



Table of contents

| | |
|---|-----------|
| 1. Introduction in the problem | 2 |
| 2. Typology of Intranets | 2 |
| 3. Table of the evolution of the manners | 4 |
| 4. Features of Intranets and archival analysis..... | 6 |
| 4.1. GENERIC FEATURES | 7 |
| 4.1.1. Directories..... | 7 |
| 4.2. FEATURES OF COMMUNICATION..... | 7 |
| 4.2.1. The e-mail | 7 |
| 4.2.2. The immediate e-mail and chat rooms..... | 7 |
| 4.2.3. Blog..... | 8 |
| 4.2.4. Wiki..... | 8 |
| 4.2.5. Audio and video conference | 8 |
| 4.2.6. Social Networks and Forums..... | 8 |
| 4.3. FEATURES OF COORDINATION..... | 8 |
| 4.3.1. Tools of workflow and processes management | 8 |
| 4.3.2. Datebook and shared calendar..... | 8 |
| 4.3.3. Organization Chart (Organigram) | 8 |
| 4.3.4. Management of project, tools of schedule and reporting | 9 |
| 4.4. FEATURES OF COLLABORATION | 9 |
| 4.4.1. Professional Workflow..... | 9 |
| 4.5. FEATURES OF CAPITALIZATION OF THE KNOWLEDGE | 9 |
| 4.5.1. Contents Management (CMS)..... | 9 |
| 4.5.2. Knowledge Management (KM) | 9 |
| 5. (Temporary) Conclusions | 10 |
| 6. Bibliography | 12 |

1. Introduction in the problem

The question of the archiving of Intranets did not make, to my knowledge, the object of any publication. While the question of the archiving of the internet appeared with the project of Brewster Kahle's internet archives in 1996¹, then was quickly considered as a problem bound to the publication and to the registration of copyright, what justified its consideration by the national libraries, on the other hand Intranets remained a " black hole " in the archival area.

The only French-speaking publication on this matter is a communication of H  l  ne Louhmeau during a session of the group PIN (Pr  servation de l'Information Num  rique)² in 2004, communication which also integrated the management of the Internets of the French administration offices [LHOUMEAU, 2004]. In the German-speaking domain, two works of diploma of high specialized school deal as the subject [SCHLIETER, 2003] and [JOHNER, 2009]. I speak about it more in detail in the conclusion.

The question of the archiving of the internet was rather quickly the object of a consensus as for its scope and for its methodology (I send back here to the works of International Internet Preservation Consortium - IIPC), although the national options vary sharply as it exists or not a registration of copyright for the electronic publications. The problem of Intranets is almost the opposite of that of Internet sites. Indeed, for the internet, the question is « where are the boundaries of the national domain in a globalized network? » while for Intranets the question is « how to access it? » because they belong by definition in the private or inaccessible domain to the public even if they are the fact of the public administration.

2. Typology of Intranets

Besides their lack of visibility, a second difficulty results from a lack of consensus about a typology of Intranets. An analysis of the literature print these last years shows classifications centered on very different criteria which we can synthetize as follows:

- a / Typology based on the functions of companies
- b / Typology based on software tools
- c / Typology based on working tools

In the case of the **typology based on the activities** of the company, there is no specificity linked to the contents or to the tools used for the Intranet, indeed we consider here the producers and them activities. Although this is little satisfactory methodologically speaking, this structuration will be the one that I shall use during this article for two reasons. In the first place, the analysis by service or producing sector sends back to a fundamental principle in archival, which is that of the

¹ See http://fr.wikipedia.org/wiki/Internet_Archive and http://en.wikipedia.org/wiki/Internet_Archive

² The « groupe PIN » is a meeting of the most french public and private actors concerned by the digital archiving problematic, web site : http://www.aristote.asso.fr/PIN/pin_groupe.html

origin. Secondly, the rare available statistical data (see chap. 3 further) are also structured according to this distribution. We find this typology in [ALIN, 2000].

For the **typology based on software tools**, we collide with the fact that it is not discriminate with regard to the Internet. Indeed, in their great majority, the used softwares are the same in both cases. The only change is the accessibility to pages managed by both systems. It is so true that certain software propose a single editing tool allowing to choose a publication either on the Internet, or on the Intranet, or simultaneously, according to the concerned pages. On the other hand, these software tools evolve very quickly. So, publications of hardly two years refer to already obsolete tools in the practice. In the same order of idea, the perimeter of tools changes. 10 years ago, the universal tool amounted to a Html editor, today, as show it the statistics of Internet sites archived by our Danish colleagues [JACOBSEN, 2007], who have some steps beforehand on this subject, we notice that only 25 % of the collected objects are Html pages (25 % of files are jpeg and 20 % are Pdf files) and it is not unreasonable to think that the contents of Intranets present the same composition in term of file formats, as confirms it the tendencies 2007-2009 loosened by the inquiry of the "Observatoire de l'Intranet", is: 3 times more videos, twice more sound files and flash animations (see [Observatoire Intranet, 2009], page 15).

Finally the **typology based on working tools**, which appears in fact as a variant of the previous option, as far as these tools are often supported by a dedicated computer application or a version in « Web service » of these applications, we find transverse tools that are directories, mail-systems, workflows, knowledge management, without that these functions are easily separable because the offer in software packages tends to make them more and more interoperable.

It is necessary to insist on the extreme volatility of the software offer. To take a simple example about editing tools, the University Hospitals of Geneva (HUG) used not less than 5 in 10 years (1999-2009). For memory: Frontpage (1999-2003), Quickplace (2000-2009), Dreamweaver (from 2005), Neospark (from 2007), and soon Jahia (from 2010).

The "Observatoire de l'Intranet" (see detail of this body in the chapter 3) proposes an analysis which mixes these dimensions by presenting a table with the big informative functions of the company (communication, coordination, collaboration, capitalization of the knowledge), which are not necessarily managed by a particular service because they are transverse functions by nature, and the types of applications used by these various functions, neighborhood a dozen by function (see the table below).

Table of the features of Intranets

| Domains | Applications | | |
|---------------------------------|---------------------------------|----------------------------|----------------------------------|
| Communication | e-mail | Direct e-mail | Chat |
| | Whiteboardc | Blog | Wiki |
| | Audio-conferencing | Visio-conferencing | Social Networks |
| | Forums | Profiling | Personalisation |
| Coordination | Workflow tools | Generic Workflow | Business Process Mangement (BPM) |
| | Shared Datebook | Organigram | Calendar |
| | Pad of adresse | Spécific phone directories | Competencies directories |
| | Projects Management | Planning tools | Reporting tools |
| Collaboration | Generic software | Specific software | Professional Workflow |
| | Shared Edition | Gestionnaire de tâches | Outils de CRM |
| | Projects spaces | Spaces of services | Community spaces |
| | Audio-conferencing | Visio-conferencing | |
| Capitalization of the knowledge | Content Management System (CMS) | Decentralised Edition | |
| | Full-text search engine | Indexed search engine | Multi-dimensionnal search engine |
| | Metadatas | Indexation tools | Knowledge Management |
| | Documentary Workflow | Workflow of validation | ARchiving tools |

Source : Observatoire de l'Intranet, <http://www.observatoire-Intranet.com/>

-> A propos de l'Intranet -> Fonctionnalités de l'Intranet

We can observe that certain applications are present in various domains, what shows well that the typology is still far from being stabilized, and that it will never be it maybe.

3. Table of the evolution of the manners

As Intranets are not public by definition, it is extremely difficult to have a global vision of their manners. Nevertheless, there is a French body which leads regularly inquiries to this subject since 1999 (I did not find similar in the English-speaking world there, although he certainly has to exist inquiries of the same type, but they are generally paying, and expensive...), it is about "Observatoire de l' l'Intranet"³. Regrettably, only the last published inquiry is on-line. I was nevertheless able to collate the figures of inquiries 1995, 2000, 2004, on 2007 and 2009 which show the following tendencies.

³ See web-site: <http://www.observatoire-Intranet.com/>

Part of the Intranet pages according to the producing services

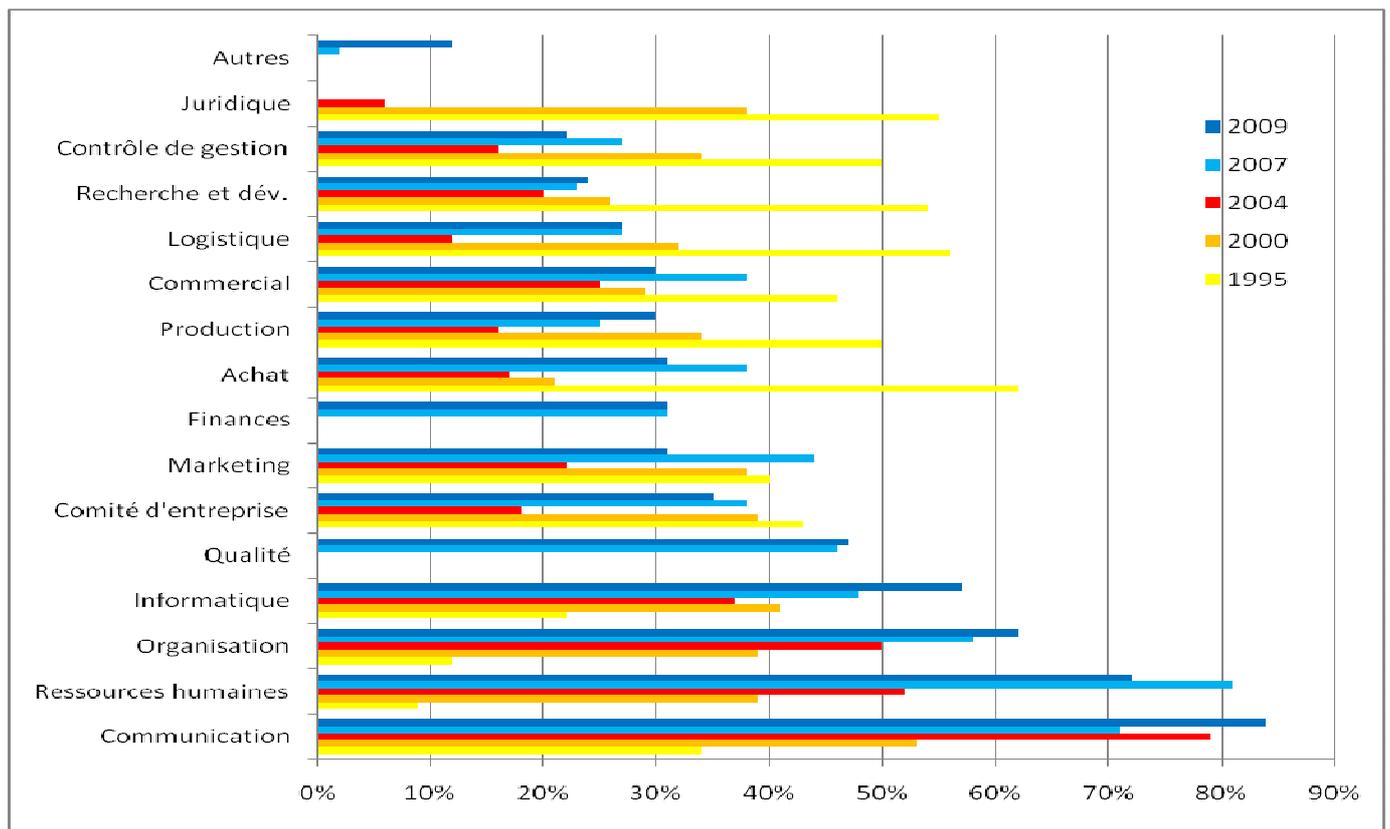
| Producing services | 1995 | 2000 | 2004 | 2007 | 2009 |
|--------------------|------|------|------|------|------|
| Communication | 34% | 53% | 79% | 71% | 84% |
| Human Ressources | 9% | 39% | 52% | 81% | 72% |
| Organisation | 12% | 39% | 50% | 58% | 62% |
| Data Processing | 22% | 41% | 37% | 48% | 57% |
| Quality | | | | 46% | 47% |
| Factory Committee | 43% | 39% | 18% | 38% | 35% |
| Marketing | 40% | 38% | 22% | 44% | 31% |
| Finances | | | | 31% | 31% |
| Purchase | 62% | 21% | 17% | 38% | 31% |
| Production | 50% | 34% | 16% | 25% | 30% |
| Commercial | 46% | 29% | 25% | 38% | 30% |
| Logistics | 56% | 32% | 12% | 27% | 27% |
| Research et dev. | 54% | 26% | 20% | 23% | 24% |
| Management control | 50% | 34% | 16% | 27% | 22% |
| Legal | 55% | 38% | 6% | | |
| Others | | | | 2% | 12% |

Source : Observatoire de l'Intranet, <http://www.observatoire-Intranet.com/>

We notice the net dominion of the communication services in the production of the Intranet pages for more than 10 years. We also denote the more and more active part assumed by the human resources services and the organization services and, in a lesser measure, that of the data processing services.

For the human resources, it is mainly about the putting at disposition of numerous forms necessary for the management up to the implementation of workflow in a recent time. For the organization, it is mainly about the publication of the procedures, in which have can add the quality processes which generate documents of the same nature. Finally, as regards the computer services, it is most of the time about the putting at disposition of the access to applications data processing via the Intranet.

Other sectors mark a generalized proportional recession while remaining in an average proportion of 25 % of the published information. The following table illustrates graphically these progress.



4. Features of Intranets and archival analysis

To avoid difficult questions we could content with assuming that the using/producing services propose on their Intranet pages of the documents/information which stand out from their traditional activities, documents to which we could apply the same rules of conservation. Regrettably, the clipping to which are subjected these documents in an electronic environment does not allow this abstract "copy and paste". So, I resume below the various types of documents/functions by putting in motto their "archivability". It is about a first sketch, opened to the discussion. The fact that I am obliged to formulate it with terms such as document/information or document/function denotes in which point the usually useful limits in a paper environment become vague in an electronic environment.

As we can see it in the higher chapter 3, the initial usage of the Intranet was mainly that of a tool of communication and internal publication. As such it interested marginally the archivists. The current tendency shows however that more and more documents formerly published under paper shape were made at first available on the Intranet simultaneously to the paper edition. The familiarity helping, it happens now that documents exist only under electronic shape. A parallel tendency shows that it is not any more only informative documents which take this way of distribution but also the normative and/or transactional documents.

On the other hand, if the Intranets of the debuts were firstly considered as instruments of publication and communication, they are more and more (at least in the companies of a certain size) designed as portals, which do not give any more exclusively access to documents/information, but also to resources which the big categories, used below, are: the

generic features, the documentary bases, the data bases, the professional softwares). The home page of the Intranet becoming then mostly the default page in the opening of the jobs of the co-workers. In this configuration, home pages (generally the first 3 levels) serve as way of points towards specific applications, which are not necessarily hosted by the structure of the intranet itself.

I detail below the present main features in Intranets, and their "archivability".

4.1. GENERIC FEATURES

4.1.1. Directories

According to the different authors the term of (phone) directory in the Intranet applies to rather different notions. He can involve the on-line putting at disposition of an image of phone directories before published (in the form of a Pdf document for example, generally updated once a year). He can involve the same version of this phone directory, enriched by features of search by name, function, localization, etc.). He can finally involve a real on-line computer directory, which is then a data base, possibly updated real time and answering generally to the LDAP⁴ standard. In this last case, it is not any more a document but a transverse software tool, adult of which one of the features is the mastery of the rights of access in the computer system. It implies that the recording linked to a co-worker has to be completely historicized (date of the beginning/end).

In this last case, the logic of archiving must that be applied to data bases.⁵

4.2. FEATURES OF COMMUNICATION

4.2.1. The e-mail

Relatively recent works (by ex: [ALIN, 2002], [CREPLET, 2003], [HUMEAU, 2005]), still consider that the e-mail is an internal function to the company and belongs to the Intranet. The current archival consensus considers rather that the electronic mail widely substituted itself for the paper mail and that it is important to identify and to process e-mails considered as "records".

I excluded thus this function of the problem of the Intranet and I send back to the works existing on this subject, and more particularly on the excellent publications of our neerlandophones colleagues of Digital Testbed Bewaring [Digitale Duurzaamheid, 2003] and of the group of expert e-David [AVG, 2007] Archives de la Ville de Genève

Archivage des sites web institutionnels (intranet et internet) aux Archives de la Ville de Genève

Décembre 2007 (non publié)

[BOUDREZ, 2003].

⁴ LDAP mean Lightweight Directory Access Protocol, see http://en.wikipedia.org/wiki/Lightweight_Directory_Access_Protocol

⁵ For ressources finding about the data bases archiving problematic, see for instance **Long-term Preservation of Databases**. In: *Electronic Resource Preservation and Access Network (ERPANET)*, 2003. <http://www.erpanet.org/events/2003/bern/>

4.2.2. The immediate e-mail and chat rooms

If the electronic mail knew a lightning progress, it is not the same as regards the immediate e-mail there. The look-out post of the Intranet indicates a 19 % penetration in 2008, but a tendency in increase (49 % in the future). We can consider that it is mainly about a function of discussion and that it has no archival interest as records, seen its short-lived side.

4.2.3. Blog

The usage of the blog of company is rather recent. It is possible whether it is only an effect of mode. We can for the moment consider it as a discussion and to apply it the same analysis as for the immediate e-mail. It is however necessary to continue to observe its manners which in combination with wikis could end in a more marked institutional value.

4.2.4. Wiki

The usage of wikis is also recent in organizations. We can distinguish two contents of wiki the management of which should be differentiated. Wikis linked to projects, which constitute the accumulation of the data bound to these. They constitute then the contextual documentation of these projects and would as such be archived at the closing of the project. Wikis constituting knowledge bases of the organization which by definition is unlimited. It also constitutes an important knowledge of the organization but their constraint of management implies a regular archiving on a chronological base (annual for example).

4.2.5. Audio and video conference

We can apply them the same rule as the immediate e-mail.

4.2.6. Social Networks and Forums

It is about tools of sharing of the dynamic competence there, and also short-lived. It does not thus seem desirable to archive them unless we want to understand the internal dynamics of the organization. In that case, the sampling of a sensibly chosen chronological edge could be imperative.

4.3. FEATURES OF COORDINATION

4.3.1. Tools of workflow and processes management

The use of these tools via the Intranet is generally made under the shape of Web service opening a window on underlying applications. We shall thus leave the principle that it is the archiving of these applications that must be managed.

4.3.2. Datebook and shared calendar

These tools, besides often integrated into e-mail software have a short-lived use and should not thus be archived.

4.3.3. Organization Chart (Organigram)

The computerized management of organigrams is potentially different from those of these managed on analog support. These last ones are often the object of an annual approval while the electronic version can be updated in the course of the institutional decisions. As far as the variations of the organization chart are an important contextual resource, it is necessary to set up rules of clear management on this subject, with the aim of an adequate archiving.

4.3.4. Management of project, tools of schedule and reporting

As for workflow, the use of these tools via the Intranet is generally made under the shape of Web service opening a window on underlying applications. We shall thus leave the principle that it is the archiving of these applications that must be managed.

4.4. FEATURES OF COLLABORATION

4.4.1. Professional Workflow

The use of these tools via the Intranet is generally made under the shape of Web service opening a window on underlying softwares. We shall thus leave the principle that it is the archiving of these applications that must be managed.

4.5. FEATURES OF CAPITALIZATION OF THE KNOWLEDGE

4.5.1. Contents Management (CMS)

Contents management system of are more and more used as tools of management of Intranet (and internet) contents as far as they allow modalities of customizable, varied publishing according to the groups of users (the same document which can appear in a different way according to the reader) what avoids creating several versions of the same document. It also allows to manage more easily multilingual sites⁶. CMS having most a structure of underlying data base we should choose to archive CMS itself rather than its Intranet version. It remains however to determine if the varied formats of publishing must be kept or not, and how.

4.5.2. Knowledge Management (KM)

The Knowledge management is not necessarily managed by a specific software but can be a constellation of applications aiming at the sharing of the knowledge (included the forums, the blogs, the wiki, etc.). He can however appear under the shape of a dedicated software, in that case the Intranet is generally only a mode of presentation and it is obviously the KM application which it is necessary to archive as such.

⁶ It is the case for instance of the web-site of the "association des archivistes suisses", witch use the CMS software Typo3, see: <http://www.vsa-aas.org/>

5. (Temporary) Conclusions

The review of the present various features in Intranets shows at present three possible scenarios:

- A) The **information/documents of short temporal values**, which do not require in priori no archiving. Let be the immediate e-mail applications, the blogs, the audio and the video conferences, the social networks and the forums, the datebooks and the shared calendars.
- B) The **information/documents only available on the Intranet**, having an archival value and for which it is necessary to develop strategies of archiving. Let be directories, wikis, organization charts (so managed digital).
- C) The **information** available on the Intranet but **resulting from the other computer softwares**. The strategy of archiving consists then in archiving the applications rather than the views which they supply on the Intranet. Let be e-mail applications, workflows, CMS, tools of Knowledge management, tools of projects management.

As far as an Intranet would not contain sources of information of the type b) and c) we can ask the question of a punctual archiving on a regular chronological snapshot base, the same type as those made on Internet sites, to keep a periodic image of the management of the information within the organization. This way of making collides then with the same constraints of conservation as Internet sites and we should envisage for the long term a warehouse of Intranet which would be the electronic equivalent of the private holdings archived at present.

If we compare this analysis with that of [SCHLIETER, 2003], who uses criteria linked to the document type strictly speaking and who holds for example « *activity reports, reviews of company and customer's reviews, the commemorative brochures of anniversaries, inaugurations of building and subsidiary and the mergers, the documentations [??], the published articles, the reports, the books of the members of the directions or the other directions, the press kits, and the advertisement materials.* », we see although the traditional documentary typology has difficulty in applying in a computer environment and, mutually, the new collaborative functions (collectively and fallaciously known under the naming *Web 2.0* or *Entreprise 2.0*) appearing in organization have not found them "diplomatic" yet.

The solution of the snapshot was also chosen by the rare institutions which made an archiving of Intranet and knowledge of which I was able to have. The list of these projects is briefly presented in the table below. Most of these projects were not the object of a strictly archival evaluation. Most of the time, data bases were excluded from these "harvesting" for technical reasons rather than archival one.

It seems that as long as we shall have no doctrine argued of the conservation of Intranets, it is this (relatively) simple solution which will be privileged.

Comparative degree of the operations of archiving on intranet

| Institution | Volume archived | Technique used | Periodicity | Remarks |
|---------------------------------|---|--|---------------------------------|---|
| Ville de Genève | Heritrix : 385 Mo HTTrack : 922 Mo | Harvesting with the crawlers Heritrix et HTTrack | One snapshot in november2007 | See [AVG, 2007] |
| Radio Télévision Suisse Romande | 5 Go/ year | Direct copy from the servers | Yearly since 2001 | Totality of the intranet |
| | 3,23 Go | Direct copy from the servers | Yearly since 2001 | Selection of relevant document without institutional decision |
| ABB | 10-15 Go | Harvesting with the crawler HTTrack sur les pages internet et sur le serveur interne | Yearly | Internet and Intranet together |
| Dresdner Bank | ??? | Harvesting with the crawler HTTrack | One unique snapshot in 2005 | Transformation from HTML to XML for archiving |
| INRIA | ??? | Direct copy from the servers | 5 versions between 1995 and2009 | Some versions are very incomplete |

6. Bibliography

Note : all the URL were checked at 28.02.2010

[ALIN, 2000] Frédéric ALIN, Denis LAFONT, Jean-François MACARY

Le projet Intranet

Paris, Eyrolles, 2000

[ALIN, 2002] Frédéric ALIN, Xavier AMAORS, Marc SALIOU

L'entreprise Intranet

Paris, Eyrolles, 2002

[AVG, 2007] Archives de la Ville de Genève

Archivage des sites web institutionnels (intranet et internet) aux Archives de la Ville de Genève

Décembre 2007 (non publié)

[BOUDREZ, 2003] Filip BOUDREZ, Hannelore DEKEYSER, Sofie VAN DEN EYNDE

DAVID : Archiving e-mail

Site de Digitale Archivering in Vlaamse Instellingen en Diensten project, 2003.

<http://www.antwerpen.be/david/website/teksten/Rapporten/Report4.pdf>

<http://www.edavid.be/davidproject/teksten/Rapporten/Report4.pdf>

[CREPLET, 2003] Frédéric CREPLET

Ingénierie de projet Intranet

Paris, Edition d'organisation, 2003

[Digitale Duurzaamheid, 2003] Digitale Duurzaamheid

Site des Archives nationales néerlandaises, [En ligne]

From digital volatility to digital permanence: Preserving email

<http://www.digitaleduurzaamheid.nl/bibliotheek/docs/volatility-permanence-email-en.pdf>

[HUMEAU, 2005] Nicolas HUMEAU

Intranet-Management

Paris, Economica, 2005

[JACOBSEN, 2007] Grethe JACOBSEN

Harvesting the Danish internet – the first two years

Netarchive.dk, mai 2007

http://netarchive.dk/publikationer/CollectingTheDanishInternet_2007.pdf

[JOHNER, 2009] Simon JOHNER

Webseiten-Archivierung – Ein Konzept zur Bewertung, Selektion und Emulation von geschäftlichen Internetauftritten am Beispiel der ABB Homepage

Diplomarbeit Fachhochschule Chur, Juli 2009

Travail non-publié

[KHOAJA, 2004] Bessem KHOAJA, Carole COUTURE

La gestion et l'archivage des sites web institutionnels

In *Archives*, Association des archivistes du Québec, vol 35, n° 3 et 4, 2003-2004.

<http://www.archivistes.qc.ca/IMG/pdf/35-3-4-KhouajaCouture.pdf>

[LHOUMEAU, 2004] Hélène LHOUMEAU

Archiver un Intranet: aspects méthodologiques. Cas du site des ministères sociaux, 2004.

<http://bibnum.bnf.fr/conservation/aristote2004/lhoumeau.ppt>

[MALAISON, 2004] Claude MALAISON

L’Intranet dans tous ses états : une approche interculturelle de ses multiples dimensions

Montréal ; I. Quentin, 2004

[Observatoire Intranet, 2004] Observatoire de l’Intranet

Résultat d’enquête – Edition 2004, 2004

<http://web.archive.org/web/20050530143130/www.observatoire-intranet.com/documents/ObsIntranet2004-Synthese.pdf>

[Observatoire Intranet, 2007] Observatoire de l’Intranet

Résultat d’enquête – Edition 2007, 2007

[Observatoire Intranet, 2009] Observatoire de l’Intranet

Résultat d’enquête – Edition 2008-2009, 2009

<http://www.observatoire-intranet.com/>

NB : le site affiche la dernière enquête publiée

[SCHLIETER, 2003] Antje SCHLIETER

Archiving websites – Archivierungskonzept für das Intranet der Dresdner Bank AG

Diplomarbeit zur Erlangung des Titels Diplom-Archivarin (FH)

Fachhochschule Postdam, Juli 2003

http://forge.fh-potsdam.de/~ABD/wa/Diplomarbeiten/Diplomarbeiten_Dokumente/Diplom_Scheidung.pdf

Résumé: http://www.fes.de/archiv/adsd_neu/inhalt/spiegelungsprojekt/dokumente/schlieter.pdf