The Role of Activities Awareness in Faceted Classification Development

Orélie Desfriches Doria

Laboratoire DICEN, Case courrier 1D6R10, Conservatoire National des Arts et Métiers, 2, rue Conté, 75003 Paris, <orelie.desfriches_doria@cnam.fr>

Orélie Desfriches Doria is a student, preparing her last year of PhD work in library and information sciences in the Conservatoire National des Arts et Métiers (CNAM) France (Paris), and is Attaché temporaire d'enseignement et de recherche (ATER) of the Research Laboratory DICEN (Information and Communication Systems in Digital Age). Her research interests focus on KOS's; uses of KOS's, especially faceted classification; methodologies for faceted classification; information organization activities in organizations; and differing points of view about documents, folksonomies, document typologies, KM, the semantic Web, and the socio-semantic Web.

Doria, Orélie Desfriches. **The Role of Activities Awareness in Faceted Classification Development.** *Knowledge Organization. 39(4)*, 283-291. 21 references.

ABSTRACT: In this paper, we propose a part of the methodological work to accompanying the development of a new type of Knowledge Organization System (KOS) based on faceted classification. Our approach to faceted classification differs from its traditional use. We develop a theoretical typology of professional documents based on their uses. Then we correlate these types of documents to specific types of KOS according to their degree of structural constraint and activities they aim to serve.

Received 7 March 2012; Accepted 7 March 2012

1.0 Introduction

The evolution of work organization models, characterized by an intensification of distant exchanges, the increasing number of coordination and communication tools and of sharing, transmission and back-up systems, results in complex informational environments. In the framework of an ANR[1] project, a new type of Knowledge Organization System (KOS) based on faceted classification is under development, aiming to reduce the cognitive cost of information management tasks in complex digital environments, particularly in working documents management. We are working on a methodology to accompany its deployment and to elaborate relevant facets relating to different trades. In this article, we present a part of this work.

The starting point of this study consists of observations on individual folder organization of documents taken from individual work stations of different research engineers who work in the R&D department of an industrial group. We focus our attention on a particular case of our work-in-progress methodology, concerning the elaboration of facets dealing with document types information, which brings up specific problems. After the development of an empirical typology of observed document types, we propose another theoretical typology to allow the management of document type information. This type of information is essential, yet difficult to process autonomously. Not being of a universal nature, the document type instead aims at representing the different terms of the type according to the context. Hence the document type in a faceted classification is considered a necessary component of document management, whose meaning, through combination with other facets, is rendered unambiguous.

In this article, the theoretical typology we present is established according to document characteristics such as usage, defined as groups they are included in, and for which they represent a support for interactions, and activities for or during which documents

are created. Finally, we propose recommendations to correlate KOS, document uses, and documentarization operations purposes.

2.0 Documentarization and heterogeneous knowledge organization systems (KOS's)

KOS's refer to "controlled languages, classification schemes, and to knowledge representation languages from Artificial Intelligence" (Zacklad 2011). In this tool category, Zacklad also includes search engines' indexes. These KOS's consist of systems of access to information, knowledge tracking, representation, and filtering systems such as thesauri, classifications, ontologies, and tag clouds. They are most frequently used for documentarization operations on documents which consist of "transcribing or recording a semiotic product on a perennial substrate, which is endowed with specific attributes intended to facilitate the practices associated with its subsequent utilization in the framework of distributed communicational transactions" (Zacklad 2006).

Documentarization is a major issue in knowledge preservation and communication by allowing "(i) to manage them along with other substrates, (ii) to handle them physically, which is a prerequisite to be able to browse semantically among the semiotic content, and lastly, (iii) to guide the recipients" (Zacklad 2006). The stored information related to the documentarization process on a technical level (content), organizational level (coordination), or location aspects (access to documents) accounts for a substantial effort that KOS endorses (Pikas 2007). In addition, we notice that different activities during the trade exercise lead to the production of distinct document types, which are not documentarized with aid of the same KOS. Despite their diversity, the latter differ in structural aspects, and also in content aspects (vocabulary, semantic), though we state that KOS's present in organizations and their structuring should be correlated to document uses and to the purposes of documentarization operations.

In most organizations, we find frames of reference that define the location where a document should be recorded according to different intentions (record management, sharing, individual use) and to document features (state of document, life cycle, department) to limit informational entropy by controlling document management. The various storage media used according to document features may present heterogeneous KOS and interfaces. Their use appears as an additional cognitive cost regarding those coming from a professional exercise in which the main activity does not consist of information management. (Desfriches Doria and Zacklad 2010).

In fact, KOS diversity, variability of storage media, of activities associated with documents and of document types make document management activities complex. Our findings differ from previous work on typologies by the scope of documents we deal with, in contrast with Zeller (2004), who is interested in all document forms (DTB, Web sites, GIS, multimedia documents, etc.), or to Gagnon Arguin (1998), who focuses her interest on proof documents for record management, or to Alberts (2009), whose work is concentrated on mail and is exploring document gender notion. We limit our studies to digital working documents that we define as individually or collectively produced or handled documents during professional exercise of various trades. The purpose of our approach does not consist of record management, but is more focused on working documents management in a knowledge management perspective.

3.0 Faceted classification

Faceted classification is represented "as a combination of complementary conceptual groups offering the ability to insert varying analysis dimensions on informational objects, to characterize and make access to information easier by offering multiple ways of navigation towards any document" (Mas et al. 2008). The notion of facet often appears as "the most consequent theoretical contribution of the century in information sciences" (Maniez 1999). Faceted classification presents a number of benefits reported in literature. The most common benefits mentioned are expressiveness, flexibility, consistency, and adaptability (Maniez 1999; Ali and Du 2004; Marleau et al. 2008). It has also been recognized by Broughton (2005) to support browsing, navigating, and information researching. This author explains that faceted classification allows browsing (which consists of quickly scanning a corpus to discover its content), thanks to its logical structure and its capacity to express complex or compound subjects. Its structure, which can be combined with user interfaces and multiple access points, enable navigation through a corpus. Finally, information research is supported by progressive filtering based on multiple search criteria (facets) (Broughton 2006), though, according to Kwasnik (1999), one must not overlook the difficulties related to establishing relevant facets, the potential incoherence in inter-facet relations, and in the

visualization of the classification scheme with regard to the internal logic of each individual facet.

3.1 A more flexible approach to faceted classification

Faceted classification is traditionally used, in a formal way, to standardize homogeneous corpus management. Homogeneity is employed here for document types, but also for content aspects. For example, in libraries, document types are almost similar, and contents are described through standards fields as keywords for book subjects. The level of specificity of indexing is established.

By contrast, in working document management, the corpus is heterogeneous in terms of form and amount. We notice that the level of specificity can vary according to specific needs, activities, and amount of produced documents. The content is not necessarily the major indexing requirement; we also meet some specific needs for describing the situation of document creation, like time related information. A study from Pikas (2007) about engineers' Personal Information Management practices reveals that they do not use the same strategies to retrieve their documents, nor do they remember the same kind of information. This study claims that the most important element while searching for a document is the time dimension, which can be conveyed with differing instances (season, precise date, period, project stage, etc.). The development of relevant facets and of the required level of specificity for the documentarization process is defined in context and in relation to activities, users habits, and volume of produced documents. We don't recommend any scale, as far as these are principles to be applied in reference to a corpus, a set of activities, a department, or a professional group. Thus the application of the principles of faceted classification in the face of the large diversity of working documents forces us to soften the principle of facets and leads us to reflect on the development of more coherent schemes adapted to diverse situations and actors.

The faceted KOS we develop allows personal faceted classification schemes, without restricting eventual constrained aspects of document description. It emphasizes the flexibility and expressiveness qualities of faceted classification and this way of using it appears as a less strict approach of faceted classification than the traditional ones. From our point of view, users or document creators are considered the most relevant people to index their documents, thus we are developing this methodology for designing faceted classification adapted to all contexts in organizations. In our preliminary study, we notice that documents do not imply identical uses according to different trades, it is not therefore necessary for them to be described in the same terms, in a constrained way by all actors in an organization.

Consequently, our approach to faceted classification which allows it to be fed and developed on the fly, is bottom-up. We can compare it to Vickery's opposition (1960) to mechanical and constrained implementation of fundamental categories to a subject. These categories should be used as a guide for suggesting potential characteristics that should not be ignored. (La Barre 2010).

4.0 Proposal of empirical and theoretical document typologies

Handling questions about document types leads us to focus our interest on the notion of facet and to confront problems mentioned before by Kwasnik (1999). The choice of relevant facets and the necessity of consistency between facets are influenced by more ancient techniques such as development of lists, taxonomies, or typologies. By typology, we mean analysis and description of typical forms of a complex reality, allowing classification. For our concerns, we need to find division criteria, or dimensions of analysis, from which we can develop a description of empirical complex data, to eventually transfer it to the development of our faceted classification.

4.1 Empirical typology of documents

The theoretical typology of documents presented in part 4.2 represent a means to avoid an increasing number of document types in faceted classification. In fact, during a deep study of folder organization on individual workstation of two research engineers from the R&D department of an industrial group, we noted more than 110 document types which make up our empirical typology. The latter already constitutes a reduction in the actual complexity of observations (Coenen-Huther 2007), given that we found several occurrences of the same document type in folder hierarchies due to the fact that workers are involved in several projects simultaneously with varying roles according to the project.

This empirical typology corresponds to the systematic listing of instances of document types, which we have reduced to a simplified form.

Documentation	Balance Sheet	Case scenario
Course	Export	Need analysis
Article	Spécifications	Quality follow-up
Visual guidelines	Requirements	Review
Methodology	Trade proposal	settings file
Thematic bibliography	Test report	Data model
Transverse bibliography	Appendix	Process description
Benchmark	Algorithm test	Synthesis note
State-of-the-Art-review	Data retrieval	Working document
Administrative document	Notes	Preleminary study
Interview	Thesis	Transverse document
Study	Letter	Outline note
Screen shot	Tutorials	Mock-up
Report	Opportunity study	Consortium agreement
Presentation	Opportunity Note	Running document
General bibliography	Need assesment study	Dashboard
Contribution	Form	Inspect form
Training	Standard	Contract
Budget	Input Data	

Table 1. Extract from the empirical typology

The figure (Table 1) presented above is an extract from the empirical typology. We can hardly accomodate 110 values for a facet with our purpose of reducing cognitive costs of information management tasks, thus we have focused our interest on other dimensions of analysis, such as document usage.

4.2 Theoretical typology functions of documents uses

Our theoretical typology is developed from the viewpoint of document usage, which depends, according to us, on groups involved in creation or utilization of these documents and on the purpose of a worker's activity considered in its entirety and to be seen in the global organization.

In the following table (Table 2), purposes are mentioned in the frame of document creation, as our goal is to enable document management rather than record management.

According to Marradi (1990), this typology, which could also be qualified as an extensional classification, originates with an item set (the document types mentioned in the empirical typology), on which we apply division criteria (purpose of activity and groups types interacting with documents). These criteria are applied to items on the base of property similarity in the item set. Thereby, empirical document types owning the same properties are grouped in a new theoretical and more abstract type.

It can be useful to notice that this typology can potentially be applied to all departments of an organization. For instance, in a Human Resource department, the purpose of the activity labeled as "accomplishment of mission in the frame of projects" can lead to pro-

Work Activities Purposes						
Types of Groups		**Individual activities**	**Execution of mission in frame of Projects**	**Coordination**	**Contribution to the Trade**	**Record Management**
Individual Work		Individual work document	Individual work document	/	Individual work document	Individual work document (Final version)
Collective Work	Project		Document of collaborative work	Project monitoring document	Auxiliary resource document	Document of collaborative work (Final version)
	Trade		Auxiliary resource document	/	Trade document	Trade document (Final version)
	Corporate	/	Auxiliary resource document (Final version)		Trade document (Final version)	Referential document

Table 2. Document typology functions of professional activities purposes and types of groups

duction of document types as "contracts." These documents will be considered, in the context of this activity, as the type "Document of collaborative work," but could also belong to the category "referential document" from the viewpoint of people from other departments of the organization. The types of groups mentioned in this theoretical typology come from the approach of Zacklad (2007). We assume that this theoretical typology can compose an adding marker for users in the stage of developing document typologies for creation of faceted classification. This can also eventually be a classification principle for consistent faceted organization within a trade.

4.3 Definitions of types from the theoretical typology

In this typology, types are nor definitive nor exclusive. For instance, a document can move from a type "Document of collaborative work" over to an official version for record, and another document of type "Document of collaborative work," like a data model, can become a "Trade Document" in other situations.

Individual work document: These documents correspond to an individual work activity, aside from any work group, or for documents created in autonomous ways, for preparing to share with a working group. For example: notes, diagram

Document of collaborative work: These documents are written collaboratively, within a group where the work of individuals is highly dependent of other workers' work, as is frequently the case in project organization. For example: State-of-the-Art-review, requirement specifications

Project monitoring document: This type of document is used to organize activities within projects over time and organizational aspects (coordination). For example: dashboard, schedule

Trade document: These documents are collaboratively written by community of practice members or professionals from the same trade and are used, individually or collectively, for professional activities; they can describe good practices; the type of discourse is often prescriptive. For example: procedure, operating procedure, good practice guide, recommendations

Auxiliary resource document: These documents are often completed, reused in other departments of organizations, or for other projects, or by other professional groups than those who produced them. They are about knowledge capitalization. They are taken up and undergo a revival of interest for activities or projects other than those during which they have been produced. They could be compared to documentation, but we distinguish them because they are internally produced. For example: maintenance documents reused in context of repairing.

Referential document: These documents can be equally accessed by all organization members. They consist of document models, formal descriptions of projects, forms, documents of activities planning of a department. They are not specialized on professional activities. For example: instructions about record management or data backup, consortium agreement, contract, organization chart, visual guidelines

External documents: In every dimension of professional exercise and in almost every case of document production, workers need some documentation. These documents come from information research activities, from external sources.

Record document: Final versions, official versions of individual work documents, documents of collaborative work, trade documents. For example: Deliverable

The following categories of Individual work documents, Documents of Collaborative Work and Trade documents can belong to a broader category from Zacklad (2006), labeled as DofA (Document For Action). These DofA are characterized by their extended state of incompletion, their perenniality, their fragmentation, their rapid circulation, by the fact that they are produced by different authors and by the non-trivial argumentative relationships between the document fragments. (Zacklad 2006). For Zacklad, DofA corresponds to various devices: textual file or annotated drawings, forum systems, blog systems or wikis, messaging systems, etc. (Zacklad 2007), while we are only focused on working documents in the frame of professional exercise.

4.4. Evaluation by reclassifying empirical types in theoretical types

To test the theoretical typology based on document uses presented above by reclassifying all empirical types inside the theoretical types, a large amount of document appears to fit in the category of Document of Collaborative Work (40 instances) while amount of documents in other categories are manageable for taxonomies that may become facet values (about 12 values by other theoretical types).

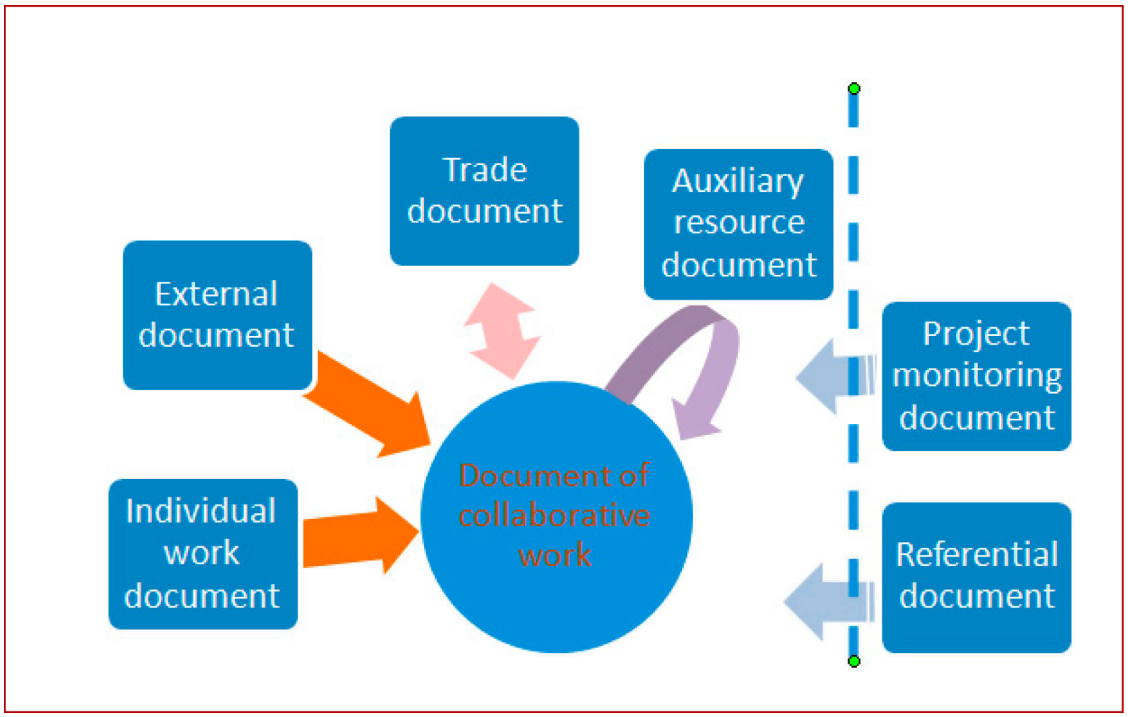

Figure 1. Major category of document types: Documents of Collaborative Work

The diagram (Figure 1) presented above illustrates that the core documents produced or handled comprises the category of Document of collaborative work. In fact, the Individual Work documents and External documents often contribute to the drafting of Document of Collaborative Work. Trade documents also frequently appear as contributions to this type of document and vice versa. Auxiliary Resource documents generally come from the category of Document of Collaborative Work and also become resources for the drafting of this latter type of document. Lastly, Project Monitoring documents and Referential documents are used to organize the drafting activities of Document of Collaborative Work. Thus, it is not surprising to note that this category gathers the most important empirical types.

4.5. Refinement with activities

One facet containing 40 values is not manageable. It appears necessary to apply a new categorization criteria. We chose the activity element, in which the specificity level can vary in terms of functions of needs, numbers of documents, and degrees of precision needed. Our tool allows the creation of activity contexts for grouping facets with relevance. This enables documentarization with an adaptable level of specificity functions for user needs, in which the functions of the prevalence of certain activities within a trade vary.

If we develop a faceted classification with activity-based contexts, we may find a facet in each context representing specific document types frequently produced during each activity. Thus we can detail document types comprising the Document of Collaborative Work category.

As observed, activities within our KOS have several roles. First, they are a means of grouping facets in a relevant context. Second, they improve information allocation in facets when the number of values is too high by refining the facets' content, while maintaining consistency in the classification scheme.

The table (Table 3) proposed below is an extraction of reclassifying operations of the Document of collaborative work category functions of specific activities. According to this example, we notice that an acceptable amount of values of facets is created in reference to specific activities. For a facet concerning the preliminary studies documents, the label could be "Preliminary Studies Specific documents." The executed choice consists of fragmenting document types in reference to activities during which they are produced.

Theoretical typology of documents	*SPECIFIC ACTIVITy*	Empirical typology of documents
Documents of collaborative work	*Design*	Requierements
		Spécification synthesis
		specification appendix
		Requierements review
		Results export
	Test	Results sheet
		Test analysis
		Diagnosis guide
		Validation interview
		Test Report
		Algorithm test
		Setting file
		Log analysis
		Case scenario
	Collection of Input Data	Data Model
		Data retrieval
		Important Input data
		Interview
	Upstream Activities	Interview review
		Interview plan
		Observation
		Observation report
		Benchmark
	Preleminary studies	State-of-the-Art-review
		Need assesment study
		Need analysis
		Preleminary study
		Process description
		Opportunity study
		Architecture study
		Quality follow-up
		Meeting review

Table 3. Extract from reclassifying one theoretical type of document by specific activities

5.0 Recommendations for types of KOS according to document types and management of information activities' purposes

As mentioned above, we recommend that KOS's used in organizations and their degree of structural constraint should be correlated to document uses and to documentarization operations' purposes. Management of information activities and especially for documentarization can be enumerated in a broad outline as follows: applying indexing instructions for record keeping with formal KOS's, systematic and scalable working documents organization with medium formalized KOS, and tagging of individual work documents through informal KOS. The degree of structural constraint of KOS is related, itself, to document types that are possibly documentarized with this KOS, and storage medias are associated to these features.

We propose to make some recommendations about KOS types functions of theoretical document types and documentarization operations' purposes. In the table (Table 4) below, KOS's degrees of structural constraint are correlated to the latter document typology. In addition, we notice that storage media associated with documentarization activities depend on the purposes of these operations and, to an extent, on the public they are addressed to.

6.0 Conclusion

Through a study of document types for developing faceted classification, we recommend degrees of structural constraint for KOS's used for documentarization of working documents.

Our tool, the flexibility of which has been mentioned before, allows us to apply varying degrees of structural constraint of KOS's to faceted classifica-

Types of Groups	Document Types	Management of information activities Types	KOS Types	Storage Medias
Individual	Individual work document External document	Individual management information activity (tagging)	Informal KOS (folksonomies)	Individual workstations Individual storage medias
Trade/ Project	Document of collaborative work Project monitoring document Trade documents Auxiliary resource document External document	Collective management information activities (sharing of documents)	Medium formalized KOS (sémiotic ontologies, faceted classification)	Shared Disks Servers
Corporate	Referential document	Information management activities for the organization (KM, Record keeping)	Formal KOS (formal ontologies , thésaurus, faceted classification with high degree of constraints)	Servers, DMS database, Intranets, Portals

Table 4. Correlation between KOS types and theoretical types of documents

tion, although it was first designed for a top-down approach for document management, for knowledge management.

The interest in considering activities in the creation of faceted classification lies in the opportunity to make the specificity degree for the classification variable, thus for indexing and then for retrieval. Users' priorities differ within a department, as does the volume of documents produced during the execution of professional tasks. We assume that the possible variation of degrees in specificity in information management tasks reduces the cognitive costs implied by those activities. Considering activities also allows for fragmenting facet values in several distinct facets, since their amounts might potentially be too large.

Faceted classification makes information management easier by providing multi-point-of-view access to documents. One can remember heterogeneous elements for retrieval, thus, if the searched documents have been indexed by the means of faceted classification, one can recognize potential elements used for the documentarization in facets. Stakes related to graphic interfaces for presenting faceted classification are involved in the efficiency and the success of this kind of system.

Note

1. National Agency for Research

References

Aberts, Inge. 2009. "Exploitation des genres de textes pour assister les pratiques textuelles dans les environnements numériques de travail: le cas du courriel chez des cadres et des secrétaires dans une municipalité et une administration fédérale canadiennes." Thèse de Doctorat, Université de Montréal.

Ali, Fauzi M. and Du, Weichang. 2004. Toward reuse of object-oriented software design models. *Information and software technology* 46: 499-517.

Broudoux, Evelyne. 2012. Quelles lectures du tagging? Modélisation, techniques et usages. à paraitre, programme de VSST 2012.

Broughton, Vanda. 2006. The need for a faceted classification as the basis of all methods of information retrieval. *Aslib Proceedings* 58: 49-72.

Coenen-Huther, Jacques. 2007. Classifications, typologies et rapport aux valeurs. *Revue Européenne des sciences sociales* 45n138: 27-40.

Desfriches Doria, Orélie, and Zacklad, Manuel. 2010. Apports de la psychologie du travail pour caractériser l'activité de gestion de l'information. In *Actes du colloque Cide 13, document numérique entre permanence et mutations,* Paris, INHA, 16-17 Décembre 2010. Paris: Europia Productions, pp.53-69.

Desfriches Doria, Orélie. 2011. Les SOC semi-formels au service de l'exercice des métiers et de l'innovation dans les organisations. In *Hypermédias*

et pratiques numériques: actes de H2PTM'11, 12-14 octobre 2011, Université Paul Verlaine, Metz. Paris: Hermès Science Publications.

Gagnon-Arguin, Louise, and Vien, Hélène. 1998. *Typologie des documents des organisations: de la création à la conservation.* Sainte Foy, Québec: Presses de l'Université du Québec.

Kwasnik, Barbara H. 1999. The role of classification in knowledge representation and discovery. *Library trends* 48 no. 1: 22-47.

La Barre, Kathryn. 2010. A semantic (faceted) web? *Les cahiers du numérique* 6n3: 103 – 31.

Leplat, Jacques. 1997. *Regards sur l'activité en situation de travail: contribution à la psychologie ergonomique*. Paris: Presses universitaires de France.

Maniez, Jacques. 1999. Des classifications aux thésaurus: du bon usage des facettes. *Documentaliste – sciences de l'information* 36: 249-62.

Marradi, Alberto. 1990. Classification, typology, taxonomy. *Quality & quantity* 24: 129-57.

Marleau, Yves, Mas, Sabine, and Zacklad, Manuel. 2008. Exploitation des facettes et des ontologies sémiotiques pour la gestion documentaire. In Broudoux, Evelyne, et Chartron, Ghislaine eds. *Traitements et pratiques documentaires, vers un changement de paradigme?: actes de la deuxième conférence document numérique et société, Paris, CNAM, 17-18 novembre 2008.* Paris : ADBS, pp. 91-110.

Mas, Sabine, Bénel, Aurélien, Cahier, Jean-Pierre, and Zacklad, Manuel. 2008. Classification à facettes et modèles à base de points de vue: différences et complémentarité. In Guastavino, Catherine, and Turner, James eds. *Information sans frontières: interactions entre la BSI et d'autres disciplines, actes du 36e congrès annuel de l'Association canadienne des sciences de l'information (ACSI)*, Université of British Columbia, Vancouver, 5-7 juin 2008.

Pikas, Christina K. 2007. Personal information management strategies and tactics used by senior engineers. In *Proceedings of the American Society for Information Science and Technology* 44: 1–21.

Vickery, Brian. 1960. *Faceted classification: a guide to construction and use of special schemes.* London: Aslib.

Zacklad, Manuel. 2006. Documentarisation processes in Documents for Action (DofA): the status of annotations and associated cooperation technologies. *Computer supported cooperative work* 15: 205-28.

Zacklad, Manuel. 2007. Une théorisation communicationnelle et documentaire des TIC. In Brossaud, Claire, and Reber, Bernard eds. *Humanités numériques 2: socio-informatique et démocratie cognitive.* Paris: Hermes.

Zacklad, Manuel. 2011. Cinq critères d'évaluation des systèmes d'organisation des connaissances. *Les cahiers du numérique*, à paraître.

Zeller, Jean-Daniel. 2004. Documents numériques: à la recherche d'une typologie perdue… *Document numérique* 8n2: 101-16.