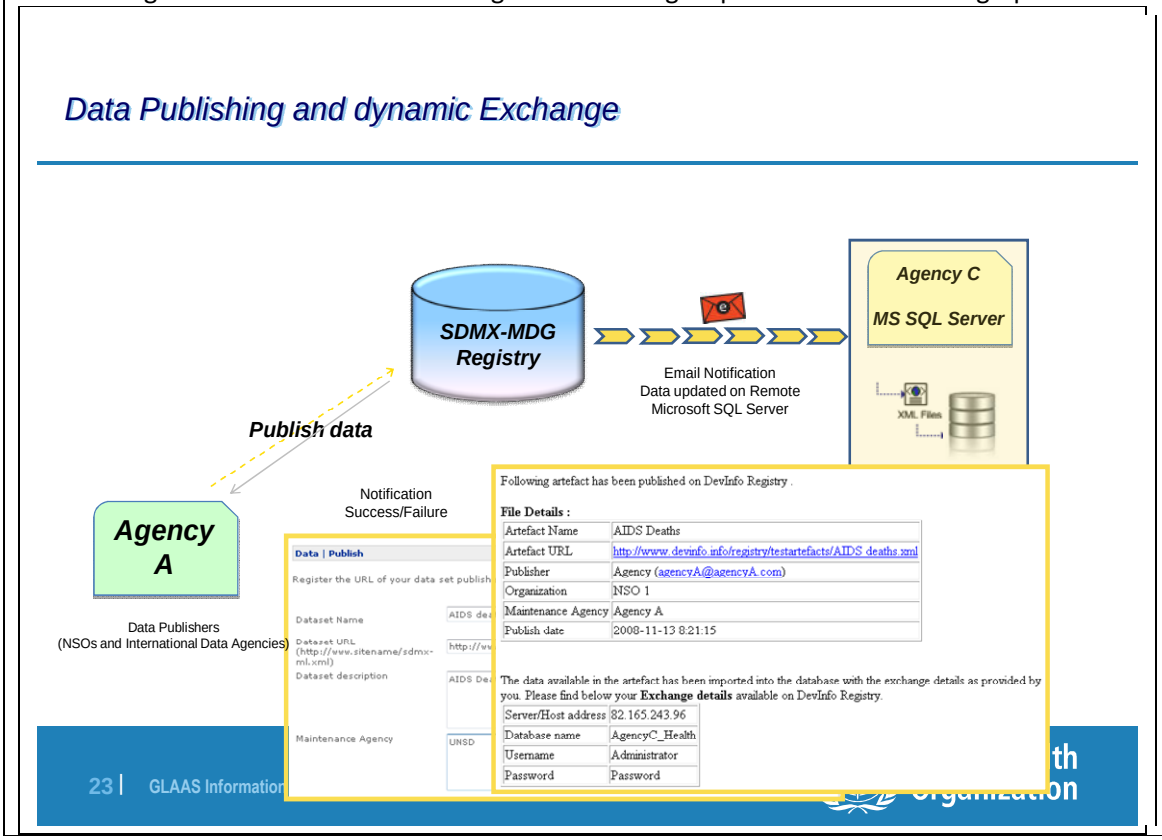
customizable data analysis tool with multi-layered data. It has predefined and customized country profiles and more in-depth analyses by users is possible. The current (Jan 2009) time line is:

- Data dissemination system by March 2009
- Integrated into GHO after March 2009
- Full system operational: end 2009
- 2010 report is based on pilot GHO

The information flow see figure below



Data management in GIMS is done through the following steps as visualised in the graphic below:



Case study 4. GLAAS

GLAAS is a UN-Water initiative led by the World Health Organization. UN-Water GLAAS is seeking a new approach to report progress in the sanitation, hygiene and drinking-water sectors in order to strengthen evidence-based policy making towards and beyond the MDGs. GLAAS seeks to integrate disparate data of the WASH sector to present a global status of progress, trends, and needs. These integrated data are currently available as a hard copy report, but will in future years be available via the Internet. GLAAS aims to be a repository of evidence to make better informed decisions. It has to be:

- Reliable
- Easily accessible
- Comprehensive
- Global
- Periodically updated

GLAAS makes use of consultants who are pre-filling the surveys and then visit the countries in order to verify the data and fill in the gaps. For the future, the idea is to have countries themselves fill in the data and start a verification process. The data is owned by countries. Data collection is through:

- Paper collection surveys
- Excel

GLAAS input are:
- ACCESS: Sanitation and drinking-water
- FUNDS: National and local government, loans, grants, commitments, private sector, households
- CAPACITY: Governance, policies, regulations, human resources

GLAAS outputs are:
- GLAAS report
- Library of country and donor profiles
- Strengthening monitoring systems capacity

3.1 Lessons learned for WASHCost

- You need to define very precisely what data you want to collect
- The core of the data needs to be good with precise data definitions
- It does not matter if you collect data on paper or otherwise
- Open source software benefit is that it is continuously being updated – though technical support might not be available on demand
- Excel is the most used technological choice for data collection
- Web enabled forms can ensure consistency of data entry – individual's knowledge of Excel means they can start to 'play' with it and make alterations to suit their particular requirements or interests
- A data auditor is needed to validate the data.
- UN Water is looking towards a 'federated database'
- WASHCost could involve and make links to Central Bureaus of Statistics in countries in order to have access to country statistical expertise for enumerators and analyzing the data.

4. System & users: evidence -based policy making, advocacy and information for researchers

Data is rarely used or disseminated in the manner in which it has been collected. It is important to make the data accessible and useful. Collecting the information is useless if nobody reads or is unable to read the information. Therefore the use of a multi-tiered approach, depending on the audience, in which you work from use back to collection, is most suitable to guarantee the success and readability of the output interface.

It is similarly necessary not to 'straitjacket' the data through analysing and presenting it ('chopping it') as you think your users want it. It is preferable to allow the possibility to see the data that is behind the analysis. This includes all the raw data – though opinions may vary as to what level of raw data to allow access to. A good example is the Query Wizard for International Development Statistics (QWIDS) from the OECD (case study 5).

One of the main issues with regard to input interfaces is to make use of existing tools on a local level. A mix of technologies and platforms can be used to release the possible tension between what is locally appropriate and what is internationally appropriate.

Other issues to take into account when designing an input interface are:

- One size does not fit all
- Use of technology profile
- Segment the market for the product
- Make use of tools which are available as excel
- First define what the different levels/positions are of the different users and shape/model data access for each of the levels (example of OECD personas)

OECD.Stat is a good example of an output interface that was not fulfilling its full potential. Only a small group of users was able to use the interface to get an overview of the data they wanted out of the database. In order to improve dissemination, the OECD decided to build a new interface that sits on top of OECD.Stat. The new interface - the Query Wizzard for International Development Statistics (QWIDS) - was developed backwards with the use of persona's. The main idea was to start with the user and with the most common requests. The development of persona's first, has helped the developer of the database to see the rationale for the data use. It also has reduced costs as it has helped to get the team to get a clear focus and it has increased efficiency as the objective has been clear from the start.

Case study 5. OECD.Stat

OECD.Stat is the OECD's statistical data warehouse. It provides a single access point for validated statistical data and metadata across the organisation. The DAC publishes statistics and reports (www.oecd.org/dac/stats/dac) on official and private flows to developing countries, based principally on reporting by DAC Members, international organisations and other countries. Membership:

- Free: Creditor Reporting System (CRS) and Development Assistance Committee (DAC)
- Pay: Universities, libraries, research institutes

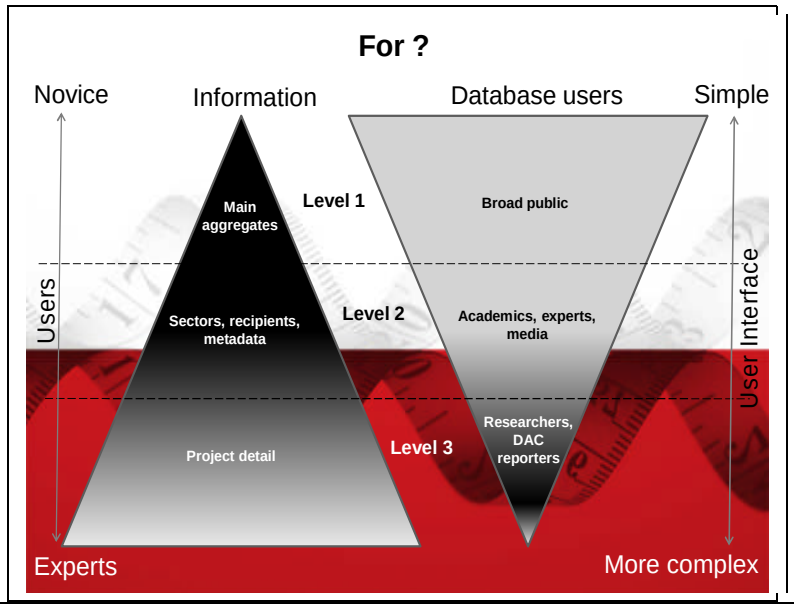
The main problem was that this database was not easily accessible for users and therefore the OECD decided to build an interface (QWIDS) on top of it. QWIDS was developed by external consultants over a period of nine months with a fixed budget, requiring 15 iterations during the development process. This involved seven developers, three OECD staff and two graphic designers. However the main repository has stayed the same (OECD.Stat) and information has to be entered in OECD.Stat before QWIDs can retrieve and disseminate it. Main facts OECD.Stat/ QWIDS:

- Data collection forms → Excel
- Database → SQL
- Input interface →text excel
- QWIDS (output interface) → C-sharp using JSP/ Java based
- Developed in nine month by 7 developers
- Two people on user friendliness/ design

Data collection is undertaken by sending out forms in Excel to member states with a request to complete and return via e-mail. The main limitation to the system is that it is a coded system which does not allow for multiple coded systems.

QWIDS was launched in 2008. The general feedback has been positive. Main functions QWIDS:

- You can save the query
- Send query as a bookmark
- Source of data is displayed at the bottom of each result-set



Case study 6. Connect International (CI)

Connect International is a Dutch NGO which works with approximately 200 communities, 650.000 population, developing Community Development Plans in countries such as Tanzania, Zambia and Mozambique. Payment for reports are:

- DGIS/ Ministries of Foreign Affairs
- Monitoring is done through program funding

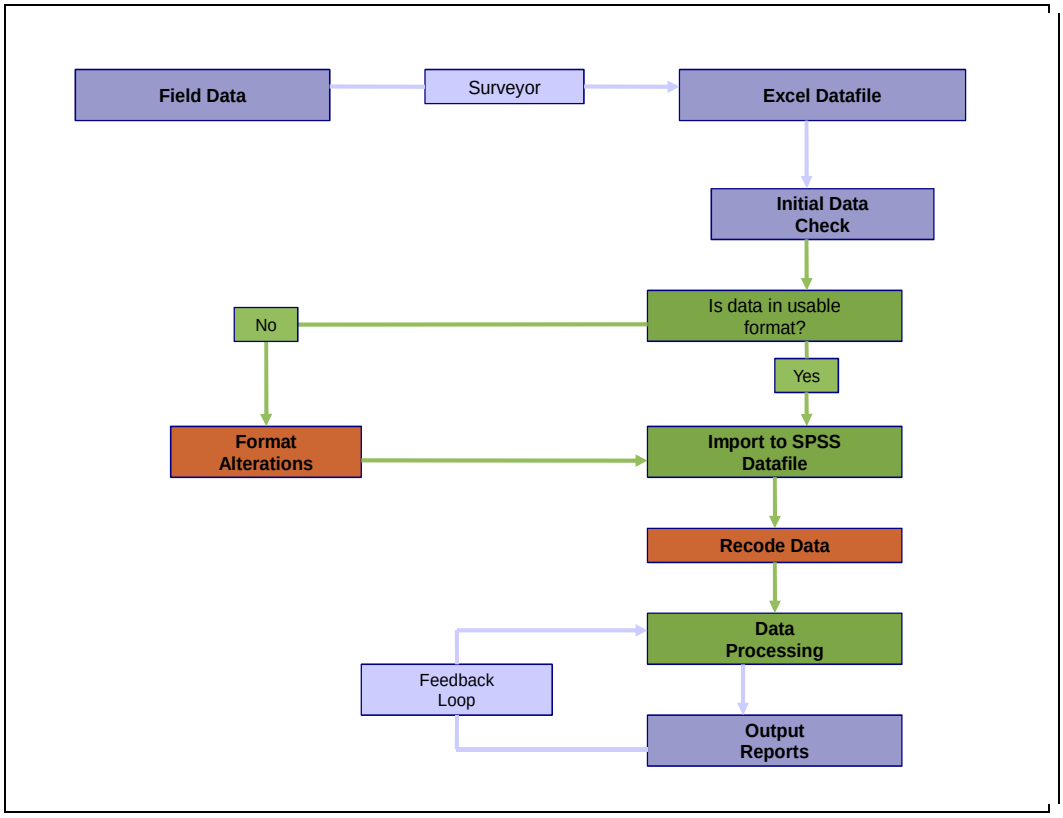
Main lessons learned for sustainability:

- Develop reports that people need
- Develop different types of reports for different types of users

Technological choices CI:

- Work in excel
- Program in SPSS
- Use only excel and SPSS else you end up with too many systems
- Not web based, send excel sheet via e-mail and then enter data into database in the Netherlands

Use a grading system with two indicators:
0 → not good
1 → good



4.3 Lessons learned for WASHCost

From the case studies the following lessons can be taken:

- Work backwards, start with: “what do you want to get out of it” so you “know what you need to put in” (= collect)
- Defining user needs (ex. persona’s OECD)
- Use a fill in the gaps approach as there is so much information already available in reports and you diminish the workload of participants
- Produce quick wins → produce quick feedback on the data collected example the overview of data collected produced by Connect International
- Most users are looking for main aggregates
- Take popular queries as a start
- Make sure the interface is interesting before you try to get people to use it
- Make it simple (is the hardest thing to do)
- Need to bring in a interface designer from the start. A good interface is worth the investment
- It is pointless to have a lot of information if nobody can access it so you need to focus on user friendliness
- Segmentation of information for different personas/users
- Possibility to zoom into interface (get the raw data if really wanted)
- Have defaults in the interface
- One source with multiple outputs

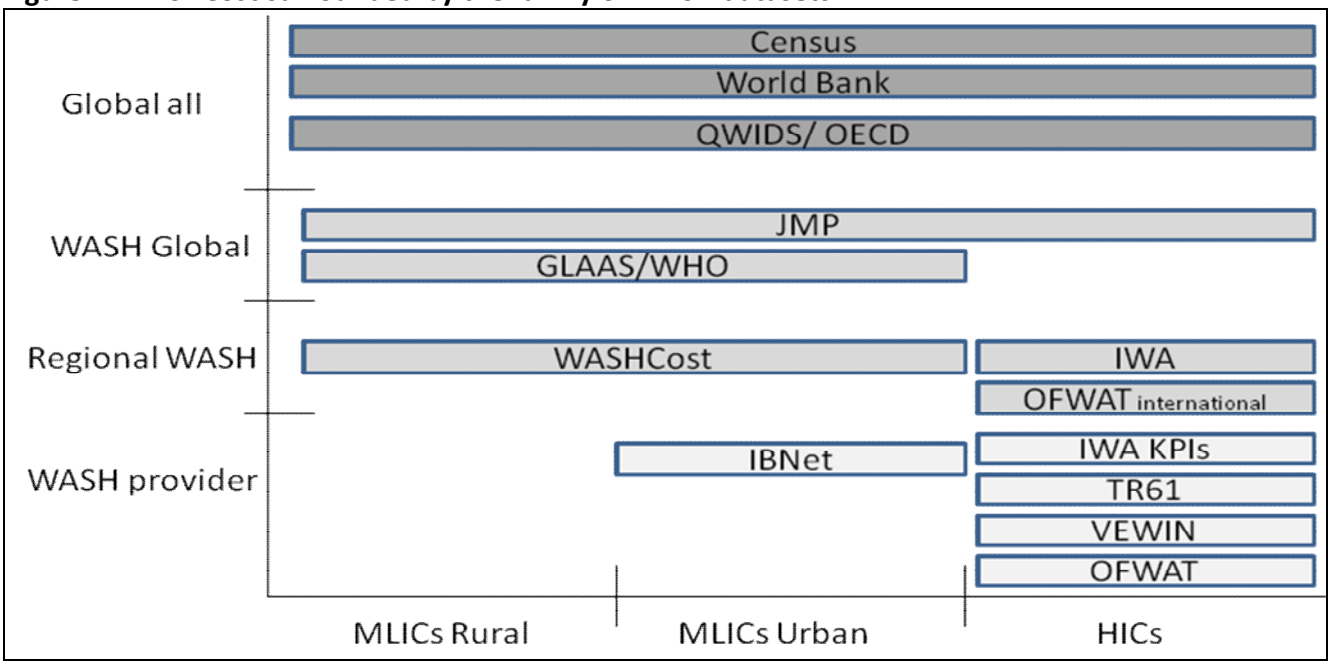
5. Summing up

There are clear lessons for the WASHCost project arising from this expert meeting and the organisers are grateful to the contributors for sharing their expert knowledge.

Many of the individual lessons for WASHCost have been noted at the end of each section. However, an over-arching lesson is that there is already a significant 'family' of (unit costs) databases in the water sector. Members of this 'family' have relations in other sectors and in the global family of development or economic datasets. Other members of the family are particularly strong in certain countries or regions or subsets of the sector. WASHCost, as the 'new addition to the family', has a clear role to play but has to ensure, as it 'grows up', that it is both paying its way and also not clashing with other members of the family if it is to be sustainable and deliver benefits to users over the long-term.

In another analogy, GLASS describes their role as 'trying to put the jigsaw pieces together'. It might well be that WASHCost 'light', which is the ongoing version of WASHCost (following the initial research and validation phase), may be one of the pieces of the GLAAS jigsaw.

Figure 2. WASHCost surrounded by the family of WASH datasets



Index figure 2:

GLAAS	Global Annual Assessment of Sanitation and Drinking-Water
HIC	High Income Countries
IBNet	International benchmarking network
IWA	International Water Association
JMP	Joint monitoring programme
KPI	Key performance indicator

MLIC	Middle and Low Income Countries
OFWAT	Water Services Regulation Authority
QWIDS	Query Wizard for International Development Statistics
TR 61	Wrc software
VEWIN	Association Dutch water companies (Vereniging van waterbedrijven in Nederland)

The figure above, illustrates our present understanding of the position of WASHCost in the family of datasets. This picture recognises that water and sanitation information is one aspect of the much larger economic development world, but also that all sectors can be analysed at different levels, service provider, national, regional and global whilst also being segmentable between high, middle and low-income countries and between rural and urban. While the World Bank, Water Sanitation Program-Africa, did not participate in the meeting, their Country Status Overviews (CSOs) project and data should be mentioned somewhere in the WASH datasets figure on this page.

WASHCost focuses on the rural and peri-urban community and service provider level in middle and low-income countries, complementing IBNet's role in understanding costs in urban utilities in those countries. This information is agglomerated at regional and global level for MLICs through JMP and GLAAS. There are other complementary sectoral databases for similar services in high-income countries and regions (TR61, VEWIN, Ofwat June returns for service providers, IWA Tariff Survey, Ofwat International comparisons etc. at regional/global level). General economic development indicators such as the World Bank's Development Indicators, primarily aimed at MLICs and OECD's (with its QWIDs interface for the DAC information) statistical data warehouse compile and make available overall information.

This figure also recognises the position of a national census. As ever in the development world, there appears to be a temptation for outsiders to generate yet more 'special units', in this case for data collection and analysis, to try and accelerate economic development. These often come at the expense of long-term sustainability of such activities. We note the existence of questions regarding whether 'the household has a bath/shower and toilet etc.' in the UK's ongoing census. We recognise that there are no unit costs being collected in such questions but it is a reminder that the more we can embed our work in ongoing activities, ensuring that 'our new child' is a fully recognised and valued member of the family, then the more effective will be the benefits for all.

Werner and Bower's classic 'Helping Health Workers learn' (1982) includes a drawing of a community member holding up a sign which reads '*Surveys tell outsiders what insiders already know'.* In the context of this report's title, 'Data with a purpose', we seek to be very clear that our purpose is to survey that data which will be directly used to benefit the insiders – and which will therefore be worth collecting long into the future.

6. List of participants and biographies

- Catarina Fonseca (IRC)
- Cesar Carmona - Moreno (EU Joint Research Institute)
- Jaap Pels (IRC)
- Jeske Verhoeven (IRC)
- Mark Baldock (OECD)
- Mark Hoeke (GLASS/JMP)
- Nick Dickinson (IRC)
- Nicoletta Forlano (UN Water)
- Richard Franceys (Cranfield University/ WASHCost advisory group)
- Simon Gordon Walker (Artesia Consulting Ltd)
- Tom de Veer and Tom Heijnen (Connect International)

Catarina Fonseca, is the WASHCost Project Director. As a senior economist at IRC, she was until recently responsible for IRC's focus area Financing and Cost Recovery coordinating the theme supervising several activities such as action research, the production of publications on key issues and the development and facilitation of training courses on the theme together with partners in the South. Before IRC she worked three years with a Portuguese NGO on participatory approaches with a gender and equity emphasis. With IRC, she has nine years of experience with community based management of water supply and sanitation services and specifically innovative finance, cost recovery mechanisms, tariff setting, microfinance and unit costs. Field experience within IRC includes assignments mainly in Africa (Mozambique, Burkina Faso, Cape Verde, Ethiopia) and consultancy work for several Foundations, Development Banks, International Organisations and Governments. Mrs. Fonseca holds a MAs in Development Studies specialized in agriculture and rural development (The Hague, Netherlands) and was originally trained as an Economist (MAs, Lisbon, Portugal). She is currently working on her PhD (Cranfield, UK).

Cesar Carmona - Moreno, EU Joint Research Institute, n/a

Jaap Pels studied Molecular sciences at the Wageningen University in the Netherlands. For twenty years he worked at the Consumentenbond (Dutch consumer association) as (senior) researcher, research department section head and information manager. He developed and honed thinking and planning from a user perspective. At IRC he is responsible for the Knowledge Management programme. Within WASHCost his focus is knowledge sharing, information management, (use of) ICT and communication.

Jeske Verhoeven has studied Journalism and International Development Studies at Utrecht University, the Netherlands. She just finished the latter study and wrote her thesis on Delivering Services: 'Accountability as a vehicle towards access to safe water'. From February 2008 until July 2008 she has done an internship for SNV in Tanzania on Baseline research on Water, Sanitation and Hygiene (WASH) and has also worked as a journalist in South Africa. At IRC Jeske is the WASHCost project assistant. She has written most of this report with inputs from the participants and the WASHCost colleagues Nick, Jaap, Catarina and Richard.

Mark Baldock studied computer science at the American University in Paris before joining OECD's Environment Directorate as IT manager. He currently holds the position of Statistical Data Manager in the Development Cooperation Directorate (DCD) at OECD. He now administers and develops the DAC and CRS databases housed on SQL servers and is responsible for online data dissemination. He recently participated in the development of QWIDS.

Mark Hoeke has a background in civil engineering and information systems, and has worked previously as an environmental consultant for the U.S. Environmental Protection Agency, and as regulatory affairs director for a large utility organization based in the United States. Mr. Hoeke currently provides technical consulting services to organizations in the areas of utility financing, utility management, technical surveys and other data collection activities, data analysis, policy development, and regulatory compliance. Mr. Hoeke provided technical data management and analysis support to the World Health Organization in the development of the 2008 pilot Global Annual Assessment of Sanitation and Drinking-Water report.

Nick Dickinson has a background in information and communication technology, capacity building and the WASH sector At IRC, Nick is channeling those into the innovative communications theme at IRC. In addition, he is currently involved in the WASHCost, RiPPLE and AKVO projects in different capacities from communication strategy development to working on the Akvopedia wiki for water and sanitation. Before starting at IRC, he worked as a consultant on various projects from an evaluation of a capacity building network of river engineers in the Nile basin (NBCBN-RE) to organising a high profile international symposium in 2007. He also has extensive experience in the development and conceptualisation of websites and the systems used to manage their information. Nick has an MSc in Water Science, Policy and Management from Oxford University. He received a BSc in Chemistry/Biology and a BA in Political Geography from University College Utrecht, both Cum Laude.

Nicoletta Forlano has a background in IT (programming and system analysis) and has mainly worked for international research organizations. She is specialized in multimedia design and heavily involved in user interface design for a wide range of applications. Nicoletta has worked for private companies, leading and participating in projects with major international institutions (National Geographic, MoMA, Getty Conservation Instituted, Silicon Graphics, Ericsson) and caries over 10 years of professional work experience in the United Nations with increasingly responsible positions in the area of multimedia, information/knowledge management and communication. In the Expert Meeting on WASH data she participated mostly as an observer on behalf of UN- Water, for which she covers the role of communication manager. She states that; 'having in mind our big task of developing a "federated" database among UN-Water members. I am truly convinced that user interface is a crucial piece in the development of any application or site, I make sure UN-Water products are user friendly and well designed.'

Richard Franceys, Director of the Water and Society programme at Cranfield University, UK, is on the External Advisory Committee of WASHCost. He has had many years experience utilising cost and service level data on behalf of customers in England through his membership of regulator Ofwat's Central Customer Services Committee, now the independent Consumer Council for Water. Dr Franceys has recently completed an urban focused study on managing costs and services for DFID, published by Earthscan as 'Regulating water and sanitation for the poor - economic regulation for public and private partnerships.'

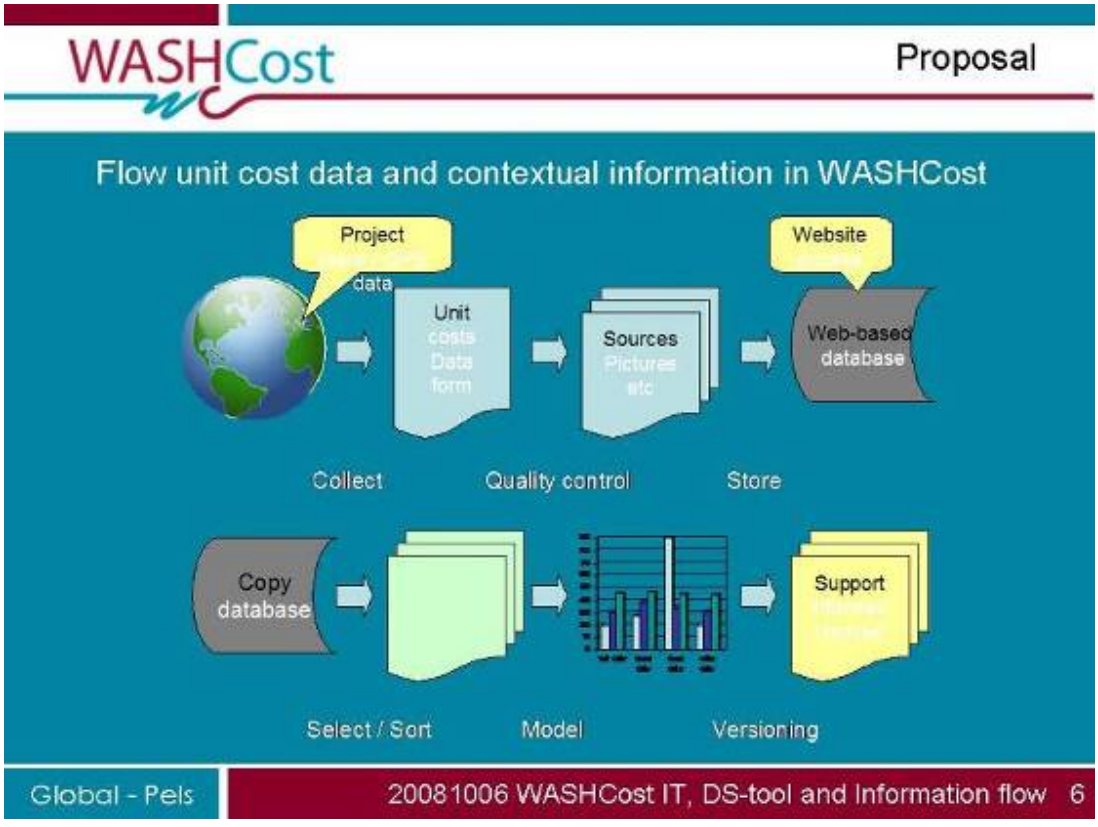
Simon Gordon Walker is an owner and Director of Artesia Consulting Ltd. Prior to this Simon spent nearly 20 years working for the Water Research Centre (WRc), where he was responsible for WRc's international business and worked as a consultant on institutional and capacity strengthening projects. In particular he worked on the development of performance improvement and benchmarking, from which the International Benchmarking Network (IBNET) was developed. Simon contributed to the WASHcost discussions and understanding of the challenges of the international collection of information and management of a complex international interchange of information for the benefit of a wide range of stakeholders for which the IBNET information would be useful. Simon was also able to describe the cost collection and management system in the UK water industry, called "TR61", which together with IBNET provides an example for understanding that will benefit WASHcost's own thinking.

Tom de Veer is co-founder of Connect International. He has more than 20 years of experience in development work as a consultant and as director of Foundation Connect International. Tom completed his study Tropical Agricultural Engineering Agricultural University of Wageningen in The Netherlands in 1988. He worked for many years in many of the large emergencies of the nineteen nineties (Sudan, North Kenya, Rwanda, Albania/Kosovo, etc.) functioning as manager of water and sanitation programs in camps, and coordinating NGO's for UN organizations in their water and sanitation programs. He started working as a consultant in 1995, being involved in evaluations, development of guidelines and writing of policies in the field of water and sanitation, emergencies, development and capacity building of local NGO's. As a director of Foundation Connect International he has been extensively involved in the development of and training NGO's in a large number of quality systems and software systems for local NGO's involved in community development and low-cost (SMART Tec) watsan technologies.

7. WASHCost DST development

The WASHCost project aims to apply Action Research and initiate Learning Alliances on unit costs for WASH services in rural or peri-urban areas and small towns in Burkina Faso, Ghana, India and Mozambique. At the end of the 5 year project a body of information must be globally available to support decision makers in the field of life cycle unit costs for WASHCost service delivery; commonly known as 'ICT' or 'DS' tool, further referred to as 'DS-tool'.

At the end of WASHCost the DS-tool – the body of information - is expected to be anchored in the WASH sector, hosted by a main player and used as framework / repository for enhanced sector performance. The picture below shows the flow of data and contextual information for WASHCost. Projects will be analysed on their unit costs underpinned with a wide variety of information sources resulting in a web-based database which can be accessed through a website. The basic three steps are collect, quality control and store.

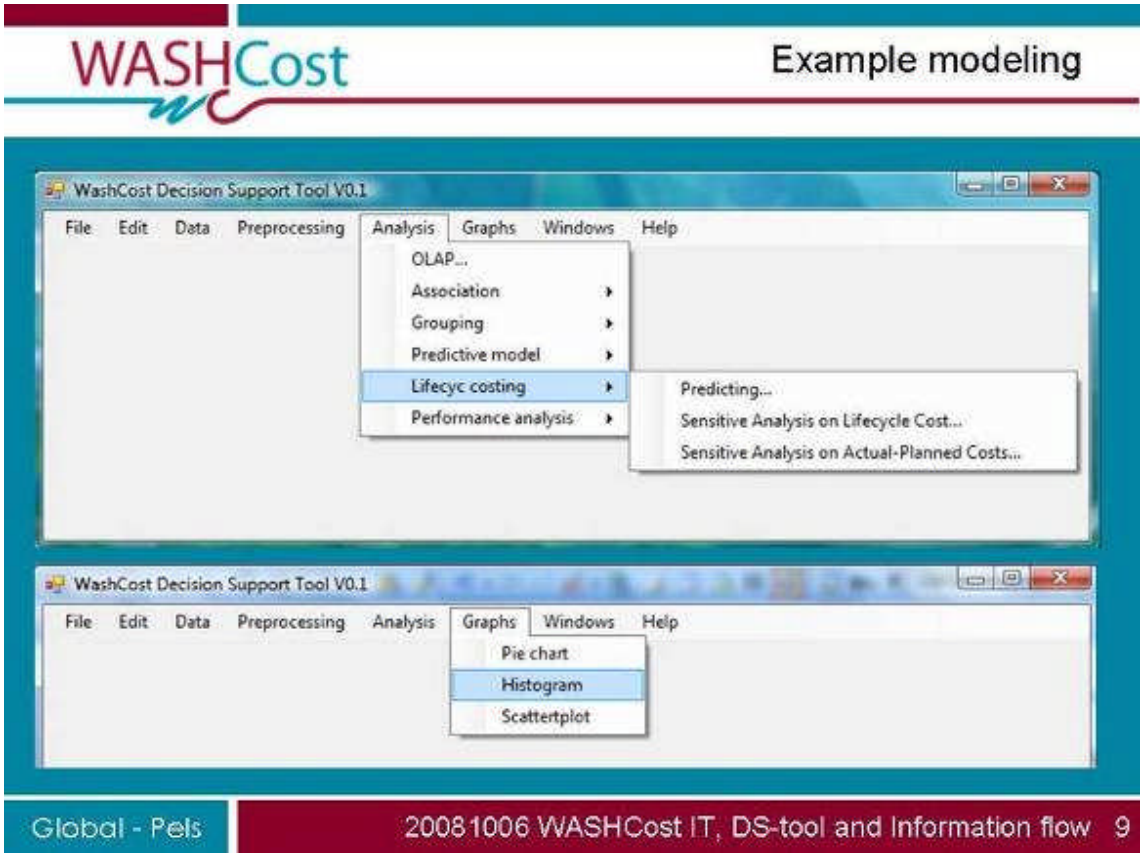


The bottom row shows how a copy of the database (containing the researched data plus underpinning information sources) can be used for further analysis – like modelling and versioning – and support decision making in regard to life cycle unit costs of WASH services.

Where are we?
From a user - technical perspective the DS tool is a website (or portal) routing access to three main sources (see also picture example front end below):
1 - a Wiki (based on SocialText) as Intranet for the WASHCost team members
2 - a WASH document repository topped of by an 'intelligent' search mechanism
3 - a database with a quality assured researched dataset[2] on WASH unit costs

Ad 1 and 2: To support information management and social learning WASHCost uses a Wiki (next to group e-mail and a newsletter) to support information package and flow. A Wiki is a collection of interlinked web-pages where each author can build upon what others have written.

A key component in WASHCost is the use of dialogue and process documentation (generating documents, video and audio) to quickly share knowledge and insights and to manage information with others. Also (background) research documents will be collected.

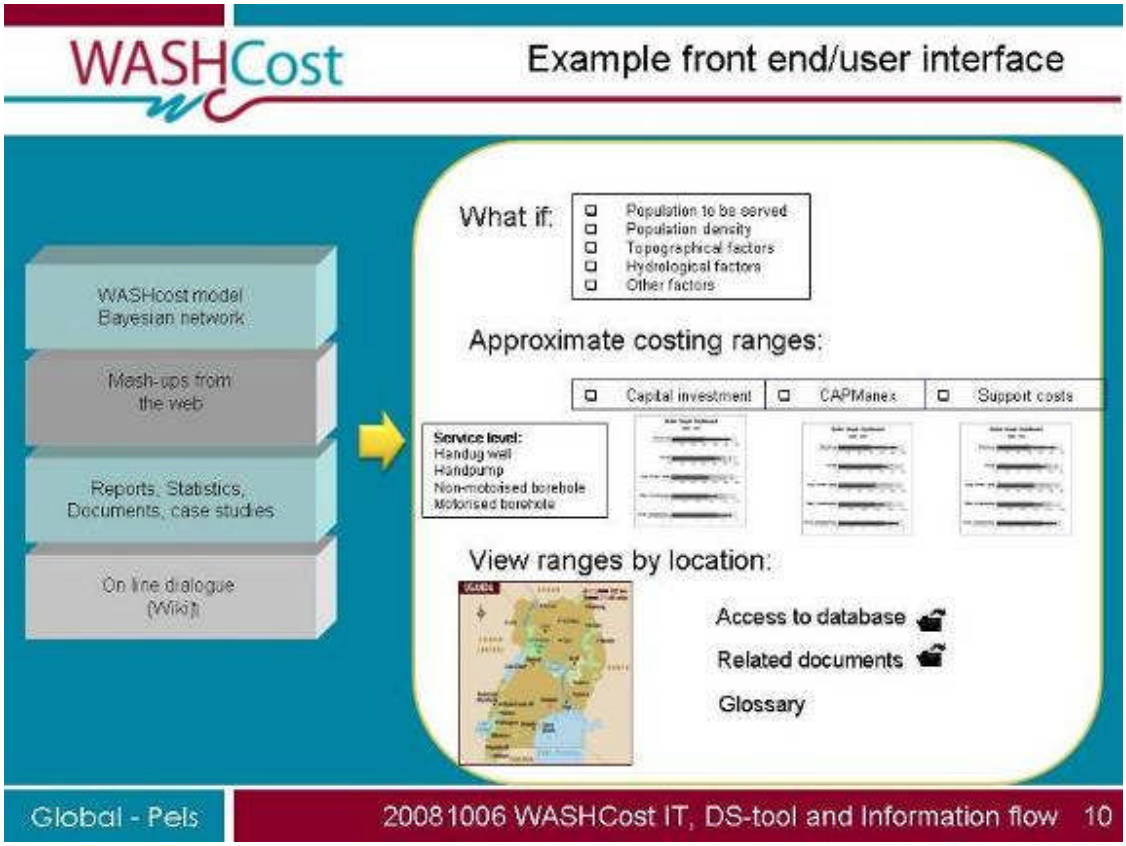


Ad 3: In the 'end of inception workshop' in Scheveningen Stone-IT (IT Company) gave a presentation on how to support WASHCost in the development of a DS tool. The aim (and milestone) was to hire an IT company by the end of 2008 to spearhead this part of the DS tool development. Up to then, only the conceptual list of meta data was available. Most lacking was a more-or-less final data collection form

[2] On the WASHCost dataset: 4 countries, 150 projects, 100 numerical variables will generate 60.000 numerical data. It needs a skilful mind to know which data may – or may not – be combined, modelled, left out, averaged etc; the data crunching will be on small amounts of data.

based on an agreed consolidated information framework (CIF). Also issues like the research protocol needs more flesh on the bone before final choices can be made in regard to the DS tool setup.

A first mock-up website from the user perspective looks like below.



What is next?

Currently, April 2009, the WASHCost country teams are busy testing the research protocol and site selection. By the end of 2009 there will be more insight into the range of data collected and the amount of numbers to crunch. In the mean time the WASHCost crew is getting an understanding what information can be collected where; also secondary information. The glossary (see mock-up above) is worked on content wise. The structure of – and navigation through is based on the wiki WASHCost uses: terms and concepts are linked through tags. When there comes more grip on the research outputs – depth, granularity, amount of data, quality etc – WASHCost can start contacting / developing user profiles. A major lesson learned from the expert meeting is that the user of the DS tool is the place to start developing the DS tool.

For more information on WASHCost, see www.irc.nl/washcost and / or www.washcost.info