

Network research by Data Graph Management for Capacity Development and Knowledge Building in Sustainable Sanitation

First David Franquesa*, Second Caterina Amengual²

*Phd student at UNESCO Chair of Sustainability at the Technical University of Catalonia.
d.franquesa@information-works.es

²Àrea d'Ecologia Interdisciplinar of the Universitat de les Illes Balears
caterina.amengual@uib.cat

Keywords: Network research, Sustainable Sanitation, Graph Databases Management, Capacity Development, Information Storage and Retrieval

Summary

The Millennium Development Goals (MDG) provide clear targets by 2015 and it turns out that sanitation is by far the largest of all the MDG targets affecting about 40% of the global population. The objective of the Sustainable Sanitation Alliance (SuSanA) is to show how Sustainable Sanitation projects should be planned with participation of stakeholders through capacity development activities. Developing the capacity of societies to collaboratively learn through change and uncertainty is fundamental for sustainability science. The aim of this contribution is to analyze the role of graph database management (GDM) for improve capacity development and knowledge building in the Sustainable Sanitation framework. We provide a theoretical model with four features of network research: link analysis, social network, pattern recognition and keyword search that we illustrate with some examples. Network research allows us to observe how the information in Sustainable Sanitation is scattered properly through the structure and also to detect the emergencies, objections and other characteristics of the network.

Introduction

In generic terms, sustainable systems are those which can be adapted to changing circumstances (Jeffrey, 1997). In order to be sustainable, a sanitation system has to be not only economically viable, socially acceptable, and technically and institutionally appropriate, it should also protect the environment and the natural resources (SuSanA, 2008). Information and educational programs, introduction of new policies and regulations and capacity building and training of professionals are needed for Sustainable Sanitation (Rockström et al, 2005).

The Millennium Development Goals (MDG) are an approach that provide targets promoted by the United Nations in 2000 to reduce poverty, hunger, illiteracy and others to ensure environmental sustainability. It was not until 2002 at the UN World Summit on Sustainable Development were water and sanitation were included as MDG7, target 10. Recently a pathway is formulated to achieve the target on sanitation through the sustainability. There are clear signs in the linkage between sanitation and all the other targets (Rosemarin et al, 2008).

Under this framework it is necessary to introduce the concept of ecological sanitation systems as safely recycle excreta and other organic waste products to crop production in such a way that the use of non-renewable resources are minimized. A change in the paradigm is necessary in the water management aspects. We need to end with the up to down vision, and start a learning process between different actors (UNDP, 2006, p. 388). Social learning is needed to start a change towards adaptive management systems to sustainability (Pahl-Wostl, 2007). Capacity development within Sustainable Sanitation it is a dynamic process where learning links up with live experiences to improve outputs, processes and products (Keen et al, 2005). Nowadays there is a lack on the outcomes monitoring when learning activities are done by different institutions at different levels. This field is taking up by improvisation, but it is not always working efficiently. How can engaging institutions record their outcomes in the learning approaches? There is some ways to state the results achieved by knowledge provision? These are some of the question that we try to answer.

In the field of Sustainable Sanitation great efforts are led to capacity development and training with special emphasis on school sanitation. Pathways for Sustainable Sanitation include a vision as an interplay between human behavior (cultural attitudes and norms) and appropriate technologies requiring stakeholder involvement in the planning and implementation steps (Rosemarin, 2005). Capacity development is understood as a process of unleashing, strengthening, creating and maintaining capacity over time. It applies to individuals, organizations and institutions. Capacity development is more than awareness of technical subjects and general organizational principles and it cannot be imported, but must be led from within the country itself (OECD, 2006). Capacity development within Sustainable Sanitation it is a dynamic process where learning links up with live experience to improve outputs, processes and products. Sufficient time and resources are necessary components to connect acquired capacity with action (SuSanA factsheet, 2009). The Sustainable Sanitation

Alliance joint organizations and institutions to work together in the learning process (SuSanA, 2008). This process is suited in the learning-by-doing approach, where both partners are learning at the same time. How can their record the effect of the appropriate knowledge and the level on the competences in an area as intangible as social learning? To provide the world with more Sustainable Sanitation solutions will require enormous efforts in the area of capacity development. New tools are needed to evaluate the achievements made in this field.

This paper argues that if the information on sustainable sanitation resides on relational data model, this imposes difficulties for decision making based on exploration of the relationships among the data, such as paths, neighborhoods, patterns, and in definitive all queries based on entities that are interconnected satisfying a given constraint. That means a bad performance on time and cost. For improve this, we present a technology based on Graph Database Models (GDMs) that implements efficiently four basic features of network research: link analyses, pattern recognition, social network and keyword search, as well as many applied research examples for decision making focused on capacity development and knowledge building activities. Changes are basically from a hierarchical structure to a network one. In the first, changes in the structure are controlled by an organization based on previous knowledge, while the second model is developed while growing the knowledge on itself.

2. Background

2.1 Holistic approach

The holistic view, proposed among others by the principles of Bellagio, is addressed in systems thinking. Systems' thinking is the attitude of the human being that is based on the perception of the real world in terms of totalities for their analysis, unlike the approach of the classical scientific method, which only sees the parts of this and so disjointed. Formally appears about 45 years ago, from the questions in the field of biology, (Bertalanffy , 1984) questioned the application of scientific method to problems of biology, because this was based on a mechanistic and causal that was weak as the scheme for the explanation of the major problems that occur in living beings.

This question raised an intellectual paradigm to better understand the world around us, the emerging the paradigm of systems theory. Systems theory integrates in the analysis of situations and in the conclusions derived from them, suggesting solutions that are considered to have several elements and relationships that form the structure of which is defined as a "system" and everything that makes up the system environment defined. The basic philosophy behind this position is the holism from the Greek (holos = whole).

Better understand the interlinkages between social and natural systems will help us take the appropriate action to act in coherence with the natural system, this is a concept linked to resilience (Holling, 2003). Sustainable Development would be one in which disturbances generated from the social system could be 'embedded' with the natural without assuming a

change of state which, in turn, affect strongly the dynamics of the social system. Achieving this goal requires action types: Coping with the impact of the shocks produced by the social system (trying to 'sync' human activities with the cycles of natural subsystem), and increase the capacity of the two subsystems to adapt to shocks (Pérez, 2008). Learning provides a basis for the joint action required to respond to social–ecological feedback (Folke, 2006, p. 253-267).

How decision support systems will help us on making sustainable development decisions if the underlying data model is too much structured and does not act effectively when our data are not structured. It seems that the theory of Ashby (1956) is still in effect, he proposed his Law of Requisite Variety, which stated that a model (i.e. a representation) can only represent some aspect of reality if it has sufficient internal variety to capture the complexity of that reality

2.2 Network research for Sustainable Development

In knowledge management social network research has been successfully with the aim of help to organizations better exploit the knowledge and capabilities distributed across its members (Davies, 2005, p. 133-149). A network map shows the nodes and links in the network. Nodes can be people, groups or organizations. Links can show relationships, flows, or transactions. A network map is an excellent tool for visually tracking your ties and designing strategies to create new connections, and also excellent 'talking documents' – visual representations that support conversations about possibilities.

Network research is hot today, with the number of articles in the Web of science on the topic of "social networks" nearly tripling in the past decades (Borgatti et al, 2009, p. 892-895). It affords to explain social phenomena in different approaches. Whereas traditional social research explained and individual's outcomes or characteristics as a function of other characteristics of the same individual (e.g., income as a function of education and gender), social network researchers look to the individual's social environment for explanations, whether through influence processes (e.g., individuals adopting their friends' occupation choices) or leveraging processes (e.g., an individual can get certain things done because of the connection he has to powerful others) (Borgatti et al, 2009, p. 892-895).

For analyzing intensity of relations among stakeholders social network analysis can be used. Networks provide a broad and inclusive framework (Davies, 2005, p. 133-149):

1. Networks can be described and analyzed at many scales, from interactions between individuals in small rural communities to international linkages between large organizations,
2. There is a range of tools available to describe and measure networks, which is relevant to the analysis, planning and evaluation of change in those networks,

3. There is an extensive and developing body of theory and research on the nature of networks, that spans many disciplines, and which is available to help inform development agencies' theories of change.

There are other areas of research and theorizing about social networks that have relevance to development aid projects. In the health sector there is already an established record of social network analysis techniques being used as part of epidemiological studies (Morris and IUSSP, 2004), as well as in studies of the effectiveness of health communications, especially in the field of HIV/AIDS (Davies, 2005, p. 133-149).

3. Graph Database Management and DEX technology

The term “data model” or “database model” has been widely used in the information management community: it covers various meanings. In the most general sense, a database model is a collection of conceptual tools used to model representations of real-world entities and the relationships among them (Silberschatz et al, 2008, p. 105-108). The term is also often used to refer to a collection of data structure types.

3.1 Limitations of the relational data model

The differences between GDM and the relational data model are manifold. The relational model is geared towards simple record-type data, where the data structure is known in advance (airline reservations, accounting, inventories, etc.). The schema is fixed, which makes it difficult to extend these databases. It is not easy to integrate different schemas, nor is it automatable. The query language cannot explore the underlying graph of relationships among the data, such as paths, neighborhoods, patterns (Angles, 2008, p.39).

The relational data model is now more than 30 years old. It is worth for a large number of scenarios and can handle certain types of data very well. For data that is semistructured and/or network oriented, the relational database offers poor runtime characteristics. Furthermore, it forces a static development cycle and is of little help to those who have to cope with a domain model that is constantly changing, even after deployment. This translates to wasted development time and money. Classical model were criticized for their lack of semantics, the flatness of the permitted data structures, the difficulties the user has to “see” the data connectivity, and how difficult it is to model complex objects (Levene and Poulouvasilis, 1990, p. 520-530).

3.2 When GDM should be applied?

The development of huge networks such Internet, geographical systems, transportation or automatically generated social network databases, has brought the need to manage information with inherent graph-like nature (Angles, 2008, p.39). In these scenarios, users not only keen on retrieving plain tabular data from entities, but also relationships with other

entities using explicit or implicit values and links to obtain more elaborated information. In addition, users are typically not interested in obtaining a list of results, but a set of entities that are interconnected satisfying a given constraint.

Cases like bibliographic database are a clear example where a more complex querying system would be beneficial. In these scenarios, the user might not be only interested in finding a specific author or publication, but to analyze the relationships within a group of authors or publication, to understand the relevance of a specific paper or any other implying the exploration of the relationships between entities (Martinez-Bazan et al, 2007, p. 573-582). Those environments impose three important problems: (i) the continuous growth of the data sources, (ii) the need for a versatile querying system that allows Information Retrieval queries with different flavors ranging from keyword search to the complex mining of patterns in graphs, and (iii) the need to integrate data coming from different sources to enrich the answers to complex queries over incomplete databases.

GDM are applied in areas where information about data interconnectivity or topology is more important, or as important, as the data itself. In these applications, data and relations among data are usually at the same level. Introducing graphs as a modeling tool has several advantages for this type of data (Angles, 2008, p.39):

1. **It allows for a more natural modeling of data.** Graph structures are visible to the user and they allow a natural way of handling applications data, for example, hypertext or geographic data. Graphs have the advantage of being able to keep all the information about an entity in a single node and showing related information by arcs connected to it (Paredaens and Tanca, 1995, p.436-453). Graph objects (like paths and neighborhoods) may have first order citizenship; a user can define some part of the database explicitly as a graph structure (Güting, 1994, p. 297-308), allowing encapsulation and context definition (Levene and Poulouvasilis, 1990, p. 520-530).
2. **Queries can refer directly to this graph structure.** Associated with graphs are specific graph operations in the query language algebra, such as finding shortest paths, determining certain subgraphs, and so forth. It is not important to require full knowledge of the structure to express meaningful queries (Abiteboul, 1997, p.1-18). Finally, for purposes of browsing it may be convenient to forget the schema (Buneman, 1997, p. 117-121).

3.3 Framework: DEX Technology based on GDM

In this subsection we present DEX as a possible technology based on GDMs, DEX is a high-performance exploration tool on large graphs for information retrieval. DEX affords the efficient implementation of four basic features of network research: link analyses, pattern recognition, social network and keyword search, afford realize different kinds of queries for graph exploration. DEX is developed by DAMA-UPC¹ and is basically characterized by three

properties: (i) data structures are graphs or any other structure similar to a graph; (ii) data manipulation and queries are based on graph oriented operations; (iii) and there are data constraints to guarantee the integrity of the data and its relationships.

In Figure 1 we can see a graph structure for the case of Bibex, a bibliographical database, where relations are authors writing scientific papers. Queries are based on authors or keywords and exploring the graph one can obtain related information and statistics.

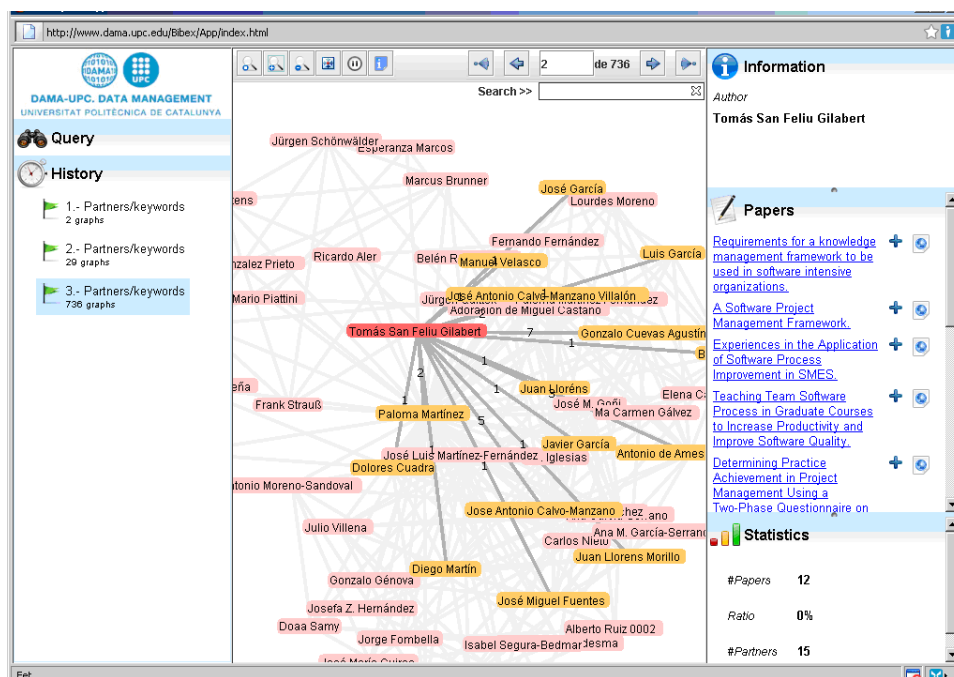


Figure 1: Shape of a query in Bibex

4. Theoretical model with DEX technology for Sustainable Sanitation

Data on Sustainable Sanitation is highly-interconnected and has a complex structure, which is difficult to capture in any current database (Excel files, Forums, many databases, etc.). Sustainable Sanitation information system is based on relational model, this model are well suited for queries based on values, like equalities or range search, but in these models the exploration of relationships must be always set explicitly by joins even if foreign keys have been declared, and it becomes really difficult to explore all the potential relationships of a node, institution or project. For self-relationships, the relational queries require recursive extensions that are more difficult to create and manage. On the opposite side, the natural mechanism of GDMs is the automatic exploration of the relationships in a graph, represented in the form of edges between nodes. It is a relationship rather than a value oriented analysis.

It seems that the problem of Sustainable Sanitation Alliance and many other growing organizations related to sustainable science that fall into their database, it is not designed to integrate data from multiple sources, or to respond efficiently queries based on entities that are interconnected satisfying a give constraint, and other issues that are not come by default in relational model. New forms of organization are emerging to deal with sustainable science process. These do not have a closed structure from the beginning; they are growing spontaneously through the learning process.

The network offers numerous resources and communication channels for Sustainable Sanitation learning partners. The resource format is diverse, so are videos that cover a specific problem, such as forums where experts to advice staff that is carrying out projects in the nodes. The data from our system come from different bakeries (e.g. db, Forums, Emails, ...). We have to collect all this data and give to the users the possibility to analyze it. The main drawback of Sustainable Sanitation is interlace all this information, place it in the partners and provides an impact assessment that causes this information when reach on-line to the users or on-site trainings.

4.1 Sustainable Sanitation data Model proposal

The following Figure 2 is an E/R diagram that conceptualize some of the entities that could be related on capacity development activities on the Sustainable Sanitation field. The principal data sets are PERSON, which contains all individuals involved in any activity; ACTIVITY, which contains basic information of each activity on capacity development or knowledge building, ROLE_P, which contains who (PERSONS) and how (ROLE_TYPE) are they participated into a activity , LINK, which contains the relations between activities using different types of links (Cites, Answer,...).

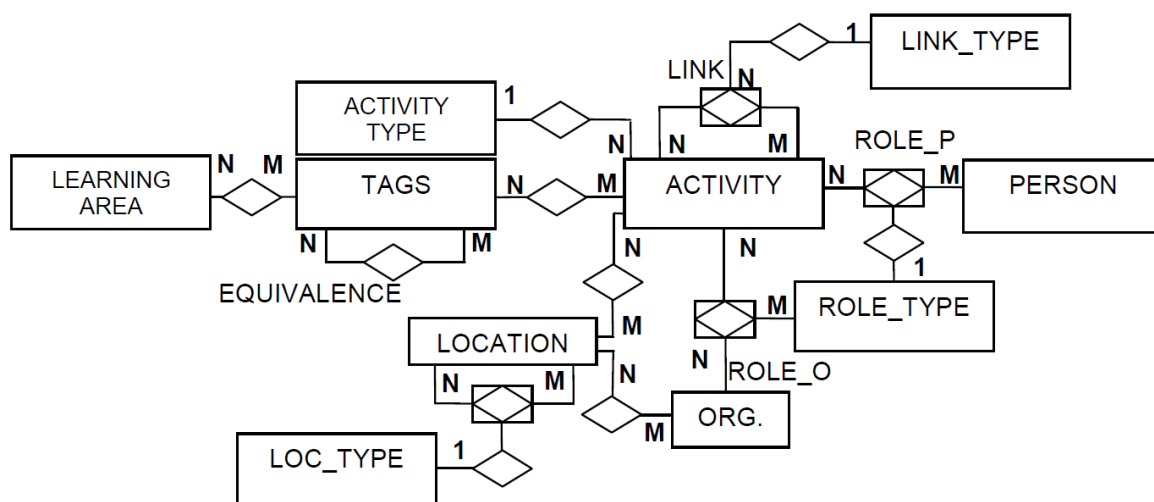


Figure 2: E/R Diagram of Sustainable Sanitation

Figure 3 is a list of instances that could be inside the information system. e.g. ACTIVITY_TYPE holds a set of tools for capacity building that can be found in Sustainable Sanitation, expressed as a relational database model.

ACTIVITY_TYPE	ROLE_TYPE	ACTIVITY	LINK_TYPE
Video	Author paper	Info id	Cites
Case study	Project Trainer	Title	Answer
Project	Project Responsible	Text	Equivalent
Training	Project Manager	Url	Recommend
Movie	Technical Expert	Date	Responds to
Paper	Map		Arises from

Figure 3: Example Instances of E/R Diagram

4.2 Use Cases

The objective of this section is to verify the capability of the model to solve different kinds of queries that practitioners and researchers need to do in their day to day work.

The examples are grouped by four features, as we have presented with the DEX model: a) Link analysis, b) Social networks, c) Pattern recognition and d) Keyword search. These features are essential when members of Sustainable Sanitation want to assess the impact of their activities in capacity development. During the development we present the results in a “map network” form for each query launched.

4.2.1 Link Analysis

In Link Analysis we are interested in exploring the relations between the nodes of the graph, navigating the edges between them, e.g. to get all the information of a Sustainable Sanitation conference, the result is a graph where you can visualize different kinds of relationships to this conference (board directors, assistants, topics, scientific panel, publications, relations among stakeholders established thanks to this meeting, etc).

Query 1 (Q1): Get all the information of a conference (ACTIVITY_TYPE = conference)

An example of link analysis is Q1, where all the information of a conference is obtained in a single graph. The root node will be a node from the data set ACTIVITY with ACTIVITY_TYPE = "conference". From this initial node, we explode a graph containing all the ROLE_P and ROLE_O (Role of persons and organizations in Sanitation conferences), persons, information of the tags used, including the chain of information referenced to the root by LINK, that can be any kind of ACTIVITY_TYPE. Note that this could pose serious problems to traditional relational systems that would have to resort to recursive queries including a large number of join operations, increasing significantly the complexity of such

queries (Martinez-Bazan et al, 2007) A simplified example of information retrieved for movie “Water and sanitation in NIGER challenges” is Figure 4.

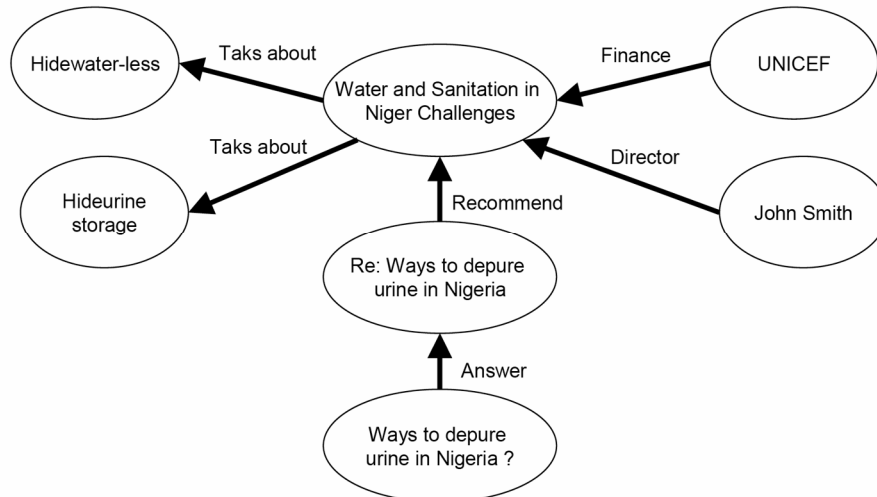


Figure 4: Q1, Conference link analysis

4.2.2 Social Network

Previous query show that DEX can be used to explode the links between different entities in a graph. Now, we depict an example where DEX is used to analyze social networks. Social Network is focused on the relationship between different groups of nodes with the same affinity. Let us consider all the technical experts in our database who have participated in the same sanitation project to form a group in a social network. Specifically, we define a partnership as the relationship between two capacity development experts who have performed in the same project. Additionally, we impose two restrictions to this query. First, we restrict just to items tagged as “sanitation projects”. Second, we restrict the participation as a “technical expert”. We apply these two conditions because (i) we want to increase the query complexity rather than always exploring everything and, (ii) our database contains a lot of items extracted from NGO, blogs, journal database, etc that could provide unrealistic relationships between technical experts.

Query 2 (Q2): Find the minimum collaboration distance between two technical experts. Q2, tries to find the minimum distance between two persons that has worked as a technical. If distance is 1 it means that both have worked in the same project; a distance of 2 means that they never worked together, but exists at least another partner who has done a project with both of them.

Query 3 (Q3): Find the full relationships network of all the partners of a technical expert. Q3 is a more complex query. Instead of looking for relationships between two

experts, we are now interested in knowing the full network of relationships of all the partners of one expert. This kind of analysis provides us with a lot of information about partnership patterns, groups of technician based on the sanitation agencies.

4.2.3 Pattern recognition

Pattern recognition defines a different kind of queries, where a lot of potential graphs can be created and explored, but only a few of them will qualify because they match a certain pattern.

Query 4 (Q4): Find all the responsible that have worked with the same technical expert in three different projects made in a period of time of five years. As an example of pattern recognition, Q4 tries to find all the coordinators with three projects in less than five years with the same technical expert, i.e., it would be some kind of 'muse' detector query. This is a complex query that not only requires pattern detection but also involves several data filters like coordinator role, technical expert role or ACTIVITY_TYPE equal to project. See Figure 5.

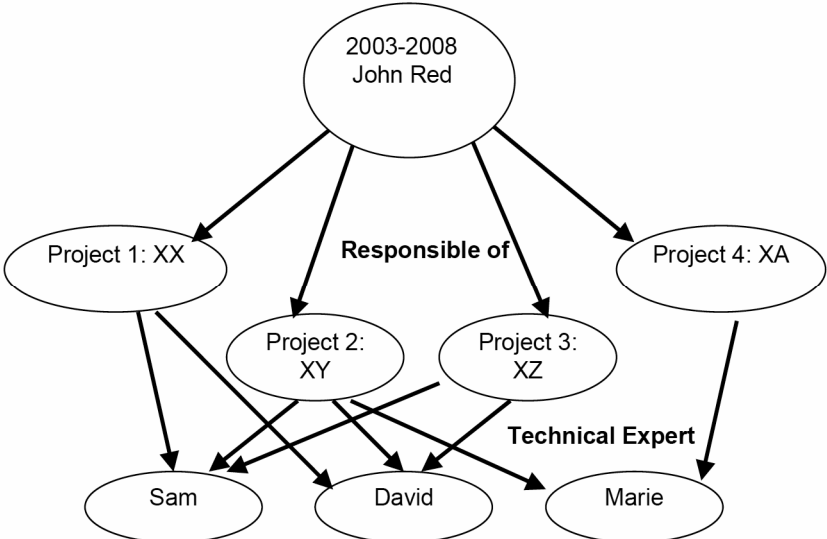


Figure 5: Q4, 'muse' detector

Query 5 (Q5): How many practitioners are doing BIOGAS trainings and projects after receive a training course? The purpose of this query is to find a training course that has led to other training courses and projects within the same topic. The training courses that appear may be considered that have had a desired impact. For example in Figure 6, Mike in 2008 was the teacher of a course on BIOGAS. In this course attended Augusto, who in turn was the teacher of a subsequent course - also on Biogas- where he attended John, and also Augusto did a project on Biogas. This query uses the feature of pattern recognition, based on exploration of the relationships among the data, such as patterns, and return all entities are interconnected satisfying a given constraint.

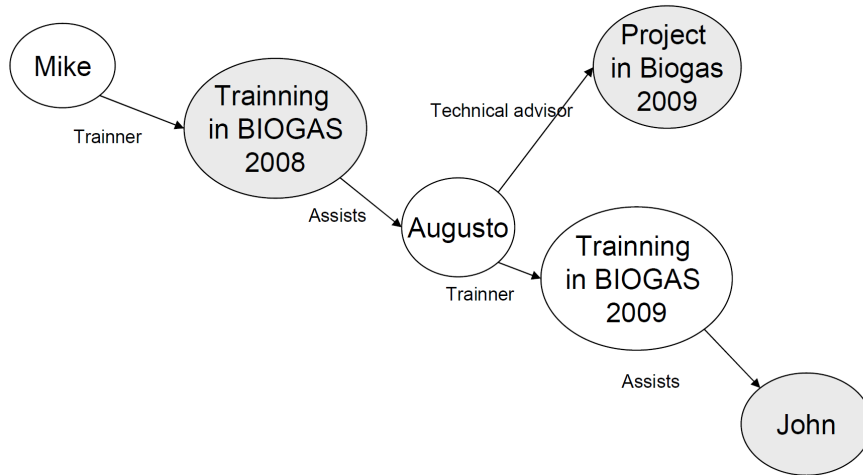


Figure 6: Q5, Mike impact training

4.2.4 Keyword search

Finally we describe Keyword search as a network research characteristic. In others queries the user knows the schema or at least part of it. We assume that the user has knowledge on how the data is structured. However, this assumption may be unrealistic in some scenarios like the web or documental database. DEX is also suitable to perform a keyword search, where the user is assumed not to know anything about the organization of the data. In conventional databases models like the relational model, this time of search requires a full indexing of all the string columns and it becomes unfeasible due to the high cost in terms of storage size and performance. DEX can take advantage of dictionaries and compressed structures (Martinez-Bazan et al, 2007, p. 573-582).

Query 6 (Q6): Return all the context information of all the entities containing the tag Biogas production, expert John Smith and Country India. Such queries are done separately in different information sources, db queries, queries in forums, consultations db project. Sustainable Sanitation does not has a tool that integrates all sources in one query, in spite of the use of Internet search engines, but narrow the search to the semantic domain of Sustainable Sanitation is difficult or is not possible, and the format of results presented to the user is only a list. The present potential for the user is the ability to perform a search across different data sources and show the results in a visual format that allows the user to navigate easily. The Figure 7 shows the result obtained.

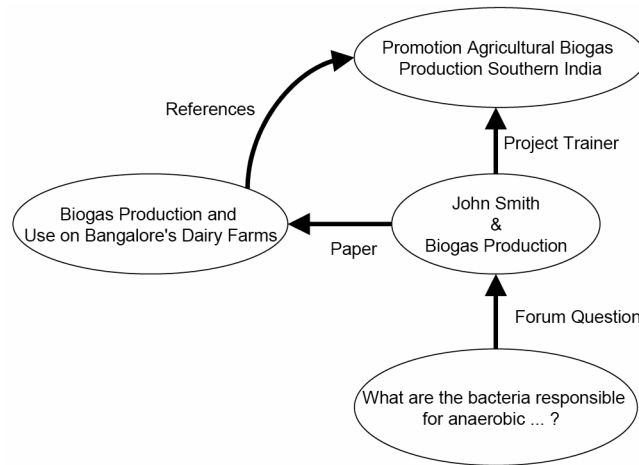


Figure 7: Q6, Biogas production search

5. Practical example in a query

We explain a case using again four kind of features: link analysis, social networks, pattern recognition and keyword search. We are based this example in a real case study on Sustainable Sanitation alliance website (Rieck, C.; Onyango, P., 2009). We suppose a responsible for Sustainable Sanitation partnership project coordination; who have to develop a project with a very tight budget, living in a city of Africa quite distant from experts on Sustainable Sanitation. In addition, the staff does not have all the necessary skills to develop the project. This is the reality in most of the deployments. We need to find an expert who can do training to our staff in the need it skills. The result would be a set experts names that could do the training or recommend somebody to do it. We can start this research with a query as: give me all the information about Projects and Trainings in “Kenia” related to “Biogas” and “Urine Diverting De-hydrating Toilets – UDDTs”.

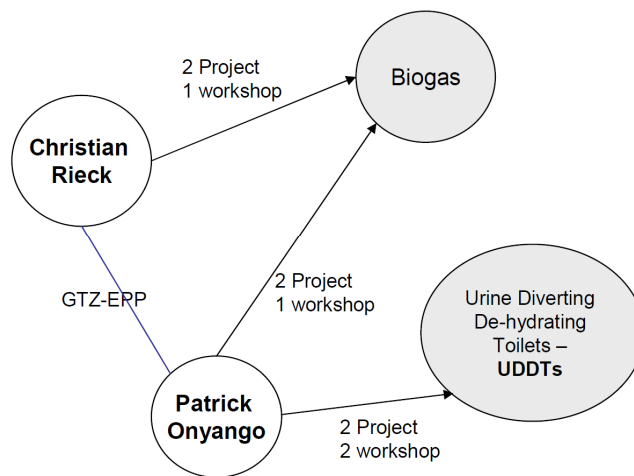


Figure 8: Projects or Trainings in Kenia about Biogas or UDDTs

Patrick Onyango seems to be the person that has done more projects and trainings about Biogas and UDDTs. Before contacting him, we decided to investigate, to validate the quality of related ACTIVITIES they have done, we use link analysis feature, just doing a click in the edges between Patrick Onyango and Biogas and UDDTs we can explore the information related to these projects and workshops, see figure 9

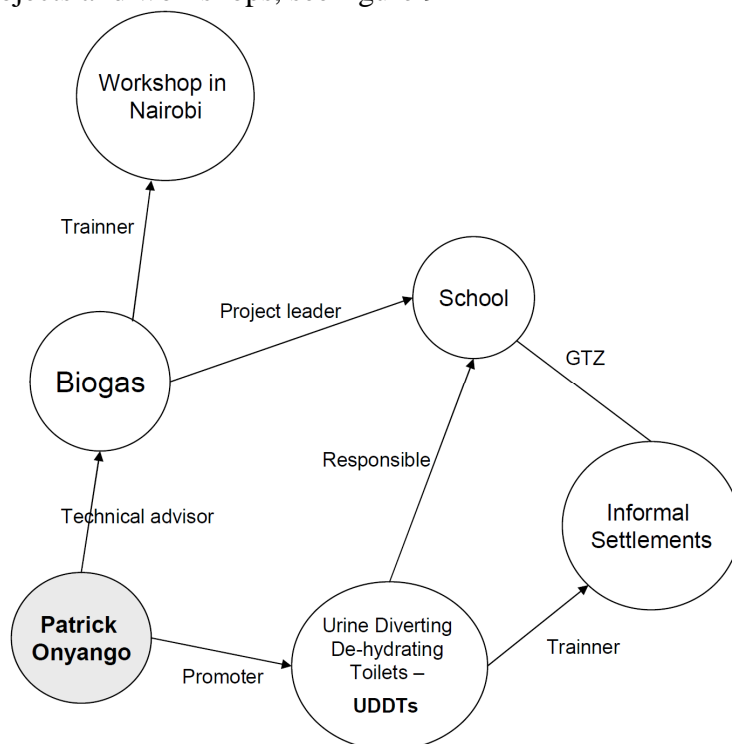


Figure 9: Projects or Trainings of Patrick Onyango

After validating the relation between Patrick and Biogas, we want their referrals. Who do we ask? Then we make a Social Network query that will return people who have been in relationship with Patrick Onyango and have been in relationship with SuSanA organizations. In Figure 10 there are the persons and their relationship with Patrick. If we want we can go on doing more link analysis in these relationships.

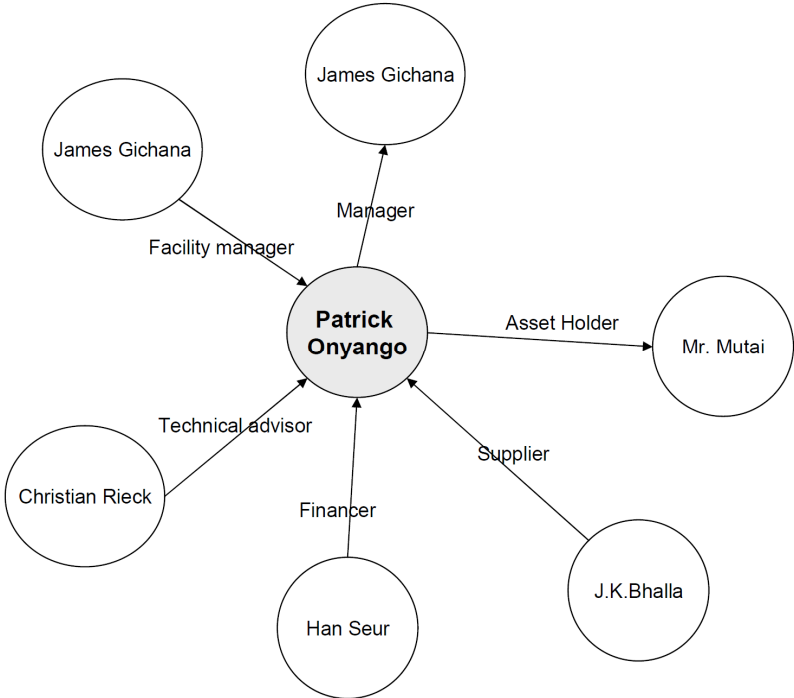


Figure 10: Social network between Patrick Onyango and members of SuSanA organization

After make this network research we can have clear idea about the scene on biogas in Kenya, among the people who is involved in different projects. The main advantage is that GDM can analyze information from different sources together, so you are not supposed to introduce manually all the information in a unique database. The search engine can track the information generated in different formats.

6. Conclusions

This paper argues that if the information on Sustainable Sanitation resides on relational data model, this imposes difficulties for decision making based on exploration of the relationships among the data, such as paths, neighborhoods, patterns, and in definitive all queries based on entities that are interconnected satisfying a given constraint.

The main feature of GDM is that data model can hold the interrelationships between the elements more efficiently than relational databases. Applying the DEX technology - one of the technologies that implements GDM -, we can retrieve information from the global World Wide Web to the local documents or databases and dump it in a graph data warehouse system, and make basic and complex network operations that can bring us interesting information about the capacity development activities and their interaction.

This approach is not a close tool; it should be seen as a new perspective on the data collection, store and query analyzing. It improves research of the network through the features of: link analyses, pattern recognition, social network and keyword search. These features can be efficiently used to manage and evaluate more easily the capacity development and knowledge building process.

Some examples of how to deal with complex questions have been exposed to express the novelty of this network research system.

This approach could help in the social learning process because can empower to someone who is just beginning on Sustainable Sanitation to deal the complexity when this person have a few knowledge on the hole scene. The possibility to manage larges amounts of information through their relations open a new paradigm based on systems thinking than can help to advance on the necessary pathway to sustainability on sanitation.

Acknowledgements

We thank to the EASY-ECO programme to give us the opportunity to share knowledge among scholars interested in sustainability assessment. Also to the open-source concept for capacity development used on the Sustainable Sanitation Alliance website², that help us to work with real data on the field. Finally to the Information Works GmbH. who supports the research of David Franquesa on the topic of information management for Sustainability.

Notes

¹ DAMA-UPC, the DATA Management group at Universitat Politècnica de Catalunya (UPC) is part of the Computer Architecture Department (DAC). <http://www.dama.upc.edu>

² www.susana.org

Bibliographic references

- Abiteboul, S. (1997). “Querying semi-structured data”, In Proceedings of the 6th International Conference on Database Theory (ICDT), LNCS, vol. 1186. Springer, p. 1–18.
- Angles, R. (2008). “Survey of Graph Database Models”, ACM Computing Surveys, vol. 40, no. 1, p. 39. ISSN 0360-0300.
- Ashby, W. R. (1956). Introduction to Cybernetics, New York: Wiley.
- Bertalanffy, L.V. (1984). General system theory: foundations, development, applications, New York. Edición: 9 ed.
- Borgatti, S., et al. (2009). “Network Analysis in the Social Sciences”, Science, vol. 323, no. 5916, p. 892-895.
- Buneman, P. (1997). “Semistructured data”, In Proceedings of the 16th Symposium on Principles of Database Systems (PODS), ACM Press, p. 117–121.
- Davies, R. (2005). “Scale, Complexity and the Representation of Theories of Change”, Evaluation, 11 (2), p. 133-149, <<http://evi.sagepub.com/cgi/content/abstract/11/2/133>>, [Accessed 19/04/2009]
- Durkheim, E (1951). Suicide: A Study in Sociology, Free Press, New York.
- Folke, C., (2006). “Resilience: the emergence of a perspective for socialecological systems analysis”, Global Environmental Change 16 (3), p. 253–267.
- Güting, R. H. (1994). “GraphDB: modeling and querying graphs in databases”, In Proceedings of the 20th International Conference on Very Large Data Bases (VLDB), Morgan Kaufmann, p. 297–308.
- Holling, C.S. (2003). Foreword: The backloop to sustainability, Pages XV-XXI in F. Berkes; J. Colding and C. Folke, editors. Navigating Social-Ecological Systems. Cambridge University Press, Cambridge, UK.
- Jeffrey, P.; Seaton, R.; Parsons S.; Stephenson, T. (1997). “Evaluation methods for the design of adaptive water supply systems in urban environments”, Water Science and Technology, Vol 35 (9), p. 45-51.
- Keen, M.; Brown, V.; Dybal, R. (2005). Social Learning in Environmental Management, Earthscan, London, UK.
- Levene, M.; Poulouvasilis, A. (1990). “The Hypernode model and its associated query language”, In Proceedings of the 5th Jerusalem Conference on Information technology, IEEE Computer Society

- Press, p. 520–530.
- Martínez-Bazan, N.; Muntés-Mulero, V.; Gómez-Villamor, S.; Nin, J.; Sánchez-Martínez, M.-A.; Larriba-Pey, J.-L. (2007). “Dex: High Performance Exploration on Large Graphs for Information Retrieval”, In Proceedings of the CIKM conference, Lisbon, p. 573-582.
- Morris, M.; and International Union for the Scientific Study of Population (2004). *Network Epidemiology : A Handbook for Survey Design and Data Collection*, Oxford, New York, Oxford University Press. ISBN 0199269017 9780199269013.
- OECD (2006) *The Challenge of Capacity Development, Working Towards Good Practice*. Organization for Economic Cooperation and Development, Paris.
- Pahl-Wostl, C.; Sendzimir, J.; Jeffrey, P.; Aerts, J.; Berkamp, G.; Cross, K. (2007). “Managing change toward adaptive water management through social learning”, *Ecology and Society* 12(2): 30. <<http://www.ecologyandsociety.org/vol12/iss2/art30/>> [Accessed 19/04/2009]
- Paredaens, J.; Tanca, L. (1995). “G-Log: A graph-based query language”, *IEEE Trans. Knowl. Data Eng.* 7, 3, p. 436–453.
- Pérez, S. (2008). “Procesos Sociales y Desarrollo Sostenible: Un ámbito de aplicación para el análisis de redes sociales complejas”, *Revista Internacional de Sostenibilidad, Tecnología y Humanismo*, ISSN 1988-0928.
- Rieck, C.; Onyango, P. (2009). “Naivasha Public Toilet Facility with Water Kiosk”. Case study of SuSanA projects. Sustainable Sanitation Alliance. <http://www.susana.org/lang-en/case-studies/region/ssa> [Accessed 5/12/2009]
- Rockström, J.; Nilsson, G.; Falkenmark, M.; Lannerstad, M.; Rosemarin, A.; Caldwell, I.; Arvisdons, A.; Nordström, M. (2005). *Sustainable Pathways to Attain the Millennium Development Goals: Assessing the Key Role of Water, Energy and Sanitation*, Stockholm Environmental Institute, Sweden.
- Rosemarin, A. (2005). *Putting Sanitation on the Sustainable Development Agenda*, 3rd International Conference on Ecological Sanitation. Durban, South Africa.
- Rosemarin, A.; Ekane, N.; Caldwell, I.; Kvarnström, E.; McConville, J.; Ruben, C.; Frodge, M. (2008). *Pathways for sustainable sanitation IWA*, London
- Silberschatz, A.; Korth, H. F.; Sudarshan, S. (1996) . “Data models”, *ACM Computing Surveys*, 28, 1, p. 105–108.
- SuSanA. (2008). *Towards more sustainable sanitation solutions, Version 1.2*. <<http://www.susana.org/images/documents/02-vision/en-susana-vision-statement-I-version-1-2-feb-2008.pdf>> [Accessed 19/04/2009]
- SuSanA factsheet (2009). *Capacity development. March 2009*. <http://www.susana.org/index.php/lang-en/working-groups/wg01> [Accessed 7/12/2009]
- UNDP. (2006). *Beyond Scarcity: Power, poverty and the Global Water Crisis*, Human Development Report, NY. p. 388.