SETIT 2007
4th International Conference: **S**ciences of **E**lectronic, **T**echnologies of **I**nformation and **T**elecommunications
March 25-29, 2007 – TUNISIA

Digital Identities and Management of Identifiers for a Socio-Semantic Web

L'Hédi Zaher, Aurélien Bénel, Rami El Sawda,
Jean-Pierre Cahier and Manuel Zacklad

Tech-CICO - Institut Charles Delaunay
Université de Technologie de Troyes / CNRS (FRE 2848)

zaher@utt.fr

Abstract: Our work aims at managing the actions of living agents on knowledge models. This management must be both prescriptive (access rights) and descriptive (actions history). Through the state of the art, we can see that the issue is still open, particularly for the descriptive part and when groups and knowledge are dynamic. Some of these problems seem to be solved by our model based on the metaphor of passports, visas and entry stamps.
Key words: Computer Supported Cooperative Work, Knowledge Management, Information Systems.

INTRODUCTION

We aim at building a Semantic Web in a broad sense. In other words, we aim at tightening the relations between documents and knowledge models, shared on a world wide medium. Among the related works, we can distinguish three approaches:

− The understanding of humans by computers (Artificial intelligence),

− The understanding of computers by computers (Interoperability),

− The understanding of humans by humans through computers (Computer supported cooperative work).

We coined this third approach as the "Socio-Semantic Web" [Cahier et al], and this is the research direction we try to follow. This approach is related to the trend saying that the novelty of information technology is in the change of media which summons new ways of human reasoning. As writing summons a "graphical reason" [Goody], information technology summons a "computational reason" [Bachimont]. Hence, an information system is not a "representation" but a "dynamic writing to interpret" [Bachimont].

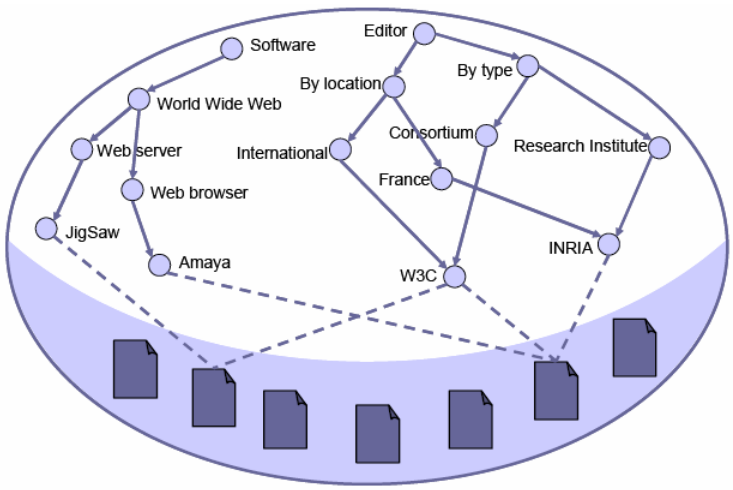

Figure 1. *Extract from a* Hypertopic *use: Multi-viewpoints modelling of the open source domain.*

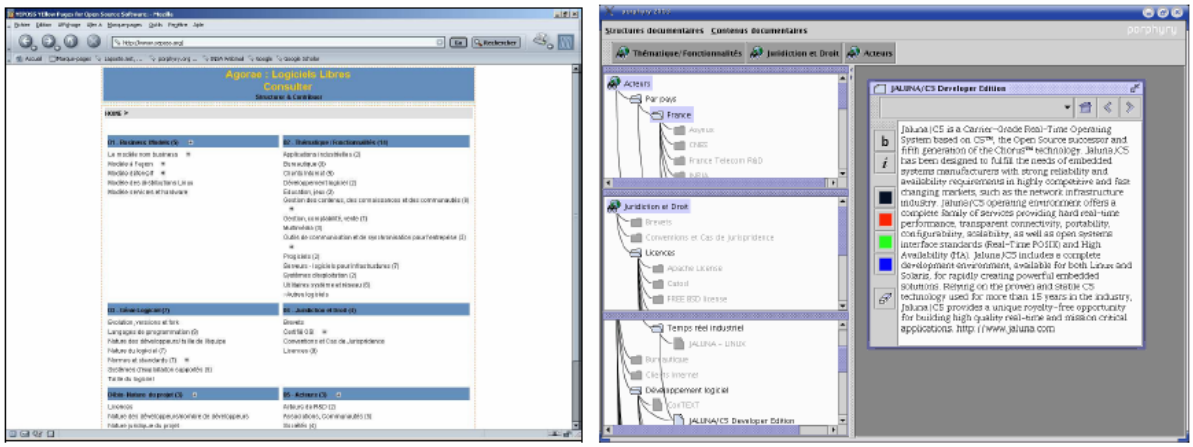

Figure 2. *Two implementations of the Hypertopic model: Agorae and Porphyry.*

We could wonder whether the different trends in the Semantic Web have consequences on the tools or only on the use. As for their inventors, the Semantic Web needs an "ontology" to be implemented. In this context, "ontology" stands for a consensual knowledge model of the domain. The problem is that in most of the domains there are several conflicting "ontologies". An explanation could be that the way knowledge is built in expert professions (Medicine, Law, Engineering, Humanities…) is not consensual but dialectic. To be called "an expert", one has to get distant from the academic knowledge (s)he learnt. Therefore, in a Socio-Semantic Web, having different viewpoints is not against the expression of semantics. On the contrary, it reveals a sense building activity.

Expressing different viewpoints is essential to the Socio-Semantic Web. To do this, the emerging community has proposed a modelling language by the name of "Hypertopic" (*cf.* Figure 1.). This language is implemented in two working prototypes: Agorae[1] [Cahier & Zacklad] and Porphyry[2] [Bénel et al, 2001] [Bénel et al, 2002] (*cf.* Figure 2.).

Experimenting on these prototypes has stressed a new problem. Because viewpoints are made by thinking agents (persons or communities), it would be probably senseless to manage viewpoints without managing those agents and their interactions with the system. The question is not really about how to model the agents' influence on knowledge[3] but what we should model.

In a way, this question has been raised for a long time about multi-user systems (either distributed or local). But did it get an answer?

1. Is it already solved?

In Internet, rigorously standardized by the IETF[4], we could think that identifiers (called identities!) management is solved by the AAA paradigm:

− *authentication* is proving "that someone (or something) is truly who (or what) they claim to be?" (e.g. by cryptography);

− *authorization* is defining "what someone (or something) is allowed - and not allowed- to do?";

− *accounting* is keeping track of what everyone (or everything) do.

However, if the authentication and the accounting were quickly controlled by standards[5], the attempts concerning the authorization[6] specified only the protocol of authorization server and not the management and the structuring of rights.

It remains to see how the various local and distributed multi-user environments tried to solve these questions.

For that, we chose a reading grid borrowing from grammarians the concepts of subject, verb and object. Although the representation modes seem to be strongly different from one environment to another, they are subsumed by a rather simple model [Shen & Dewan] in which a function associates to any triplet (subject, verb, object) a value from {granted, denied, unspecified} (*cf.* Figure 3.).

[1] Agorae demonstration: <http://www.yeposs.org/>
[2] Porphyry download: <http://www.porphyry.org/prototypes/expert/>
[3] Nested conceptual graphs, modal logics…

[4] IETF: Internet Engineering Task Force.
[5] IETF RFC 2138 and IETF RFC 2139 (Radius).
[6] RFC 3588 (Diameter).

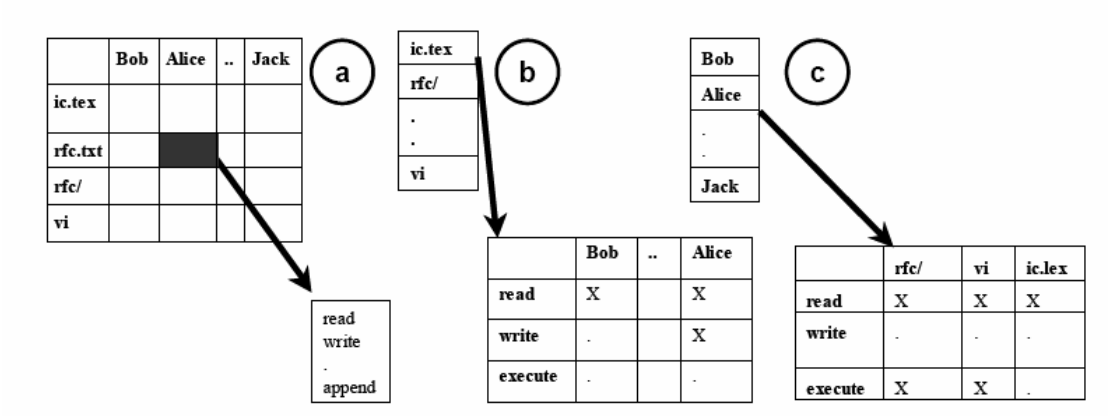

Figure 3. *Different representations of subjects, verbs and objects associations in multi-users environments.* ***a)*** *protection matrix,* ***b)*** *access control list,* ***c)*** *capacity list.*

1.1. Situational Clues

One of the most known examples of this model is given by Unix in a simplified version. For each object (*i.e.* for each file, this article will not deal with semaphores, the other object type in UNIX), only three subjects are considered (the user himself, the group and the others). The creation date is also stored for each object. One would be tempted to believe that the main clues concerning the situation are present: "Who said what? Whom? When?". The conditions seem to allow the human readers interpretation.

As the execution of the *chown* command can change the file "user" (if it is carried out by the starting user), it is impossible to consider the "user" as the "author of a document". It would be a document falsification. This indicates that the user concept is related to the object owner and not to the object author.

This example, still simple and self-evident, brings heavy consequences. Subjects' management in Unix systems, and generally in computer systems, assumes and adopts a prescriptive goal. It is stuck to protect privacy and to avoid malicious acts and errors. We can oppose to this prescriptive goal a descriptive one that tries to ensure objects authenticity (in a philological meaning).

Computer systems are often claimed to bring trust, mutual awareness and filtering on digital objects. But how could we ensure all of these without knowing a bit more about the subjects who made them?

Even if the problem of describing subject-verb-object relations were solved, the issue of structuring subjects and verbs would remain outstanding.

1.2. Structuring Subjects

In multi-user environments, the first attempts at structuring the subjects aimed at managing generic subjects in order to simplify access control. Those generic subjects are called "roles" or "fields". In today's content management systems (CMS), it is

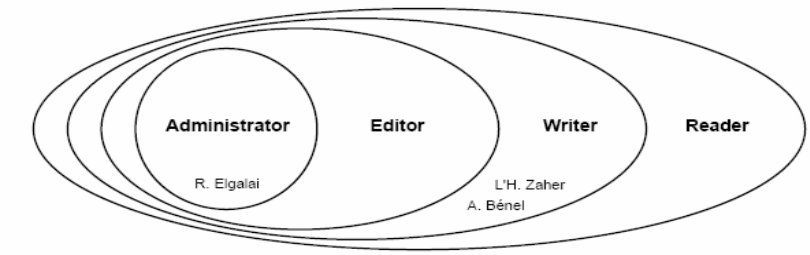

Figure 4. *Using "roles" (Multics rings) to structure subjects.*

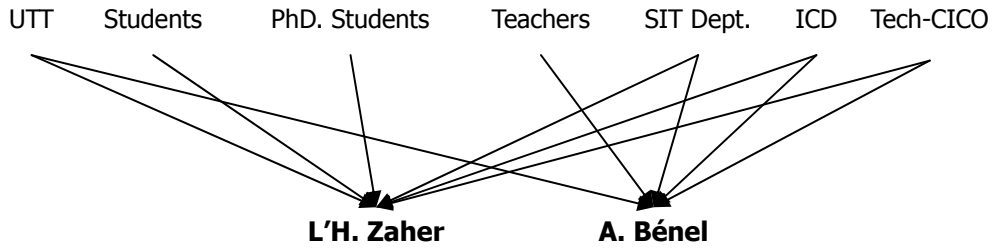

Figure 5. *Using Unix group to structure subjects.*

usual to see roles like "reader", "writer", "editor", "manager" (*cf.* Figure 4). Like the eight Multics "rings", roles are defined as access rights sets, which overlap from the most permissive to the strictest [Schroeder & Saltzer]. In several systems, when a user fears to do errors, he may change to a stricter role (on the fly or at logging time).

The principal defect of this approach is to structure subjects as if we structured verbs on objects. Not only subject, verb and object cannot be considered anymore as independent dimensions, but also, we get a cyclic definition of the access rights.

Another curious organization of the subjects is given by the dichotomy root / user on Unix. The conscientious administrator of a Unix system will have to change identity according to the task to be carried out. Moreover, two system administrators will share the same root account. This quasi-schizoid behaviour could be explained by the ambiguity between structuring subjects and structuring verbs or objects. Besides, this ambiguity is present on the first reference on the tree-like file systems [Daley & Neumann] where authors wrote about a file organization in directories (yellow pages of people) and not in folders (documents container). Indeed, the first structured file system had on only one level which distinguished private spaces from public space [Tannenbaum]. Just as the term "directory" is still used, confusion between the structuring of the objects and that of the subjects seems to remain.

A more interesting way to structure subjects is based on user groups. Each user may belong to several groups. A rather awkward illustration is given by Unix. But the lack of structure among groups is really a problem. It is impossible for instance to tell, once and for all, that every member of the "Tech-CICO team" is member of the "ICD" (*cf.* Figure 5.).

As for the naming model of LDAP8 (RFC 2253), it preserves the users' multi-membership principle while structuring the groups in a tree (*cf.* Figure. 6). However, several problems remain. First of all, the tree structure prevents from defining a group in the intersection of two others. For instance, it will not be possible to represent the fact that the ICD Lab is included both in the CNRS Institution and in the UTT University.

Moreover, for architectural reasons, the object modelling the user cannot be referred in another directory. For instance, if the user "L'Hédi Zaher" is registered in UTT's LDAP[7] directory and is the CNRS one, it will be modelled by two completely distinct and independent objects. In the same example, it would be also impossible to search for the "Tech-CICO PhD students". Let us note that the interconnection of directories, even if it were solved from the technological point of view, would not be without posing ethical and legal problems (cf. the French law on "Data processing and Freedom").

1.3. Structuring verbs

Many are the multi-user environments for which the structuring of verbs is based on the Unix composition of the elementary verbs: "to read", "to write", "to execute". Despite they look like Turing's machine operations, are we sure these verbs are atomic? Is their list exhaustive? In fact, this list existed in Multics. At the time, the list contained a fourth verb: "to append", i.e. "to write at the end of the file". In other words, this verb is now implicitly included in the verb "to write"… just as the verb "to delete". The expressivity of the model is consequently extremely reduced. It is undoubtedly better, as in a certain number of FTP servers, to assign rights on each command which the protocol offers.

Another reproach with the verbs structuring model in Unix is to use the same elementary verbs for directories and files. It is counter-intuitive to say that executing a directory stands for opening it.

Moreover, we have to keep the same number of actions for different types of object. It worth noting the object oriented solution which Zope offers. The constructor of each class of object defines methods

[7] LDAP: Lightweight Directory Access Protocol.

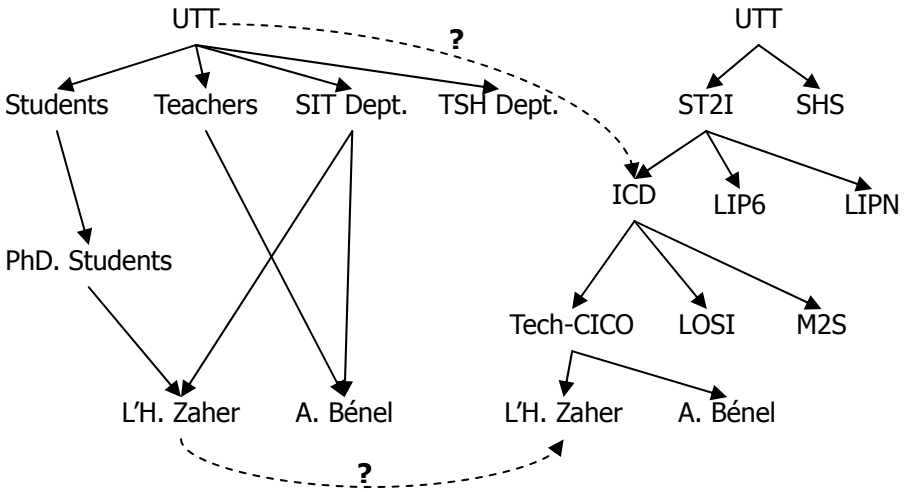

Figure 6. *Structuring persons and groups with LDAP directory.*

exposed to users, corresponding to potentially allowed elementary operations. These verbs depend then on the type of object and are completely customisable (for instance within the framework of a workflow: to validate, to sign, to accept, to increment).

2. What would we need?

2.1. Identities Building

In the preceding section, we did a survey of identifiers' management in multi-user environments and highlighted how they mismatched with the needs of cooperating human beings. This can be explained by the inadequacy of Smith's work model to the mental activities of sense making [Zacklad].

Indeed, instead of being based on a technical division of work as a productivity guarantee, the mental activity seems to require a kind of redundancy of the actors, cooperation, and debate. Such activities also require redefining the concept of "problem solving". Indeed, the problem situation, its statement, and even the problem itself are not any more given "data", but are built progressively. In particular, mental activities do not only produce a deliverable but also forge the identities of people and the involved communities. Such framework generates instability. On the one hand, each person belongs systematically at several (conflicting) communities. On the other hand, the person permanently seeks unity. From this instability, emerges a constant evolution of personal and community identities (fusion, articulation, negotiation, alliance).

2.2. Redefining the "role" notion

Some current works also propose for information technologies a concept of "role" which is more sociological than the traditional set of access rights. For the team of Thomas Hermann, four characteristics must be taken into account [Hermann et al]:

− the *position* in the hierarchy, in a logic of differentiation and responsibility (however the rules and the memberships are dynamic),

− the *function*, because the hierarchy is always adapted according to the circumstances (more or less unconsciously),

− the *behaviour* corresponding to "the role interpretation" during the task,

− the *sanction* of a variation between the observed behaviour and the behaviour awaited by the group. The sanction can be negative (exclusion...) or positive (momentary acceptance, recognition of the new role).

3. What do we propose?

In the preceding section, we saw a few sociological notions about identity management. We do not want to reduce these complex notions to software artefacts, but they can be an inspiring conceptual frame.

In order to build a more precise model, we will draw a metaphor to a real world organizational object which is known to work with many actors and for a long time: the passport. Contrary to its recent use by a well-known software editor, this metaphor might bring a real conceptual shift in identifier management (*cf.* Figure 7).

Then, we shall list the main characteristics of our "passport model" and illustrate each of them with an example from a university information system.

As an example, some people from our University need to access to the knowledge models of a workshop and a project. Their "passport" is issued by the

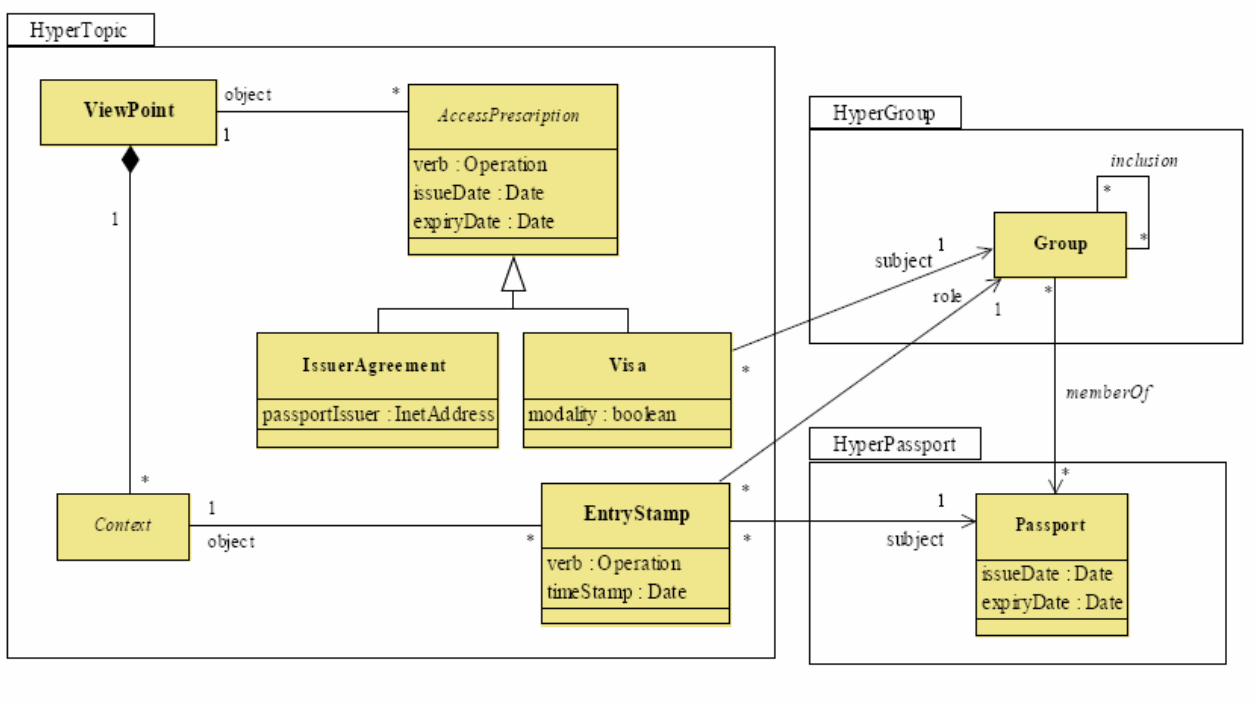

Figure 7. *Identities management model based on the passport metaphor [UML Class diagram].*

university administration, in order to be used by the managers of the knowledge models.

A passport is made by an authority for other ones. As an example, some people from our University need to access to the knowledge models of a workshop and a project. Their "passport" is issued by the university administration, in order to be used by the managers of the knowledge models (*cf.* Figure 8).

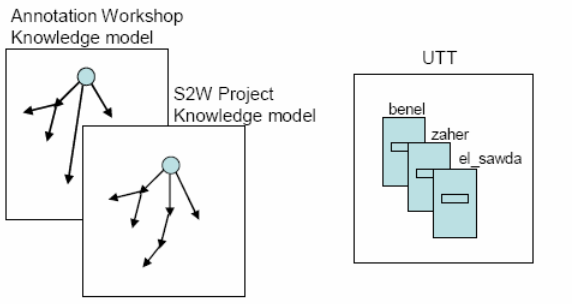

Figure 8. *A passport is made by an authority for other ones.*

A person can have different passports from different emitting authorities or (in very special cases) depending on the receiving authorities (because some authorities trust some authorities more than others). As an example, we can have a job account and a few personal accounts in order not to get disturbed at work, not to use a server for something it is not supposed to be used for, and to preserve our private life (*cf.* Figure 9).

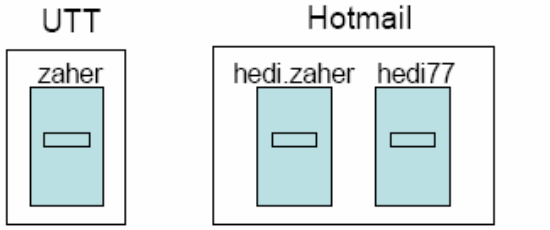

Figure 9. *A person can have different passports from different emitting authorities*

A passport is a temporary mark of trust. It must be regularly renewed (so that it cannot be illegally used by others for a long time and because there can be no mutual trust if there is no possibility of sanction). For example a university can have confidence in a student while he is in this university because the administration knows that in case of any fault he commits, it will be able to take disciplinary measures but in the case of a student who has finished his studies, it will not be possible to sanction him. To avoid this kind of problem, his account will not be renewed if he is not going to be registered at the university again cf. (Figure 10).

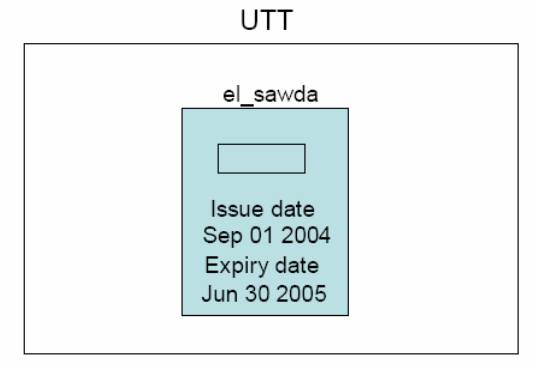

Figure10. *A passport is a temporary mark of trust.*

Sometimes passports are sufficient (depending on the border and on the issuer of the passport). But in this case, controls and negative sanctions could be carried out later. For instance, the university gives on-line access to some expensive magazines to all its members (students and staff). This means that the only fact to have a passport from the university gives the access to read those magazines (*cf.* Figure 11).

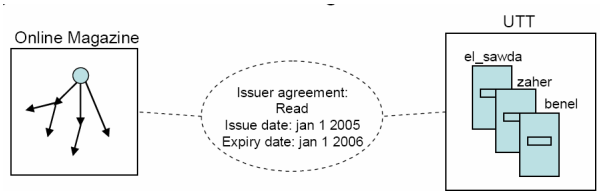

Figure 11. *Passports May be sufficient.*

People can travel with several groups. Groups can be built by the members themselves (not by the country they came from). For example, groups can be built freely in the university (independently from the administration and from one another) (*cf.* Figure 12).

When a visa is given to a group, the issuing authority accepts the group as listed. For instance, a teacher can give access to his (her) handouts by setting a visa for a group defined by the doctoral school (*cf.* Figure 13).

The visa is a temporary mark of trust. As an example, exercises solutions should not be available from a year to another, but only when the students have already tried to make the exercises alone (*cf.* Figure 14).

The history of visas is kept. By reading this history, we could learn for example that students have been lately authorized to read the administrative reports, and even that students representative are now involved in their writing (*cf.* Figure 15).

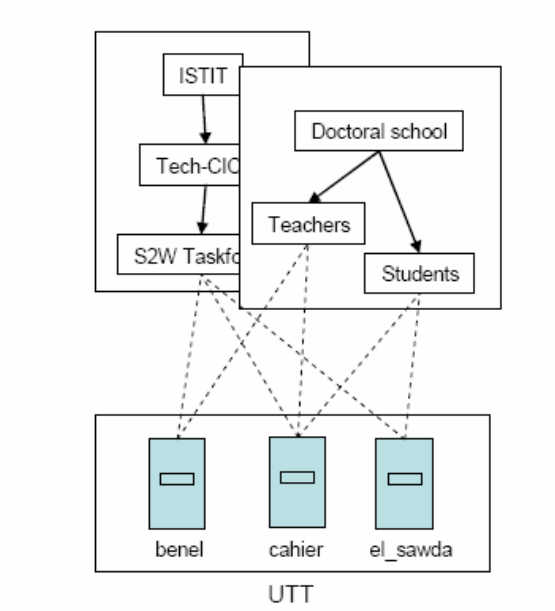

Figure 12. *Groups can be built by members themselves.*

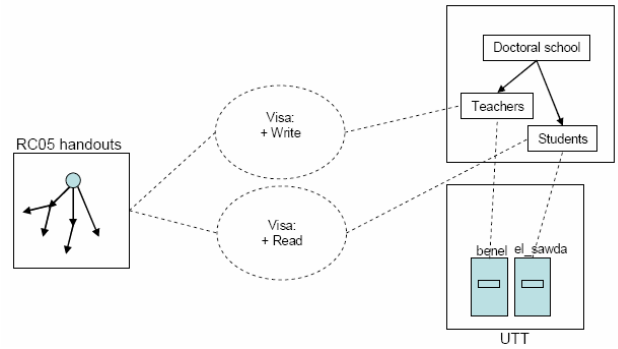

Figure 13. *When a visa is given to a group, the issuing authority accepts the group as listed.*

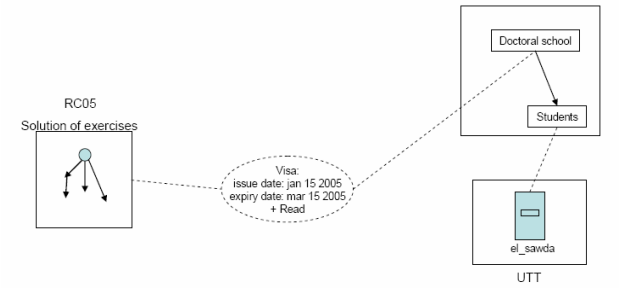

Figure 14. *The visa is a temporary mark of trust.*

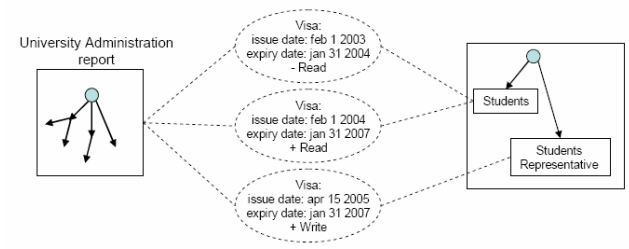

Figure 15. *The history of visas is kept.*

An entry stamp describes the actual accesses. On an entry stamp the data are far more precise than on a visa (which border, which day, which person, in which group). The history of entry stamps is kept. In our example, we can see that someone has read the same part of the handouts twice, and with different roles. If this person had not got a visa (his issuer has an agreement), the writers of the handouts would have been notified and they could have given a visa (positive or a negative) to him (*cf.* Figure 16).

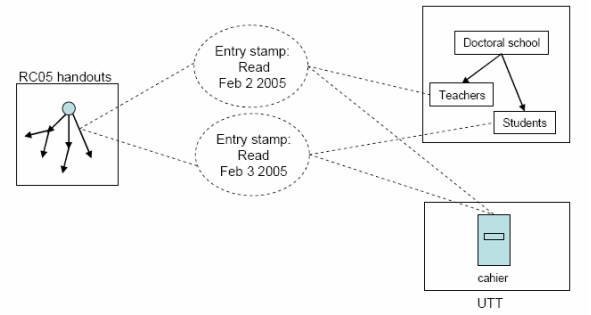

Figure 16. *An entry stamp describes the actual accesses.*

4. Conclusion

This paper dealt with managing the actions of living agents on knowledge models. In a way, this question has been raised for a long time in multi-user systems, but, as we saw through a state of the art, the question is still open.

First, the situational clues which are needed to understand the objects seem to be most of the time unusable. Secondly, the structuring of subjects is far from being adequate. To understand this last point, we focused on a few sociological theories saying that the structuring of subjects should have been both dynamic and multiple.

Lastly, we proposed a model based on the metaphor of passports, visas and entry stamps. Some of the problems seem to be solved by the model, but some other remains.

For example, we store the history of objects with regards to subjects, but, in order to avoid an infinite regression, we do not store the history of subjects. Will it be enough? It would have been also interesting to have an identity delegation feature, but we still do not have a model for this. Last but not least, if we can really record the history of every operation done in the socio-semantic web, we will need powerful visualization techniques soon to get synthetic view of this history.

REFERENCES

BACHIMONT B. « L'intelligence artificielle comme écriture dynamique : De la raison graphique à la raison computationnelle ». In PETITOT J. (ed), *Au nom du sens*. p. 290-319. Paris : Grasset, 1999.

BENEL A., EGYED-ZSIGMOND E., PRIE Y. CALABRETTO S. & MILLE A."Truth in the Digital Library: From Ontological to Hermeneutical Systems". In Proceedings of the fifth *European Conference on Research and Advanced Technology for Digital Libraries*, Darmstadt, 2001. Lecture Notes in Computer Science #2163. p. 366-377. Berlin, Springer-Verlag, 2001.

BENEL A., CALABRETTO S., IACOVELLA A. & PINON J.-M. "Porphyry 2001: Semantics for scholarly publications retrieval". In Proceedings of the thirteenth *International Symposium on Methodologies for Intelligent Systems*, Lyon, June 26-29, 2002. Lecture Notes in Artificial Intelligence #2366. p. 351-361. Berlin, Springer-Verlag. , 2002.

CAHIER J.-P. & ZACKLAD M. "Towards a Knowledge-Based Marketplace model for cooperation between agents". In Proceedings of *COOP'2002 Conference*, St Raphael. Amsterdam, IOS Press, 2002.

CAHIER J.-P., ZACKLAD M., MONCEAUX A., « Une application du web socio sémantique à la définition d'un annuaire métier en ingénierie ». In JAULENT M.-Ch. (ed), Actes de la conférence *Ingénierie des Connaissances*, Lyon, 2004.

DALEY R. C. & NEUMANN P. G. "A general-purpose file system for secondary storage". In Proceedings of the *AFIPS Fall Joint Computer Conference*, p. 213-229. New York, Spartan Books, 1965.

GOODY J. The Logic of Writing and the Organization of Society. Cambridge : Cambridge University Press, 1986.

HERMANN T., JAHNKE I. & LOSER K.-U. "The role concept as a basis for designing community systems". In F. DARSES, R. DIENG, C. SIMONE & M. ZACKLAD (Eds), Proceedings of *Cooperative Systems Design: Scenario-based Design of Collaborative Systems*. p. 163-178. Amsterdam, IOS Press, 2004.

LEONHARDT J.-L. « Note de lecture : Léviathan, traité de la matière, de la forme et du pouvoir de la république ecclésiastique et civile ». Rapport de recherche, Maison de l'Orient et de la Méditerranée CNRS, Lyon. 26 p. To be published in « *Essai sur les modèles de la raison de*

l'homme (de science) », 2004.

LEWKOWICZ M., MARCOCCIA M. The participative framework as a design model for newsgroups: PartRoOM In F. DARSES, R. DIENG, C. SIMONE & M. ZACKLAD (Eds), Proceedings of *Cooperative Systems Design: Scenario-based Design of Collaborative Systems*. p.243-257. Amsterdam, IOS Press, 2004.

LONGCHAMP J. *Le travail coopératif et ses technologies.* 319 p. Paris: Hermès Science, 2004.

SHEN H., DEWAN P. Access control for collaborative environments. In Proceedings of *the ACM Conference on Computer Supported Cooperative Works*. p. 51-58. New York, ACM Press, 1992.

SCHROEDER M. D. & SALTZER J. H. A hardware architecture for implementing protection rings. *Communications of the ACM*. Vol. 15, Number 3. p. 157-170. New York, ACM Press, 1972.

TANNENBAUM A. S. *Operating Systems, Design and Implementation.* London, Prentice Hall International, 1992.

ZACKLAD M. Documents for Action (DofA): Infrastructures for Distributed Collective Practices, In *Distributed Collective Practice: Building new Directions for Infrastructural Studies (Workshop of the CSCW 2004 conference)*, 6 p, Chicago, 2004.