

# Helena Francke

## Towards an Architectural Document Analysis

*The text that follows is a slightly revised extract from my doctoral thesis (Francke 2008) [1]. It draws on the notions of information architecture and document architecture to identify a set of concepts for analysing documents as sociotechnical artefacts. The analytical tool was developed for the purpose of analysing scholarly journals, which has tinged it to some extent, but it is likely applicable also for analysis of other document types and genres. As part of identifying concepts, I refer to several of the more established ways of conceptualising information architecture and document architecture. As such, the text may provide an introduction to these two areas, as they looked a few years ago when the thesis was prepared. The article may thus be of interest both as a first introduction to the two perspectives and from a theoretical and methodological point of view.*

### Abstract

Information architecture (IA) and document architecture (DA) provide two, partly overlapping, perspectives on the creation of document structures. This article suggests how the architecture of a document can be analysed from these two perspectives. Literature on IA and DA has been examined in order to identify central ideas that are of relevance for analysing the architectures of digital documents. The article contains an overview of how IA and DA have been used and defined. The article shows how a model for analysing documents as sociotechnical artefacts can fruitfully draw on parts of the theoretical and practical complexes of IA and DA. The aspects that are identified as particularly important from IA are organisation systems, navigation, and labelling. From DA, logical structures, layout structures, content structures, and file structures are all applicable aspects. It is discussed how these various aspects may be interpreted in order to support an analysis of the organising principles of documents.

The architecture metaphor provides a model for understanding and conceptualising documents as sociotechnical artefacts. Specifically, information architecture and document architecture equip us with tools for the analysis of documents as artefacts. Originally, they are two ways of thinking about the production of documents and texts, and in this article I will investigate how the perspectives they represent can also be used as analytical tools in studying the organisation of epistemic content [2] in documents in general, and on web sites in particular. Using the architecture metaphor as a point of departure implies a view of documents as similar

in certain respects to buildings. To use metaphors as a research tool in this manner in order to emphasise certain perspectives of a phenomenon is not uncommon (Alvesson & Sköldbberg 1994, 141 ff.). Sabine Maasen and Peter Weingart state that

*(f)rom the 1960s onwards, scholars increasingly were of the opinion that metaphors indeed served important discursive ends. While explanations and evaluations still vary enormously, ever few scholars doubt the considerable, if not constitutive power of metaphors (2000, 25).*

The constructive power of metaphors entails that they make us attentive to some, although not all, aspects of a phenomenon, which in my case is the document's organisation. I will begin by looking at how appropriate the architecture metaphor is when applied to documents, before I proceed with a discussion of how looking at document and information architectures can be useful as a means to bring out the organising principles of documents. In doing so, I will draw on similarities with traditional library and information science knowledge organising systems.

Before I continue, I need to clarify my use of the terms “structure” and “architecture”, however. “Architecture” is used here as a term at a higher hierarchical level than “structure”. That is, a number of structures make up an architecture, structures which describe the relationships between elements in the document. A document can be described from different architectural perspectives. The inspiration for the two architectures that are in focus in this case – document architecture and information architecture – comes from two separate traditions, described more closely under their own separate headings below. They represent different perspectives from which to view a document, and occasionally the two architectures can share the same document structures, so that these can be described in terms of either document or information architecture. As both terms spring from existing traditions, I have chosen to continue referring to them as document architecture and information architecture respectively, even though a differentiation between an architecture of “document” structures and one of “information” structures makes little sense in this context. Although the terms are retained, both traditions are treated here as dealing with the material organisation of the document's epistemic content, though they do so from two different points of view.

## The History of the “Text/Document as Architecture” Metaphor

The juxtaposition of the metaphorical fields of architecture and text date back

to antiquity. Plato, in *Gorgias*, observes that writer and builder alike handle “problems involving the ordering, framing, and fitting together of materials” (Cowling 1998, 140; Plato 2000, 502a-504a). In Roman rhetorical pedagogy, architecture and the building are used as metaphors both at the grammatical level, at the disposition level and as an image of the entire process of the rhetorical activity: *inventio* “dispositio” *ornatus* (Cowling 1998, 140 ff.). David Cowling points out that in this context, the metaphor is usually applied to the process of constructing a literary composition rather than to the product, but ascribes it to the didactic purpose of the use (Cowling 1998, 141).

These ancient uses of the architecture metaphor primarily stress one or several of the aforementioned rhetorical stages as shared between the art of constructing a building and constructing a verbal text (which was usually oral): *inventio* “gathering or constructing material; *dispositio* “ordering it according to some form of structure (which may be a hierarchical arrangement); and *ornatus* “decorating it in accordance with the intended style and audience. The metaphor “the text is a building” continued to be used, with slightly different connotations, in, for example, Christian exegesis and medieval and renaissance poetry (Cowling 1998, 143 ff; *passim*). In transporting the metaphor from the domain of text to the domain of document, the potentially material nature of the metaphor’s literal meaning, the building, is transferred to the document [3].

In more recent times, the architecture metaphor has been used in different computer-related discourses. Harold Lorin has described how the metaphor was initially used in the sense of processor architecture “to mean the view of a computing system as seen by a programmer or automated code generator [but how] informal usage often blurs the distinction between design and architecture” (Lorin 1986, 256). With time, the metaphor has also come to be used in other noun phrases such as systems architecture, software architecture [4], information architecture, and document architecture, where it often refers to models of document types or information networks or the study and application of these [5].

## Information Architecture: Uses and Definitions

Information architecture is the label often used for a practice that claims to apply a user’s perspective. It is a production activity focused on the functionality of large web sites [6]. As such, it involves people with a background in a large number of practices or disciplines, such as graphic design, library and information science, journalism, usability engineering, marketing, human-computer interaction, and so on (Rosenfeld & Morville

2002, 18 ff.) However, the central task for the information architect, as presented in the information architecture literature, is to work as a project manager, getting the stakeholders to work together in creating a web site “an information environment” that will be user-friendly and meet the requirements of the project (see e.g. Morrogh 2003; Rosenfeld & Morville 2002; Van Dijck 2003; Wodtke 2002). Morrogh explains how “(i)nformation architecture is primarily about the design of information environments and the management of an information environment design process” (Morrogh 2003, 6). Another main task for the information architect is to maintain the overall vision of the architecture of the web site. This is an aspect that has strong similarities to an area in architecture called “wayfinding”, which aims to facilitate people’s orientation in large, often unfamiliar buildings, such as airports (Van Dijck 2003 91; Muhlhausen 2006). This aspect of information architecture deals with “organization, labeling, and navigation schemes within an information system” (Rosenfeld & Morville 2002; Head 2001). It has similarities to such knowledge organising tools as tables of contents and back-of-the-book indexes in print, but traditional knowledge organisation tools such as thesauri, controlled vocabulary, and classification are also used (Toms 2002, 860) [7]. Furthermore, it is the aspect of information architecture that is most closely linked to a library and information science competency, and the one that I will cover in this article.

The focus in information architecture is on facilitating the use of web sites (or other types of documents). Much of the literature is of a “cookbook” nature, offering advice on how best to design a web site, or give suggestions on how users can be involved in evaluating the site. However, it is also possible to use aspects of information architecture to analyse documents of a particular type or genre. It is this type of analysis that is in focus here.

The one thing that the information architecture community seems to agree on when it comes to defining information architecture is that they cannot agree on a definition (Reiss 2000, 2; Morville 2002). Richard Saul Wurman is usually acknowledged as the person who coined the term information architecture in 1975 (Morrogh 2003). Since the late 1990s, information architects have been engaged in a lively debate on an appropriate definition. However, despite the disagreement, many of the attempts to define the area include one or several of the following three ideas (Rosenfeld & Morville 2002, 4; Information Architecture Institute 2005; Wodtke et al. 2001):

- The combination of organization, labeling, and navigation schemes within an information system.
- The structural design of an information space to facilitate task

completion and intuitive access to content.

- The art and science of structuring and classifying web sites and intranets to help people find and manage information.

In this article, focus will be on information architecture as it is visible in a document, rather than on the production and maintenance process. I am interested in the strategies that are employed to make the web site's architecture comprehensible to the user, such as organisation and navigation schemes and structures, categories, labels, and search facilities. Although both document architecture and information architecture are to some extent media-specific, I believe that in many cases it is also possible to speak of these two types of architectures when studying other media than digital ones. For instance, many information architecture functions for helping readers navigate in books have developed throughout history, including pagination, tables of contents, back-of-the-book indexes, and references within the book and to other documents.

A great deal has been written to advise both practicing information architects and more or less experienced web designers on information architecture practices. Louis Rosenfeld and Peter Morville are the authors of one of the most influential handbooks on information architecture, *Information Architecture for the World Wide Web* (1998, in a 2nd ed. 2002, and in a 3rd ed. 2006). With a background in library and information science, they stress the knowledge organisation aspects of information architecture that were outlined above. They provide a substantial discussion and taxonomy for the organisation and navigation of web sites. Andrew Large, Jamshid Beheshti, and Charles Cole (2002) suggest an "applied information architecture" that is adjusted to the characteristics of the users. The authors categorise different approaches to interface design and retrieval, focusing particularly on portal design. Peter Van Dijck mainly addresses designers, and he places information architecture in a wider perspective, but he divides what he specifically refers to as information architecture into organization schemes, categories, labels, and sitemaps (2003, ch. 3). These authors all present partly overlapping categories of information architecture and the treatments these categories are given by the authors will form a point of departure for discussing how information architectures can be analysed in "primarily" web sites. In the discussion, I will also draw on several other sources within, as well as outside, the information architecture community. My focus will be on the categories organisation systems, navigation and searching, and labeling.

## An Analytical Model of Information Architecture

### Organisation Systems

The organisation of the pages of a web site, and of the elements on each page, displays some similarities to the classification of document collections. In most cases, if the web site is large enough, a need will arise for gathering pages with similar topics or functions and make them accessible through a superior label. Depending on the characteristics of the web site, this classification can be made at different levels of granularity. In most cases, it is also important to allow for the possibility that the web site will develop and grow, so that such growth will be possible within the chosen organising principle. Rosenfeld and Morville divide organisation systems into organisation schemes and organisation structures. They describe the difference as follows: “[a]n organization scheme defines the shared characteristics of content items and influences the logical grouping of those items. An organization structure defines the types of relationships between content items and groups.” (2002, 55) The decision to apply a particular organisation scheme, or a hybrid of different organisation schemes, will depend on the type of material intended to be included on the web site; on the producers’ anticipation of who the users will be and what they will do on the site; on the impression the producers wish to make; and on the producers’ (consciously or unconsciously applied) ethical and political values (Bowker & Star 1999, 321; Ørom 2003).

There are two main types of schemes identified by Rosenfeld and Morville: exact and ambiguous ones. The exact schemes are, for instance, alphabetical, geographical, and chronological [8]. However, these schemes are perhaps not quite as uncontroversial as Rosenfeld and Morville make them out to be (2002, 56). Experiences with library classification systems show, for instance, that our conception of geographical divisions may be dependent on a geopolitical standpoint (Hjørland 2004). Such dependence on epistemological standpoint is more obvious in ambiguous schemes. In this category, Rosenfeld and Morville mention topical, task-oriented, audience-specific, and metaphor-driven schemes [9]. Large, Beheshti, and Cole further divide metaphor-driven schemes into organisational (e.g. a library web site organised into different rooms traditionally found in the library), functional (e.g. the desktop metaphor transferred from the real world to the computer screen), and visual (e.g. the “Ask Jeeves” character who “helps” you find what you are looking for in the search engine by that name) depending on how the metaphor is used (2002, 834). Wodtke points out that categorisation is highly dependent on context (2002, 118 f.), and there seems to be an

awareness among several of the authors on information architecture of the influence social factors have on how the web site is organised. Among researchers, Kimmo Tuominen, Sanna Talja, and Reijo Savolainen are only some of many representatives that have noted that classification languages “neutralize and conceal the inherent messiness of reality; the fact that different perspectives and theories provide each document or term or concept with its own specific meaning” (2003, 563; Bowker & Star 1999). At the same time, the use of classification systems (not necessarily previously existing ones) and controlled vocabulary is stressed on several occasions in the information architecture literature, because such structuring of a messy reality is partly the aim of information architecture (Rosenfeld & Morville 2002, ch. 5; Wodtke 2002, ch. 5-6).

It is not only necessary to categorise the material on the web site; the categories need to have a structure, be related to each other in some way. Four basic approaches to this are identified in the information architecture literature: a top-down, hierarchical approach, where navigation takes place up and down through the hierarchical tree structure (Rosenfeld & Morville 2002, 65 ff.; Lynch & Horton 2002; Large, Beheshti & Cole 2002, 834 discuss this in terms of navigation), a bottom-up, relational database approach, which often takes the form of a server-side script-based solution (Rosenfeld & Morville 2002, 69 ff.; Large, Beheshti & Cole 2002, 834), a horizontal, linear approach, where every page only gives access to a previous and a subsequent page in a linear order (Lynch & Horton 2002; Large, Beheshti & Cole 2002, 834), and a web or network approach, where most pages are accessible from most other pages (Rosenfeld & Morville 2002, 73 f.; Lynch & Horton 2002). The different organisation structures provide different possibilities and restrictions in how the web site can be navigated and in how its parts can be conceptually related to each other. When looking at organisation schemes and structures, such aspects as exclusiveness or inclusiveness and breadth or depth are important. The analogy with knowledge organising systems such as classification schemes and subject heading lists is relevant. Is the system specific enough to adequately describe the documents? Is it general enough to be useful for clustering similar documents? Is it too general, so that retrieval will be adversely affected by too many different documents in the same category? Will these clusters be of reasonable size, or do they need subcategories in order to facilitate retrieval? The answers to these questions have to do with the purposes of classification and indexing, which are also of relevance to how the hyperlinks are organised on the web site, to how links and headings are labelled, and to the types of (semi-)automated searching that is possible to perform.

## Navigating and Searching the Site

Barring the use of an external search engine, the submission of a known URL to the browser, and other methods for accessing a web site which are beyond the planning of the information architect [10], a web site can be navigated in two ways: through hyperlinks or through an internal search function. However, the distinction between these two is not unproblematic. Submitting a search to the system will usually produce a hypertext-based list of search results. On the other hand, an index or site map, along with the design of the entire page, may be generated automatically from a database when the user activates a certain link. Thus, the line between webmaster-produced links and database-generated results is, in fact, fuzzy and unclear.

In Hypertext 2.0, George Landow indicates that the function of linking is at the root of the change in textuality that he attributes to hypertext; the link is “the element that hypertext adds to writing and reading” (1997, 11). While linking is central to theoretical views on hypertext, it is also particularly important in web-based information architecture. However, in practice, as well as in theory, the function of the link is not undisputed. The objective of information architects, according to Rosenfeld and Morville, is to provide a navigation system that allows users to find their way through a web site almost without noticing the linking system (2002, 106). The user should be able to access the desired data without detours – these are associated with a negative sense of getting lost. The rationale for this is quite easy to understand: research shows that people navigating the web are impatient. If they do not find what they are looking for directly, they move on (Weinreich et al. 2006, 137 f.). The emphasis on navigation often results in navigation bars, pull-down menus, icons and so on being introduced on all or most of the pages on a web site, creating a highly structured navigation system.

However, this is not an undisputed view of how a navigation system should be constructed. Mark Bernstein has criticised the need for structuring tools that organise and facilitate navigation (Bernstein 1998). He claims that the emphasis on a “centre,” in the form of entrance and navigation pages, creates a premature sense of closure. Furthermore, focus comes to be too much on those parts of a web site that are announced on the entrance page, too much screen space is occupied by navigation tools, and the user is lured away from a certain page or content space too quickly (Bernstein 1998). Bernstein uses metaphors from (landscape) architecture to describe how he, as a user, experiences highly structured, as well as more chaotic, navigation systems. He concludes that what he is looking for is “the artful combination of regularity and irregularity that awakens interest and maintains attention”

(Bernstein 1998), something he considers characterises the garden or park. The choice between regularity and irregularity is not one of right or wrong, of good or bad, but rather is presumably influenced by such factors as the web site's intended function and the community of practice that the designer and anticipated audience belong to [11].

Rosenfeld and Morville (2002) distinguish between navigation systems and searching systems, and further separate between different types of navigation systems: these can be global (site-wide), local [12], or contextual in nature (2002, 112 ff.). Web sites can, according to the two authors, contain these different types of navigational structures simultaneously, and in fact many do. Rosenfeld and Morville's categorisation of navigation systems can be compared to the one made by Karlsson and Malm between hierarchical and decentralised navigational structures (2004, 15), which describes the possibilities or restrictions involved in a user's vertical and lateral navigation of the site (Rosenfeld & Morville 2002, 111). Karlsson and Malm view web sites as mainly or partly hierarchical or decentralised. In a hierarchical structure, the user's path through the web site is limited to a tree structure where one proceeds from a higher level to one immediately beneath it or vice versa, but where paths between pages at the same level or between levels not directly adjoining are missing. Such both lateral and vertical movement is possible on web sites with a global navigation system, but a more restricted version could be accomplished with only a limited local navigation system or contextual links. Karlsson and Malm's decentralised structure, which is characterised by a navigation system which only to a very small degree limits the user's navigation of the web site, could, in its most elaborate form, be viewed as a combination of the global/local structure and the contextual structure. According to Rosenfeld and Morville, the contextual structure consists mainly of editorial "as distinguished from architectural " links, for instance in the form of links in the web site text (2002, 117).

Navigation is facilitated by different navigation elements, which can be integrated or remote. On a page, these take the form of, for example, navigation bars, which can be textual or graphic. The bars on the page can be constant, for instance, placed in a frame in order to keep them on the screen at all times, or be more interactive as in the case of pull-down menus, popup windows, cascading menus, and so on. (Rosenfeld & Morville 2002, 119) Remote or supplemental navigation elements "provide an alternative birds-eye view of the site's content" (Rosenfeld & Morville 1998, 63), such as through an index, a site map, or a guided tour (Rosenfeld & Morville 2002, 121). To help users keep track of where they have been and where they are moving to on the web site is also seen as important. This is achieved through different techniques, such as sign-posts and bread-crumbs, and it is also possible to use "see also" options to facilitate navigation to related or

similar sections of the site (Wodtke 2002, 96 ff.).

A search function on the site can be adapted for different types of information needs, such as known-item searching, existence searching, exploratory searching, and comprehensive searching (Rosenfeld & Morville 1998, 102 f.; Chowdhury 1999, 183 f.). Search functions differ from each other in the complexity of the search engine and in the underlying data in which the search is made. Aspects that influence the search function include:

- if the search is made in the documents' full-text or in some kind of metadata, which can be taken from a controlled vocabulary or authority files.
- Which types of data can be searched, i.e. which metadata fields can be searched.
- How the query can be formulated, including Boolean logic and proximity searching, truncation, and limitations (Chowdhury 1999, 169 ff.)

It is important to remember that information needs is a relative concept. It will be different for different people in the same situation, change for the same person over time and with different situations and tasks, and not least change during the information seeking process. Another aspect of searching which influences the usefulness of the search is how the results are presented, that is, which data are included in the results list. (e.g. Chowdhury 1999, 181)

## Labeling

It is not enough that the categories and links on the web site are identified by the information architect; they also need to be assigned labels to aid users in finding their way on the site [13]. Depending on the purpose and character of the content on the web site, the labels can be taken from an existing controlled vocabulary, such as a subject heading list or thesaurus; a local controlled vocabulary can be created in order to keep the labels on a large web site consistent; users or domain-specific literature may be studied in order to choose vocabulary employed within a particular domain; the labels can be taken from the text on the site; or labels for categories can be invented by the site designer, or borrowed from other web site designs (Rosenfeld & Morville 2002, 95 ff.). Rosenfeld and Morville stress that labels should be thought of as labelling systems rather than individual labels and advocate consistency in the system (Rosenfeld & Morville 2002, 93 f.).

The notion of labels can be wider than only referring to links, whether these are part of a running text or placed separately. Headings are one example of how labels help structure a web page. If appropriately labelled, headings serve as indications of the subject of the text and images displayed below (Rosenfeld & Morville 2002, 83 ff.), and will facilitate the user's navigation through the document. Icons are another example; these can be efficient tools for quickly indicating a topic or category to users, provided that the icons make use of a pictorial language that is well-known to the users (Rosenfeld & Morville 2002, 91 f.).

The above aspects of information architecture have been selected from the information architecture literature as interesting structures to study in digital documents of a particular type or genre. They have an impact on the way in which the user meets the epistemic content of the document. Furthermore, such practices as labelling and creating organisation schemes provide an additional interpretative layer to the content, a layer which will also influence how the content is perceived. Structures of the information architecture function together with the document architecture to provide a specific view of the epistemic content in the document.

## Document architecture: Uses and Definitions

There is no absolute agreement on how the term “document architecture” is used in the literature that addresses it, but mostly it is connected to a conception within text processing of documents as being structured into logical parts. The earliest paper reported in the ISI Web of Science as discussing document architecture describes an architecture called COBATEF, the context-based text formatting system (Peels, Janssen & Nawijn 1985). It is a system that automatically recognises text elements in machine-readable documents and formats them according to type. It can work either with documents that already contain some form of markup, or it can process the text and recognise certain basic text components based on context, such as for instance punctuation (Peels, Janssen & Nawijn 1985, 348, 363). The system works with a process-based model where structuralisation of the document results in a logical structure, which then goes through a process of formatting to gain a physical structure in preparation of presentation. The division into logical and physical structures is one that is also made in the two most well-known standards operating with the concept of document architecture, the Open Document Architecture, ODA [14] (ISO 8613 1986) and the family of languages related to the Standard Generalized Markup Language, SGML (ISO 8879-1986) and the Extensible Markup Language, XML.

One thing that distinguishes these two standard complexes is that they are focusing on interoperability, on a platform- and software-independent document transfer between systems. Both of them also constitute abstract document architectures, which require the individual implementer or group of implementers to create more detailed application profiles. ODA is directed towards the production, storage, and exchange of office documents created by word processors and desktop publishing programs. Much of the research conducted in the late 1980s and the 1990s concerned extending the ODA model to handle, for example, multimedia content, maps, and video (e.g. Kameyama & Tominaga 1989; Kameyama, Hanamura & Tominaga 1991; Lubich 1991; Appelt & Scheller 1995; Huang, Chu & Chang 1996). An automated document processing system, along the same lines as COBATEF, is suggested by Chiu Yu, Yuan Tang & Ching Suen (1993). However, these researchers completely leave the idea of marking up logical structures, focusing instead on a mathematically based language for identifying vertex points of rectangular and non-rectangular pattern segments in the document.

Hannes Lubich (1991) reviews a number of German studies on differences between ODA and SGML, and concludes that while SGML has focus on the logical structure (with related standards taking care of the presentational format), and is readable and processable by humans, ODA has focus on layout and requires machine reading and processing (Lubich 1991, 60). SGML and its subsystem XML have been very successful in the areas of publishing and have become widely known to the general public with the breakthrough of the Internet. In the ISO Standard for SGML, document architecture is defined as “(r)ules for the formulation of text processing applications” (ISO 8879-1986, 10). The document architecture thus serves a very instrumental purpose, and, in line with Lorin’s description above, concerns primarily the relationship between the different components of the SGML document and profiles for exporting them. As was noted above, the use of document architecture in SGML is also based in a primarily logical understanding of the document as a “collection of information that is processed as a unit” (ISO 8879-1986, 10) [15]. More specifically, an SGML document is described as consisting of

*data characters, which represent its information content, and markup characters, which represent the structure of the particular data and other information useful for processing it (ISO 8879-1986, 18).*

In ODA, where document architecture is defined as “the set of rules for structuring an interchanged document” (ISO/DIS 8613 1986, 1:6.2), a distinction is made between logical and layout structure (ISO/DIS 8613 1986; Campbell-Grant 1999, 273) [16]. These structures are described as follows:

*(t)he logical structure divides and subdivides the content of a document into increasingly smaller parts on the basis of the human-perceptible meaning of the content, e.g. into chapters, sections, paragraphs, figures. The layout structure divides the content into sets of pages, individual pages and areas within pages such as columns. Each structure represents a different but complementary view of the content of a document. Each structure is a hierarchy of objects, represented by sets of attributes that define the properties of the objects and their relationships (ISO/DIS 8613 1986, 1:6.2).*

It is evident here, that ODA is mainly focused on documents created in a word processor environment, whose intended end-result is in the print medium. The textuality expressed in the quotation is further an example of the so-called OHCO, the view of text as an “ordered hierarchy of content objects” (Renear 1997, 118) [17]. OHCO textuality has been criticised because it restricts the interpretive views that can be manifested of a text. If the logical structures overlap, it is not possible to show this without violating the hierarchical and linear order. One example from SGML of when this could become a problem is where an editor is interested in marking up the text of a verse play both according to who speaks the lines and according to metrics or rhymes or lines.

SGML and XML distinguish between logical structures and physical structures. Logical structures are referred to in the XML Recommendation as “declarations, elements, comments, character references, and processing instructions” [18] (Bray et al. 2004, 2) and a set of rules for describing their syntactic relationship. It is then a matter for every XML application to specify the semantics “in the sense of element and attribute names” that can be used to mark up the text. Although logical structures are not the same in SGML/XML as in ODA, the overall similarities are apparent.

The term “physical structure” is used for describing the entities (units) that make up a document, and is not the equivalence of ODA’s layout structure [19]. An entity can be described as “a placeholder for content, which you declare once and can use many times” (Ray 2001, 45). This describes one type of entity, the parsed entity. A simple example could be if one were expecting to have to write, for instance, the full name of SSLIS [20] many times. An entity could be declared (such as SSLIS for the long phrase “The Swedish School of Library and Information Science”) and subsequently referred to in the text (by writing “&SSLIS;”). The parser would exchange the entity reference for the entity content (the replacement text) when displaying it. Another type of entity is that which, for example, holds the place for an imported file or part of a file, such as an image file (Ray 2001, 45 f.) This type is called an unparsed entity (Bray et al. 2004, 4). Finally, all XML documents also have a document entity which is where the XML parser starts its processing. The document entity could contain the entire document, but

it could also have references to other entities (as explained above) (Bray et al. 2004).

After this exposition of how the document architecture concept is understood in computer science, I will turn to how the concept may be reformulated to work as an analytical tool in document analysis. In doing so, I will draw on some of the fundamental distinctions that have been outlined above.

## An Analytical Model of Document Architecture

I try in this section to connect document architecture less to a specific system or standard, while still not straying too far from how the term is used in SGML and ODA. The view of a document architecture that I adopt here is that it consists of a set of structures at different levels, and the interaction within, and to some extent between, these structures. Document architecture can then be used as a model for studying how a work is organised and manifested in a material artefact (Dahlström & Gunnarsson 2000; Dahlström 2006, 81). This means that it is just as natural to discuss the structural rules found in a style guide for academic writing in terms of document architecture as it is to do so with markup languages. For instance, the logical structure is quite clear in this advice from The Harbrace College Handbook for how to model a formal outline of a term paper:

```
Thesis:
I. Major idea
  A. Supporting idea
    1. Example or illustration for supporting idea
    2. Example or illustration for supporting idea
      a. Detail for example or illustration
      b. Detail for example or illustration
  B. Supporting idea
II. Major idea (Hodges et al. 1990, 383)
```

Similarly, style guides often contain very specific instructions for layout structures, based on such logical elements as paragraphs and headings. In markup languages based on SGML/XML, logical structures are marked up explicitly with tags, but it is also possible for a human being to interpret logical structures based on how they are displayed in a presentation medium through various visual codes (e.g. Gunnarsson 1997). This is possible because we have learnt to connect a number of the most common layout structures with logical structures, so that we can, for instance, distinguish between

ordinary paragraphs and block quotes. However, this can cause trouble when we are confronted with a set of conventions with which we are not familiar, for example because they belong to a community of practice of which we are not members. This could be the case for a Swedish reader of a German book, where new paragraphs are not marked by an indentation or by extra space between the paragraphs, but merely by the text beginning on a new line, which may result in a shorter last line of the paragraph (Hellmark 2000, 86).

As explained above, document structures of different types and at different levels can be identified in a document. The structures offer complementary views of the interaction between epistemic content and materiality in the document. In this article, I will distinguish between four levels of structures that are part of the document architecture: logical structures, layout structures, content structures, and file structures. An example of how these aspects can be operationalised is discussed further in Francke (2008, ch. 5).

Logical structure is understood here in much the same way as in the quotation from ODA above, namely as dividing and subdividing “the content of a document into increasingly smaller parts on the basis of the human-perceptible meaning of the content, e.g. into chapters, sections, paragraphs, figures” (ISO/DIS 8613 1986, 1:6.2) [21]. These parts can be explicitly identified in some way through markup or templates, or visually identified through layout. In digital documents, the correspondence or divergence between the markup/template and the visual layout provides for interesting study. In particular in HTML documents, the differences can be significant and can play an important part in how the logical markup can be used for automated tasks (Hansson et al. 2003).

Layout structures are closely associated with logical structures and are connected to their presentation. Layout structures can be formalised as in stylesheet languages or in guidelines for “the physical characteristics of the printed manuscript”, as the heading reads in *The MLA Style Manual and Guide to Scholarly Publishing*, where the positioning of different logical elements is specified (Gibaldi 1998, 128 ff.). These formalised layout structures are expressions of institutional agreements on what are appropriate or acceptable layout structures in certain situations, and they can be found in digital stylesheets, markup, templates, or in realised form on the page, screen, and so on [22]. Logical structures and layout structures are often closely interdependent and thus tightly interlinked. It is interesting to study, for instance, which elements are given much prominence on a page or screen in relation to other elements and how the different logical elements are displayed in relation to each other. Layout structures also take the form mentioned in the ODA definition quoted above, that is, controlling the

different areas of a document, such as pages, columns, and frames.

To discuss the text of a document in terms of content structures has been quite popular in connection with the Semantic Web (see e.g. Berners-Lee, Hendler & Lassila 2001). One of the examples mentioned in SGML/XML contexts, where the idea of marking up texts according to content or semantics is regarded as promising, is the marking up of product and personal names, for example in terms of <<doctor>, <medicine> or <chemical formula>. The relationship between different types of content is in focus for the work on Topic Maps, which can be viewed as one attempt to formalise content structures (Pepper & Moore 2001). HTML is mainly constructed to describe a document's logical structures, but it does contain some possibilities to mark up and define specific types of metadata [23]. These are often not rendered in the web browser, but if they are included in the HTML file, they are available to systems that allow field searches, to browsers that include a function for displaying metadata separately, or directly to users through an ASCII/Unicode interface. This sort of metadata may be considered to be marked up according to content. An alternative view is to see these elements as part of the logical structure. This can be particularly relevant in cases where the metadata are displayed in the browser window, as is the case with the <title> element. Whether one wishes to consider marked up metadata as part of the logical or the content structure is a question of perspective, of whether, for the specific purpose, the metadata are considered part of the logical text or the bibliographical paratext. An argument for the latter could be that the markup is governed by an intellectual interpretation of epistemic content as well as the function of the marked up text, and that the function it serves in this context is slightly different from that of the logical markup.

Furthermore, the various files that together constitute a document can be considered to form a structure at the file level [24]. An ordinary HTML-based web page, for instance, may consist of an HTML file, several GIF image files, a reference to the HTML document type definition (DTD), and a stylesheet, which provides the rules that give the browser instructions on how to display the document on the screen. The structure formed by the files that make up a document can be quite complex and there are clear connections to the more complex concept of physical structures in SGML/XML. If focus for the document analysis is on the composition of a document's entities, SGML/XML can provide a basis for a more advanced study of physical structures. The study of file structures is primarily of interest when studying digital documents and can be quite technical. Concerning print and manuscript documents, Dahlström offers some examples that may be thought of as equivalent, among them the distinction between presentation medium and signs such as paper, ink, and pencil lead, as in a book where a reader has made her own annotations (2006, 82 f.) [25].

What the viewer sees in the presentation medium could be considered the realisation of these different types of structures. I will term this realisation of structures “texture” [26]. In this context, the terms structure and texture are used metaphorically, with an emphasis on structure as “the coexistence in a whole of distinct parts having a definite manner of arrangement” (OED), and on texture as “the visual or tactile surface characteristics and appearance of something” (Merriam-Webster Online). In this way, the logical, layout, content, and file structures of a document are accessible to us through its texture, which may be more or less informative or helpful in our interpretation of the text. For instance, when viewing a HTML file in Windows Notepad, there may be white space which will help us distinguish between words, and white space can also facilitate the interpretation of, for example, logical structures by visually creating separate elements. If the same file is viewed in an HTML editor, the program may facilitate our navigation in the text by colouring tags, attributes, and content differently. If viewed, finally, in a web browser, the markup that governs the visual presentation may cause the browser to show a complex visual layout with images, colours, text in different fonts and sizes, and so on (as indeed will WYSIWYG editors). Thus, the structure is realised as different textures depending on the technology used to view the files. I would therefore argue that “realised” layout is part structure and part texture; part organisation of the elements in an information space (for example in the form of stylesheet instructions) and part realisation of this organisation in a presentation medium. Realised layout deals with both the mirroring of document structures and with the artistic design of their presentation. Depending on the relative emphasis put on each of these two aspects, the layout may be perceived as simple, straightforward, visually complex, or perhaps even ambiguous, dull, or incomprehensible.

We need to take document structures “logical, layout, content, and file” into consideration when creating a document. But it is also possible and, I would argue, interesting to examine the design of these structures in existing documents and in that way to study how document architectures are implemented in practice, not only how different standards and guidelines prescribe or recommend that they should be implemented. Which architectural styles do the documents that make up our “information environment” represent? Knowledge of document architecture has been used to some extent for improving information retrieval in marked up document collections (see e.g. INEX 2005), but in these cases the document architecture has often been very similar (and thus predictable) in the included documents. To study document architectures in less predictable documents, such as the open access scholarly journals I had in focus in (Francke 2008), presents a different type of challenge and allows for gaining knowledge of what characterises these documents.

## Conclusions

The aspects of information architecture and document architecture that have been outlined here, and that can form the basis for an architectural document analysis, can be summarised thus:

```
document architecture
logical structures
layout structures
content structures
file structures
```

```
information architecture
organisation system
organisation schemes
organisation structures
navigation
browsing
searching
labelling
```

These aspects are the result of my interpretation of how information architecture and document architecture can be modelled in order to serve analytical purposes. In my interpretation, I have begun in the notion of a document as a material artefact (Chap. 3 in Francke 2008) rather than in the underlying document views of the respective traditions. It is my impression that a material document concept agrees well with the ideas behind both information architecture and document architecture and that stressing materiality serves to open up the use of these tools to other media than the digital one. Constructing a document includes shaping the design of its document and information architectures. But our interaction with documents is also influenced by their architectures. Most people in the industrialised world engage in different forms of documentary practices on a daily bases. Often, documents we interact with belong to specific document types or genres, such as newspapers, blueprints, or scholarly articles. That restricts the variation in document and information architecture we come to expect from the documents and when their social and technological circumstances change, we may be facing a change in architecture which in turn may come to influence our documentary practices. The emergence of

new genres and document types as well as changes to existing ones are reason for analysing document structures “ material aspects of documents ” in order to gain further understanding of artefacts that are an important part of many people’s information practices.

## References

- Alvesson, M., Sköldbörg, K. (1994). *Tolkning och reflektion: vetenskapsfilosofi och kvalitativ metod* [Interpretation and Reflection: Philosophy of Science and Qualitative Method]. Lund: Studentlitteratur.
- Appelt, W., Scheller, A. (1995). “HyperODA: Going Beyond Traditional Document Structures”. *Computer Standards and Interfaces* 17.1: 13–21.
- Berners-Lee, T., Hendler, J., & Lassila, O. (2001). “The Semantic Web”. *Scientific American* May. [http://www.sciam.com/print\\_version.cfm?articleID=00048144-10D2-1C70-84A9809EC588EF21](http://www.sciam.com/print_version.cfm?articleID=00048144-10D2-1C70-84A9809EC588EF21).
- Bernstein, M. (1998). *Hypertext Gardens: Delightful Vistas*. Eastgate Systems. <http://www.eastgate.com/garden/>
- Bowker, G. C., Star, S. L. (1999). *Sorting Things Out: Classification and Its Consequences*. Cambridge, MA. The MIT Press.
- Bray, T., Paoli, J., Sperberg-McQueen, C. M., Maler, E., & Yergeau, F. (2004). *Extensible Markup Language (XML) 1.0 (Third edition): W3C Recommendation 04 February 2004*. W3C. [2005-05-22]
- Campbell-Grant, I. R. (1999). “Introducing ODA”. *Computer Standards and Interfaces* 20.4–5: 269–278.
- Chowdhury, G. G. (1999). *Introduction to Modern Information Retrieval*. London: Library Association.
- Coplien, J. O. (1999). “Reevaluating the Architectural Metaphor: Toward Piecemeal Growth”. *IEEE Software* 16.5: 40–44.
- Coplien, J. O., Devos, M. (2000). *Architecture as Metaphor*. Proceedings of the World Multiconference on Systemics, Cybernetics and Informatics, Orlando, Florida, July 2000. 737–742. <http://users.rcn.com/jcoplien/SCI2000Arch.html>.
- Cowling, D. (1998). *Building the Text: Architecture as Metaphor in Late Medieval and Early Modern France*. Oxford: Clarendon Press.
- Dahlström, M. (2006). *Under utgivning: den vetenskapliga utgivningens bibliografiska funktion* [The Editor’s Text: Bibliographic Functions in Scholarly Editing]. (Skrifter från Valfrid, 34). Borås: Valfrid.
- Dahlström, M., Gunnarsson, M. (2000). “Document Architecture Draws a

Circle: On Document Architecture and Its Relation to Library and Information Science Education and Research". *Information Research* 5.2. <http://InformationR.net/ir/5-2/paper70.html>.

Dillon, A. (2002). "Information Architecture in JASIST: Just Where Did We Come From?" *Journal of the American Society for Information Science and Technology* 53.10: 821-823.

Foucault, M. (2002). *The Archaeology of Knowledge*. [1969]. London: Routledge.

Francke, H. (2008). (Re)creations of Scholarly Journals: Document and Information Architecture in Open Access Journals. Valfrid: Borås. <http://hdl.handle.net/2320/1815/>.

Frohmann, B. (2004). *Deflating Information: From Science Studies to Documentation*. Toronto: University of Toronto Press.

Gibaldi, J. (1998). *MLA Style Manual and Guide to Scholarly Publishing*. 2nd ed. New York: Modern Language Association of America.

Gunnarsson, M. (1997). "Introduktion till textkodning (mark-up languages) [Introduction to Text Coding (Markup Languages)]." University Borås. <http://www.adm.hb.se/personal/mg/ko2/textk.htm>.

Hansson, J., Francke, H., Dahlström, M. & Gunnarsson, M. (2003). "Documents in Library and Information Science: Sociotechnical Dimensions in Document Genre and Architecture Studies." Paper presented at DOCA M'03, Berkeley, 13-15 August, 2003. <http://thedocumentacademy.org/resources/2003/papers/boras.paper.html>.

Head, A. J. (2001). "Information Specialists at the Intersection of Information Architecture and Usability." Lecture delivered at Florida State University School of Library and Information Science, November 10, 2001. <http://www.ajhead.com/lecture.html>.

Hellmark, C. (2000). *Typografisk handbok [Typographical Handbook]*. 3rd ed. Stockholm: Ordfront.

Hjørland, B. (2004). "Theory of Knowledge Organization and the Feasibility of Universal Solutions." Eighth International ISKO Conference, London, Friday July 16th 2004, Session 9B. <http://www.ucl.ac.uk/isko2004/sysweb/9bHjorland.ppt>.

Hodges, J. C., et al. (1990). *Harbrace College Handbook for Canadian Writers*. 3rd ed. Toronto: Harcourt Brace Jovanovich.

Huang, C.-M., Chu, Y.-F., & Chang, Y.-I. (1996). "An ODA-like Multimedia Document System." *Software: Practice and Experience* 26.10: 1097-1126.

INEX (2005). "Initiative for the Evaluation of XML Retrieval." DELOS Network of Excellence for Digital Libraries. <http://inex.is.informatik.uni-duisburg.de/2005/>.

Information Architecture Institute (2005). "About us." [http://iainstitute.org/pg/about\\_us.php](http://iainstitute.org/pg/about_us.php).

ISO/DIS 8613/1-6 (1986). Information Processing — Text and Office Systems — Office Document Architecture (ODA) and Interchange Format.

ISO 8879-1986 (1986). Information Processing — Text and Office Systems — Standard Generalized Markup Language (SGML). 1st ed.

Kameyama, W., Hanamura, T., & Tominaga, H. (1991). "A Proposal of Multimedia Document Architecture and Video Document Architecture." IEEE International Conference on Communications, 1991, ICC '91, Conference Record, 23-26 June 1991. 511-515.

Kameyama, W., Tominaga, H. (1989). "Extended Document Architecture for Maps." GLOBECOM '89, "Communications Technology for the 1990s and Beyond," IEEE Global Telecommunications Conference, 27-30 November 1989. Vol. 2. 975-979.

Karlsson, L., Malm, L. (2004). "Revolution or Remediation? A Study of Electronic Scholarly Editions on the Web." *Human IT* 7.1: 1-46. <http://www.hb.se/bhs/ith/1-7/lklm.pdf>.

Khalfallah, H., Karmouch, A. (1994). "Architecture and Synchronization of Multimedia Data for Presentational Applications." GLOBECOM '94, "Communications: The Global Bridge," IEEE Global Telecommunications Conference, 28 November-2 December 1994. Vol. 2. 891-895.

Kittay, E. F. (1989). *Metaphor: Its Cognitive Force and Linguistic Structure*. Oxford: Clarendon Press.

Landow, G. P. (1997). *Hypertext 2.0: The Convergence of Contemporary Critical Theory and Technology*. 2nd rev. ed. Baltimore, MD: Johns Hopkins University Press.

Large, A., Beheshti, J., Cole, C. (2002). "Information Architecture for the Web: The IA Matrix Approach to Designing Children's Portals." *Journal of the American Society for Information Science and Technology* 53.10: 831-838.

Levy, D. M. (2001). *Scrolling Forward: Making Sense of Documents in the Digital Age*. New York: Arcade Publishing.

Lorin, H. (1986). "Systems Architecture in Transition: An Overview." *IBM Systems Journal* 25.3-4: 256-273.

Lubich, H. P. (1991). "A Proposed Extension of the ODA Document Model for the Processing of Multimedia Documents." Proceedings of TRICOMM '91, "Communications for Distributed Applications and Systems," IEEE Conference on Communications Software, 18-19 April 1991. 59-72.

Lynch, P., Horton, S. (2002). *Web Style Guide: Basic Design Principles for Building Web Sites*. 2nd ed. New Haven: Yale University Press. <http://www.webstyleguide.com/>.

Maasen, S., Weingart, P. (2000). *Metaphors and the Dynamics of Knowledge*. London & New York: Routledge.

Morrough, E. (2003). *Information Architecture: An Emerging 21st Century Profession*. Upper Saddle River, NJ: Prentice Hall.

- Morville, P. (2002). "The Definition of Information Architecture." *Semantics: Peter Morville's Column about Information Architecture and Strategy*. Semantic Studios. <http://semanticstudios.com/publications/semantics/00010.php>.
- Muhlhausen, J. (2006). "Wayfinding is not Signage." *SignWeb.com*. <http://signweb.com/index.php/channel/6/id/1433>.
- Peels, A. J. H. M., Janssen, N. J. M., & Nawijn, W. (1985). "Document Architecture and Text Formatting." *ACM Transactions on Office Information Systems* 3.4: 347-369.
- Pepper, S., Moore, G., eds. (2001). *XML Topic Maps (XTM) 1.0*. TopicMaps.Org Specification. TopicMaps.Org. <http://www.topicmaps.org/xtm/1.0/>.
- Periasamy, K. P., Feeny, D. F. (1997). "Information Architecture Practice: Research-based Recommendations for the Practitioner." *Journal of Information Technology* 12: 197-205.
- Platon (2000). *Gorgias*. Skrifter: Bok 1 [Works: Book 1]. Stockholm: Atlantis.
- Ray, E. T. (2001). *Learning XML*. Sebastopol, CA: O'Reilly.
- Reiss, E. L. (2000). *Practical Information Architecture: A Hands-on Approach to Structuring Successful Websites*. Harlow: Addison Wesley.
- Renear, A. (1997). "Out of Praxis: Three (Meta)Theories of Textuality." *Electronic Text: Investigations in Method and Theory*. Ed. Kathryn Sutherland. Oxford: Clarendon Press. 107-126.
- Rosenfeld, L., Morville, P. (2002). *Information Architecture for the World Wide Web*. 2nd ed. Sebastopol, CA: O'Reilly.
- Toms, E. G. (2002). "Information Interaction: Providing a Framework for Information Architecture." *Journal of the American Society for Information Science and Technology* 53.10: 855-862.
- Tosca, S. P. (2000). "A Pragmatics of Links." *Journal of Digital Information* on 1.6. <http://jodi.tamu.edu/Articles/v01/i06/Pajares/>.
- Tuominen, K., Talja, S., Savolainen, R. (2003). "Multiperspective Digital Libraries: The Implications of Constructionism for the Development of Digital Libraries." *Journal of the American Society for Information Science and Technology* 54.6: 561-569.
- Van Dijck, P. (2003). *Information Architecture for Designers: Structuring Websites for Business Success*. Mies: RotoVision.
- Weinreich, H., et al. (2006). "Off the Beaten Tracks: Exploring Three Aspects of Web Navigation." *Proceedings of the 15th International Conference on World Wide Web*, Edinburgh, Scotland. New York: ACM Press. 133-142.
- Wodtke, C. (2002). *Information Architecture: Blueprints for the Web*. Indianapolis, IN: New Riders.

Wodtke, C. et al. (2001). "Defining the Damn Thing." *Elegant Hack* Web blog June 2001. <http://www.eleganthack.com/blog/archives/00000069.html>.

Yu, C. L., Tang, Y. Y., & Suen, C. Y. (1993). "Document Architecture Language (DAL) Approach to Document Processing." *Proceedings of the 2nd International Conference on Document Analysis and Recognition*, 20-22 October 1993. 103-106.

Ørom, A. (2003). "Knowledge Organization in the Domain of Art Studies: History, Transition and Conceptual Changes." *Knowledge Organization* 30.3/4: 128-143.

## Footnotes

[1] The text is republished with permission from the publisher, Publiceringsföreningen Valfrid.

[2] This is the term Bernd Frohmann prefers to information or meaning. He defines it in relation to a scientific document as "what we grasp when we understand a sentence, diagram, graph, data set, computer-generated image, or any truth-telling inscription in any media form" (2004, 24).

[3] e.g. Kittay (18) for a discussion of how metaphors transfer "relations which pertain within one semantic field to a second, distinct content domain" (18, 36).

[4] See Coplien & Devos (2000) for a discussion on similarities between the role of building architects and software architects. Both in software architecture (Coplien, 1, 41) and in information architecture (Morrogh, 2003, 3), the analogy seems to focus more on the architects' tasks and concerns than on the architecture of the artefact. For a critique of the metaphor's applicability in software architecture, see Baragry & Reed (2001).

[5] For approaches to document and information architecture within Library and Information Science, see e.g. the two special issues of *Journal of the American Society for Information Science and Technology* (on Document Architecture vol. 48, no 7 (17) and on Information Architecture vol. 53, no 10 (2002)).

[6] Information architecture is usually discussed in connection with large web sites, but it is possible to use the metaphor for other media as well (see Morrogh, 2003; Dillon, 2002, 823), in which case it occasionally comes closer to document architecture. There are also other views of information architecture than the one discussed here. For instance, in business administration, information (systems) architecture can be understood as "a set of high level models which complements the business plan in IT-related matters and serves as a tool for IS planning and a blueprint for IS plan implementation" (Periasamy & Feeny, 17, 18)

[7] This is the aspect of information architecture that Andrew Dillon has called "little IA" (2002, 822). I do not claim that this is all that information architecture is, but it is an important aspect and the one most relevant to the current project. There

are other aspects of information architecture not mentioned here, such as user-centred and graphic design. In some contexts the information architect is responsible for developing all the ICT structures in an organisation, not only the web site.

[8] Van Dijck (2003) mentions time-based and geographical organisation.

[9] Van Dijck (2003) mentions subject-based/topical, task-based and audience-based schemes.

[10] Just like the “random access” of a book, the possibility to open it on any page, is beyond the planning of author and publisher.

[11] Susana Pajares Tosca (2000) suggests a “pragmatics of links”, which may incorporate both of the standpoints described above. She identifies two different types of approaches to planning hypertext links. One achieves the effect described above as desired by information architects where, in Tosca’s words, the goal is “Minimal processing effort + Maximal (informational) cognitive effects”. The other type of approach, more appropriate for what Bernstein is after, strives to gain “Increased processing effort + Maximal (lyrical) cognitive effect” (Tosca, 2000, section 5).

[12] E.g. navigation systems within subsites on the web site.

[13] Although see Bernstein’s (18) position discussed above, and Rosenfeld & Morville (18, 5 ff.).

[14] In early versions, the O stood for Office.

[15] No definition of “information” is given.

[16] A similar distinction is made in Khalfallah & Karmouch (14, 81).

[17] This is also the case with SGML/XML.

[18] Declarations can be used to refer e.g. to a DTD; comments refer to the possibility of including notes that are not parsed by the XML parser; a character reference “refers to a specific character in the ISO/IEC 10646 character set” (Bray et al., 2004, 4.1); and a processing instruction is “a container for data that is targeted toward a specific XML processor” (Ray, 2001, 55).

[19] In SGML and XML, layout and presentation is handed over to a separate style language, such as DSSSL, CSS, or XSLT.

[20] The Swedish School of Library and Information Science at the University of Borås and Göteborg University.

[21] Content in this case can be understood as inscriptions or functions in various modes of representations.

[22] See also the discussion on texture and layout below. The DTDs or schemas of

markup languages are similarly expressions of institutional agreements.

[23] There are still in HTML also elements whose function is to control layout structures, but these are deprecated.

[24] This is similar to what Dahlström terms document layers (2006, 82). He notes, quite correctly, that in some cases it may be useful to consider the individual files as documents in their own rights. In fact, electronic documents pose a number of questions that spring from the fact that their storage and presentation media are different than, for example, print. As both the proper hardware and software are needed in order to render a computer file readable, David Levy reflects that “(u)nder such circumstances, is the file really ‘the document’? Or should I say that the document consists of the file plus the requisite technical environment? Or must I also include the perceptible forms as well?” (2001, 157; cited in Dahlström, 2006, 72, note 214).

[25] In fact, when we discuss print documents, document layers is probably a more suitable term to use than file structures. It does, however, have a broader span than file structures, and also includes aspects that would here be considered part of the layout structures.

[26] In much the same way as Foucault uses the term (Foucault 2002, 115).

## Cite as

Francke, H. (2009) Towards an Architectural Document Analysis. *Journal of Information Architecture*. Vol. 01. Iss. 01. Pp. 19–44. <http://journalofia.org/volume1/issue1/03-francke/>.

