DKN PROGRESS REPORT, SEPTEMBER, 2005

Diaspora Knowledge Network Project

Progress Report, September, 2005

General Overview:

The following three ideas serve to organize work on the DKN project:

- Members of Scientific and Technical diaspora are potentially in a position to mobilize the resources which are available to them in their host countries for use by the scientific and technical networks in their home countries. The first goal of the DKN project is to identify and model the mechanisms leading to this mobilization. The result will be a ***mobilization scenario***.

- The second goal is to use the software under development by the different teams taking part in the DKN project to ***computer-support*** the implementation of this mobilization scenario.

- The third goal is to publish a book in order to provide an understanding of how computer-supporting diaspora knowledge networks can contribute to scientific and technical development spanning the North-South divide.

1.) A three step "mobilization scenario"

1.a.) Defining "diaspora projects" that span the North-South divide

- In order to define a diaspora project, what *information sources* are used – personal network of relationships, scientific publications, grey literature reports, congresses and workshops,… and what *access channels* – face-to-face meetings, telephone mediated communications, computer mediated communications, information based technologies (libraries, databases, documentation centres,…?
- What resources have to be assembled in order to build a viable project?
    - o Human skills;
    - o Technical infrastructure (computers, telecommunications, networks, databases, standards,...);
    - o Institutional support (local support in the labs of diaspora members, government support, civil society support,…)
- What incentives exist for encouraging members of a S and T diaspora to engage in an effort to mobilize resources for their home countries?

With respect to the first two questions, a report produced by Ximena Castro-Sardi and William Turner summarizes the work which has been done to identify information sources, information channels and resources for project definition. Problems of defining a case study area and building an interview guide are discussed (annex 1).

With respect to the problem of incentives, Jean-Baptiste Meyer is pursuing the work he initiated in the IRD Expert Group Review of Scientific Diasporas
(http://www.limi.fr/Individu/turner/DKN/Completed_research/IRD-plaquette_EN.pdf).
To this end, Jean-Baptiste and Ximena Castro-Sardi took part in a July 2005 symposium on diaspora and international cooperation in Colombia entitled "Towards the construction of a Knowledge Society

for the XXI century in Colombia". It was an occasion to meet people and begin a document analysis (see annex 3).

1.b) Mobilizing support behind a diaspora project

Activity in a DKN generates a constant flow of documents which mediate the interactions of people located in different places. These documents will appear as messages directed to the DKN Bulletin Board, as contributions to the DKN Forum or as a more formal contribution classed in one of the information bins of the DKN Document Management System (studies, articles, questions for debate, content summaries of chosen articles or books and bibliographic references, institutional reading and background texts).

Computer assistance is needed in order to transform these "repositories" – the information bins of the document management system and the message archives produced through the on-going use of the DKN Bulletin Board and Forum – into an active interaction space which can be used for mobilizing support behind a diaspora project. In the DKN project this computer assistance will take the following forms:

- Use of "broadcast" and "pull" technologies to implement document-mediated awareness of events in the different geographical locations comprising a DKN;
- Use of computer-assisted project planning procedures as a means of building trust in the efficiency of collective action;
- Use of graph analytical techniques to detect the semantic structures underlying text mediated interactions in order to help improve information flows within a DKN
- Use of automatic document classification techniques as a means of annotating documents according to their specific contribution to on-going activity
- Use of visualization techniques for displaying different points of view on themes of interest to a DKN

More will be said about these technologies in point 2 below.

1.c) Running a project and consolidating it over time :

We assume that successful DKN project management will require managing three, largely independent processes:
- *A social process* linked to the dynamics of document-mediated interactions over Internet;
- *A technical process* of constantly improving the DKN technical infrastructure;
- *And an organization process* concerning trade-offs in the way of managing the cognitive and social dimensions of ICT use, trade-offs which can be evaluated when attempting to strike a balance between computer and human assisted access to DKN resources.

We also assume the need for *building a common language* given the different points of view occasioned by the need of working across the North-South divide.

See point 3 below for the DKN method on testing these assumptions.

2.) Computer-supporting the mobilization scenario

The first 6 months of the project are dedicated to:

- better understanding the mechanisms for identifying the information sources, information channels and resources needed to define a diaspora project; (point 1.a above)
- modelling the role of document-mediated interactions in mobilizing support behind the project (see point 2.a. below);
- defining a software integration plan to implement and test the model (see point 2.a. below).

The last 6 months of the project are dedicated to:

- Implementing the software integration plan in a case study area;
- Drawing conclusions with respect to the goal of managing the social, technical and organizational processes (see point 2.b. below):
- Generalizing results to other areas of diaspora activity.

2.a.) The software building blocks of the DKN platform

Interacting through documents at a distance implies using documents:

- To make people aware of interaction opportunities;
- To collectively construct the meaning of on-going activity (sense-making through storytelling);
- To build-up trust by documenting progress towards meeting collective goals and objectives;
- To overcome cognitive diversity while avoiding the dangers of organizational closure as a result of intellectual in-breeding;
- To manage the process of knowledge creation through debate and negotiation.

Each of the above points will be looked at briefly in what follows in order to explain the different software packages that are to serve as building blocks for the construction of the DKN platform.

2.a.1.) Using software to make people aware of interaction opportunities.

The VECAM system is used in the DKN project for this purpose. It combines "broadcast" and "pull" technologies:

- "Broadcast technologies" are intrusive in the sense that a message sent by a registered member to the VECAM Bulletin Board or to the VECAM forum is automatically delivered to all the other registered DKN members.
  - o The VECAM "Bulletin Board" provides members with news on current events and information on activities of interest to DKN membership.
  - o The DKN "forum" allows a member to post a message in a variety of languages, consult an archive of all the messages posted and carry out an information search on this archive by sentence or key-word. Each posting is registered chronologically and can be threaded with other messages, however, it can't be deleted.

- "Pull technologies" demand making a special effort to consult a document management system (DMS) in order to monitor a situation.

o The VECAM "DMS space" is structured according to the following rubrics: studies, articles, questions for debate, content summaries of chosen articles or books and bibliographic references, institutional reading and background texts.

It is expected that members of a DKN project will use the Bulletin Board and the Forum, and file documents in the different information bins corresponding to the VECAM rubrics as a way of staying in touch with what is going on in a project. This is a strong hypothesis and, in order for it to hold, it appears that two conditions have to obtain:

- The first is the need for a "user advocate" (who can be the member of the S and T diaspora who has taken the initiative to launch a DKN project, but who will more likely be an information professional capable of helping in detecting suitable information sources, human, material and social resources as well as in overcoming problems of "channel" use in mobilizing support for the DKN project. (see the report on the VECAM "facilitator" meeting);
- The second is finding a negotiated equilibrium between the "intrusiveness" of broadcasting techniques and the "non-use" of polling techniques.

(See the report ("Annex 2") by Ximena Castro-Sardi on the **i-Twinning Facilitators Evaluation Workshop (1-3 June 2005 – Panama)** which was organized by VECAM to look at the role of User-advocacy in organizing interactions among associations working in Latin America and Europe on similar types of projects.

2.a.2. Using software for sense-making through storytelling

Empirical evidence shows that despite what is often a heavy investment to build and maintain document management systems and message archives, these are frequently underused as sources of information for building projects and finding human and material resources. Storytelling is seen as a means of overcoming this problem and increasing the vigor of document flows in a DKN. The steps in a storytelling scenario are the following:
- Documents are filed in information bins or posted on bulletin boards or a forum by their authors;
- These documents are automatically classed and annotated as contributing to efforts aimed at consolidating mainstream interests, better understanding peripheral questions or as contributing new and original content;
- Their authors, who are members of the DKN working in geographically distributed locations, are thereby identified as "stakeholders" working in favor of a particular kind of investment which could potentially affect the course of collective action.
  - o Investing in mainstream concerns suggests growing consensus within the group as to what is important and what isn't;
  - o investing in peripheral issues suggests the determination of some actors to enlarge the scope of collective action which implies modifying resource allocation strategies in the group;
  - o investing in new areas of emerging interests suggests a lack of consensus as to what is important and what isn't.
- Storytelling focuses on resource allocation strategies, highlights different actor options ranging from building consensus to growing contestation, and identifies who is leading the charge against established positions or seeking to consolidate them.

It is expected that this type of story can be used reflexively by the advocate of a DKN project (whether the diaspora member who has taken the initiative to launch a project or an information professional) to encourage a new cycle of document production.

LIMSI has research underway based upon the idea that an information professional will have to become a storyteller in order to help transform information repositories (DMS and Message archive) into interaction arenas for creating new knowledge. A document filed in an information bin of a DMS or posted on a Bulletin Board or Forum needs to be contextualized in order to take on meaning. When contextualization is done through storytelling, the "event" of filing a document or posting a message is analyzed in terms of its potential impact upon the course of on-going, collective activity. Three categories serve for determining this impact:

- The list of subjects identified as being of mainstream importance to a DKN
- The list of subjects classed as being of peripheral interest
- The list of new, emerging subjects potentially important for changing the direction of collective action.

A document can be classed using a variety of similarity measures as contributing to one of these three categories. Storytelling consists in using this knowledge to weave what an author says in his document into an interpretation of where the group is going collectively: towards a more focused understanding of mainstream issues; towards a better understanding of peripheral issues; or towards a more open position with respect to new and emerging subjects.

2.a.3.) Using software for confidence-building

Empirical studies show that technology does not appear to do well in supporting tight coupling among distant team members (strong task interdependencies) but does do well in supporting loose coupling through division of labour and modularity of task assignments. Tight coupling implies time and interpersonal (face-to-face) interactions to build up the confidence needed to accept being highly dependent upon others for carrying out what one has to do. TOPICA-TECH is designing software for project management which supports the loose-coupling of DKN members.

Loose-coupling implies that all members of a DKN understand the story which is being played out on stage, that is, the tasks, the roles and the flow of activity which moves a group towards the objectives which it seeks to achieve. This understanding is achieved through task decomposition. Division of labor allows a team to divide its work into manageable chunks, set up procedures for controlling that everything is going as planned and build up trust in the ability of partners to deliver on schedule. The PIC software developed by TOPICA-TECH enacts loose-coupling in two ways.

- *Project management:* people are assigned responsibility for a given activity, for the actions that have to be taken in order to carry out the activity and for the specific operations to fulfill in connection with each action.
  - o People have to be notified of their task assignments, and they have to agree on the conditions of their involvement.
  - o People can delegate their responsibility to others but once again the designated person has to accept the task assignment
  - o Finally, the system allows for a distinction between being operationally responsible for carrying out a task and hierarchically accountable
    - ▪ We expect that observing how decisions are constantly revised in the course of a project, leading to new decisions on task decomposition, taking-on or delegating responsibility, meeting milestones and assuming hierarchical responsibility will enable us to better understand the dynamics of confidence-building in a group. The PIC system provides us with a formal framework for carrying out these observations.
- *Access to information for decision making*
  The PIC system's workflow is organized around milestones, deliverables that have to be produced at each milestone and documents which describe how the fabricated product, service or process is to be used. Evaluation of the project depends upon meeting the constraints of this highly structured workflow. The PIC system distinguishes between *two information circuits* for moving forward: the people that need to be *consulted* on the decisions that have to be taken at each step along the way; and people that are simply *informed* of the situation.

- ***Being consulted*** implies obligations to do what has to be done in order to keep the project on track, leading back to questions of assuming responsibility, delegating actions and operations, being accountable;
- ***being informed***, places the individual in a gatekeeping role, with the possibility of deciding either to do nothing and let things go on in the way foreseen, or to intervene in an attempt to open the system up to new ideas or actors.

2.a.4.) Using software for building social ties

Formal project planning procedures establish the framework for telling the story of an organization; interpersonal interactions produce the events which make the story happen. These events are linguistic in nature; language is the medium through which social ties are built. LIMSI is developing graph analytical software in order to detect semantic structures underlying text mediated interactions. It is expected that these structures will help in evaluating cooperation opportunities, identifying people to contact and questions to raise for working across borders.

The PIC software formalizes information circuits, task assignments, responsibility, delegation mechanisms and, for this reason, is extremely useful for thinking through the dynamics of confidence building in a DKN social setting. However, its starting point is the assumption that the social setting is given: that people know what they want to do, can translate their understanding of the problem into a three-tiered task division focused on activities, actions and operations, have identified the people that will be taking part in the project, have an idea of their skills and can consequently assign them responsibility, delegate jobs and anticipate upon their capacity to meet milestones. But, in fact, the formal mechanisms set up through PIC are the outcome of a great deal of time and energy expended in such things as contacting people, coordinating schedules, getting responses to questions, transmitting information, getting others to see the point, resolving conflicts, making sense out of conflicting sources of information etc. In other words, before applying the PIC formal management procedures, a lot of informal, person-to-person work has to go into building the social bonds for working together over distant locations. Many of the messages posted to Bulletin Boards or to Forums serve this purpose.

Graph theory is used to identify people or groups with some pattern of contact or interaction between them. For example, when analyzing the content of a DKN information repository, we would expect that the authors of the documents stored in that repository will share an interest in a common set of subjects. Shared interests are materialized graphically by a network of word associations that are repeatedly encountered in the different documents of the repository. Dense word association networks imply that people position themselves in much the same way using a "coded" language as a means of affirming their group identity, in the same way as there is a street language, an "in-group" language, and so on. In a DKN, creation of a "coded language" implies what in management studies is classed as *organizational closure*: despite working at a distance over the North-South divide, given the international character of science and technology, we might expect to encounter situations of organizational closure in which all parties adopt the same way of conceptualizing a problem. For example, an often cited case is when "western science" dominates to the point of making it impossible for developing countries to put their local, practical problems on the agenda of cooperative research efforts. However, another situation exists which is of particular interest to establishing a diagnostic of the degree of organizational closure in a DKN situation. Sparse word association networks (holes in the network) imply that actors' interests are heterogeneous which can have either positive (inter-disciplinary) or negative (balkanization) consequences on the group. It has been found that knowledge disseminates best in a semantic structure if satisfactory trade-offs are found to avoid intellectual in-breeding (dense networks), too great a degree of heterogeneity (sparse networks) and if weak links can be successfully exploited to promote cross-fertilization of knowledge and avoid balkanization.

2.a.5.) Using software for knowledge management

Conflict is essential for finding equilibrium between intellectual in-breeding and disruptive heterogeneity. Debate aimed at finding that equilibrium improves group performances when discussions concern such things as how tasks should be managed, the relevancy of information, appropriate frameworks for interpretation etc. However, we also know that these "positive cognitive conflicts" are often highly correlated with "negative relationship conflicts". An important goal of knowledge management is to maintain a healthy level of "cognitive conflict" while avoiding relationship conflicts. The DKN project will experiment the concept and techniques for managing "ontological disagreements" as a way of achieving this goal, using software under development by TECH-CICO at the Technological University of Troyes. Ontological disagreements can take several forms:

- At the most fundamental level, there can be disagreement about the kinds of things that (can, should) exist in an application domain ; how do we determine the limits of the subject area we are working in ?
    > We can obtain an initial answer to this question by simply listing all the objects identified as relevant for storytelling, building confidence and social ties (points 2.a.2, 2.a.3 and 2.a.4 above).
- Secondly, there can be disagreement about the way each object on the list is defined.
- Finally, there can be disagreement about the way each object is (should be) used to structure on-going interactions (establishment of hierarchical, part/whole, see also types of relationships)

The goal of Tech-CICO is to produce software for visualizing "the ontological disagreements" at each of these three levels of granularity using a modified version of the TOPIC Map representation scheme under development in the *Semantic Web* research context. Achieving this goal should help stimulate and focus "positive" cognitive conflict and debate while, at the same time, avoiding that the discussion deteriorates into relationship types of conflict.

2.b. Testing use of the DKN Platform

2.b.1.) Building the DKN testbed:

- A case study area has been defined : plant genomics is an area where cooperation exists between France and Colombia (see annex 1)
- A strategy for building an information repository has been defined based upon a snowballing technique (question 2 of the interview guide, annex 1)
    - o The VECAM software will be experimented in order to adapt the current information bins of the Document Management Software to DKN needs. These bins are currently a) studies, b) articles, c) questions for debate, d) content summaries of chosen articles or books and bibliographic references, e) institutional reading and background texts.
    - o The Forum and Bulletin Board will serve for exchanges between members of the DKN team. These exchanges have already begun and an archive exists with several documents contributed over the summer and in preparation for the September 22-23 meeting.
- The third question on the interview guide seeks to better understand the information channels used by people in the case study area to access information concerning the form of France-Colombian cooperation in the plant genomics field. Information channels can be such things as face-to-face meetings, telephone mediated communications, computer mediated communications, information based technologies (libraries, databases, documentation centres,… The goal is to evaluate what role the DKN information repository might play in promoting cooperation in the area.
- The fourth question on the interview guide seeks to better understand how to structure the content of the information repository through questions concerning how people monitor their

field of activity and identify who is doing what. Structuring content implies adopting metadata and subject classification schemes.

- The fifth and final question on the interview guide concerns incentives for getting involved in a DKN project. The goal here is to better understand why people might be willing to take part in French-Colombian scientific cooperation projects. This understanding will help enrich the process of sense-making through storytelling, because storytelling implies giving meaning to events, and this is facilitated by an understanding of the motives and intentions of people provoking those events.
- The PIC software is currently being experimented as a tool for organizing the DKN project itself. Over the next three months, it will serve team members for reporting on the progress they are making in each of their individual projects; for specifying how their work affects, or relies upon, the work of others; and for organizing the projects general workflow (milestones, division of labour, etc.)
- Finally, LIMSI and Tech-CICO will implement their classification, graph analytical and knowledge management software in order to process the information repository constructure for the plant genomics case study area.
-

2.b.2.) Experimenting the use of the DKN platform

The "**Colombia Nos Une**" Programme of the Ministry of Foreign Relations has attracted through its Network *of Colombian Students Abroad* (**Red de Estudiantes y Profesionales Colombianos en el Exterior**, http://es.groups.yahoo.com/group/redestudiantescolombianos/) around 1200 members in more than 50 countries, creating around 35 geographic and thematic nodes in less than 7 months. We will attempt to work with this network in order to experiment the use of the DKN platform in two areas: plant genomics and an area of research yet to be defined in the social sciences.

The procedure which we will adopt is that of a *task assignment*: we hope that in each of the two areas – the biological and social sciences – one or more members of the Colombian student diaspora will be willing to *become DKN user advocates*, and engage in a mobilization scenario in which:
- they take the initiative to form a team,
- build agreement among team members on an appropriate subject area for cooperative research between Colombia and France and on how to organize themselves for doing the work;
- write a 5 page team proposal setting out the scientific issues that will be addressed, why they are important, how the work will be organized and results that can be expected.

To form a team, the DKN user advocates will be asked:
- to use the DKN bulleting board in order to introduce themselves and present their project;
- to use the DKN forum as a means of actively interacting with candidates for membership in the project;
- to post the Interview guide as a device for structuring Forum discussions on how to form a team and on the information resources needed for the project;
- to file what they consider reference papers in the different information bins of the DKN-DMS.

To negotiate and build agreement on goals and working procedures, the DKN user advocate will be asked to test the usefulness of DKN software:
- for collectively constructing the meaning of on-going activity (sense-making through storytelling);
- for overcoming cognitive diversity while avoiding the dangers of organizational closure as a result of intellectual in-breeding (graph analytical techniques);
- for handling the process of knowledge creation through debate and negotiation (managing ontological disagreements).

To write the five page collective proposal, the DKN user advocate will be asked to use the PIC software.

Each of the three phases of the user study is expected to last 2 weeks. Measures of user satisfaction with the use of the DKN platform are yet to be determined.

3.) DKN project management

The DKN project is a multidisciplinary project which brings together a team made up of:

- Social scientists (William Turner, Claude Henry)
- Political scientists (Jean-Baptiste Meyer, Ximena Castro-Sardi)
- Knowledge engineers (Manuel Zacklad, Jean-Pierre Cahier)
- Cognitive scientists and computational linguists (Benoît Habert et Dominique Béroule)
- Information scientists (Mathilde de Saint Leger, Marie-Josèphe Pierrat, Fred Sultan)
- Computer scientists (Thomas Hirsch, Emmanuel Cohen and Hedi Zaher)

A division of work which corresponds to the chapter breakdown of a book is presented in what follows as a means of maximizing the benefits which can derive from the diversity of perspectives brought together in the DKN project:

Chapter 1 : (Jean-Baptiste Meyer) « Modeling the dynamics of Diaspora knowledge networks ». Prescriptive introductory chapter providing guidelines for answering the following questions : what conceptual, organizational, methodological and practical steps have to be taken to install DKN structures as channels for North-South scientific and technical cooperation?

Chapter 2: Ximena Castro-Sardi and Bill Turner, "Doing socially robust, technically reliable research in Diaspora Knowledge Networks"

This chapter will summarize the results of our case studies using the DKN plateform

Chapter 3: Claude Henry, Fred Sultan and Ximena Castro-Sardi

"Document-mediated awareness of Collective Practices"
This chapter will address the question of how documents can be used to make people geographically distant from one another aware of interaction opportunities. A comparison of "broadcast" technologies and "pull" technologies will be made and the role of "user advocates" (facilitators) will be discussed.

Chapter 4: Dominique Beroule, Mathilde de Saint Leger and Marie-Josèphe Pierrat

« User advocates as storytellers"

In order to *tell the story* of what is taking place in a DKN each document has to be made meaningful by assigning it a role in the on-going organization of collective practices. Classification and annotation techniques are being developed at LIMSI which help user advocates tell the story of collective activity by identifying documents (and consequently their authors) that:
- consolidate the mainstream activity of a social network (add to our understanding of subjects of central concern to the network)
- enlarge the scope of a subject area (add to our understanding of questions of peripheral concern to the network);
- contribute to the emergence of new subject areas.

Hypothesis: in order to transform "repositories" - document management systems and message archives produced through the use of bulletin boards and forums - into active interaction spaces, user advocates need tools for helping them tell the story of what is taking place in a social network.

Chapter 5 : Thomas Hirsch, Emmanuel Cohen, Manuel Zacklad

"Building trust in Diaspora Knowledge Networks"

Empirical studies show that technology does not appear to do well in supporting tight coupling among distant team members (strong task interdependencies) but does do well in supporting loose coupling through division of labour and modularity of task assignments.

Tight coupling implies time and interpersonal (face-to-face) interactions to build up the confidence to accept being highly dependent upon others for carrying out what one has to do.

However, loose coupling can lead to "swift trust" by respecting the following conditions:

- interact through rule-based norms (fixed milestones, workflow engagements,...)
- have moderate levels of interdependence (as defined by the RACI feature of the PIC software)
- have category based information processing routines (as developed though project decomposition and follow-up procedures in PIC

Chapter 6: Benoît Habert

"Building social ties through document mediated interactions"

LIMSI is experimenting different graph analytical techniques to detect semantic structures underlying text mediated interactions. Graph structures are characterized by zones of high density and zones of weak density (holes in the network). It remains an open question for researchers studying social networks to know if information circulates better in structures with holes (implying non-redundant information and the importance of weak links in connecting unique pieces of knowledge) or in densely interconnected networks (implying redundant information). This type of question will be addressed in this chapter.

Chapter 7: Jean-Pierre Cahier, Hedi Zaher, Manuel Zacklad

"Visualizing ontological disagreements"

Conflict is an essential part of the knowledge production process. It improves group performances when discussions concern such things as how tasks should be managed, the relevancy of information, appropriate frameworks for interpretation etc. However, these "positive cognitive conflicts" are often highly correlated with "negative relationship conflicts". An important goal of knowledge management is to maintain a healthy level of "cognitive conflict" while avoiding relationship conflicts. The DKN project will experiment the concept and techniques for managing "ontological disagreements" as a way of achieving this goal.

Chapter 8: (Jean-Baptiste Meyer, Ximena Castro-Sardi, William Turner)

"Knowledge production, competitive advantages and scientific development"

The « diaspora option » suffers from the fact that knowledge production is now part of government policy aimed at actively creating comparative competitive advantage by developing and effectively managing technological assets – including human skills. Science policy is an adjunct to innovation policy. In the DKN context, knowledge production is often defended in rather idealistic and vague terms (for decision-makers !!!) turning attention to things like sustainability of economic, social and cultural assets, or the new requirement for knowledge to be *socially robust* as well as *technically reliable*. If DKN structures are channels for cooperation of this sort, are they viable? Can the discourse be heard and understood without developing indicators for translating the argument into something

people concretely understand (competitive advantage is a powerful concept because it can be easily understood by lay people). Problems of accountability in the use of public resources have to be addressed !