An ecological look at scholarly documentation

by Magnus Enger

DOK-3021 Spesialstudium: dokumentinstitusjon

Vår 2004, Universitetet i Tromsø

Version 1.1 (2004-11-23)



See: http://creativecommons.org/licenses/by-nc-sa/2.0/

Changelog

- Version 1.0 (2004-05-22) http://www.enger.priv.no/dokvit/spes-1.0.pdf
 - · Original version
- Version 1.1 (2004-11-23) http://www.enger.priv.no/dokvit/spes-1.1.pdf
 - Candidate-number replaced by full name on front page
 - Version-number added on front page
 - CC-license updated to http://creativecommons.org/licenses/by-nc-sa/2.0/
 - · Changelog added
 - Index updated

Table of contents

Changelog	2
Introduction	4
Documents and memes	4
What are memes?	4
Documents as carriers of memes	5
The ecology of scholarly documentation	6
What is ecology?	6
Ecosystems	6
Three genera of scholarly documentation	7
Primary documents	8
Secondary documents ("metadata")	8
Systemic documents	9
Relationships	10
Actors	11
The documentary environment	12
Energy transfer and food chains	12
Habitats, niches and niche-construction	14
Co-evolution	15
Succession	16
Why?	16
Abbreviations and acronyms	18
References	19

Introduction

In Enger (2003), I proposed that documents could be regarded as carriers of "memes", units of cultural transmission or imitation. I elaborated on some concepts connected with evolution, and tried to connect them with the concept of documents and documentation, as set forth by Lund (2001). In this essay I will continue along the same lines, but the main emphasis will be on an analogy to ecology, the study of the relationship between organisms and their environment.

It is very common for authors to borrow from the biological nomenclature; the terms "to evolve" and "evolution" seem to be the most popular. But these words are often used as mere synonyms for "change", without any apparent regard for the theory behind them. Part of the rationale for my approach in this essay is to see what happens if one tries to take "evolution" and some of its attendant terms seriously, and apply them to the study of documentation in general, and the study of scholarly documentation in particular.

Documents and memes

When thinking about evolution we are used to think of genes as the basic unit at the heart of the processes that make up evolution.¹ Genes, in cooperation with the environment, determine the shape of individuals, and the succession of individuals over time constitute a species.

What are memes?

In 1976 Richard Dawkins introduced another kind of basic evolutionary unit, in order to argue that genes are not the only entities undergoing evolutionary processes. He labelled this new unit a "meme":

We need a name for the new replicator, a noun that conveys the idea of a unit of cultural transmission, or a unit of imitation. 'Mimeme' comes from a suitable Greek root, but I want a monosyllable that sounds a bit like 'gene'. I hope my classicist friends will forgive me if I abbreviate mimeme to meme. [...] Examples of memes are tunes, ideas, catch-phrases, clothes fashions, ways of making pots

¹ This is made very explicit by Dawkins (1989), with the title *The selfish gene*, in which the author stresses that evolution can be most fully understood by viewing genes as if they have a will of their own and behave selfishly.

or building arches. Just as genes propagate themselves in the gene pool by leaping from body to body via sperms or eggs, so memes propagate themselves in the meme pool by leaping from brain to brain via a process which, in the broad sense, can be called imitation. (Dawkins 1989 [1976] p. 192)

As we can see, Dawkins stresses *imitation* as a central concept for the propagation of memes, but other writers have broadened the view of memetic processes somewhat, to include artefacts as spreaders of memes:

[...] memes can interact with artefacts as well, in their efforts to find new hosts. [...] The simple model of memetic replication through communication becomes more complex when what might be called a 'communicative artefact' steps into the middle of the communication process. In this case message-senders create artefacts rather than signals (e.g., written messages rather than speech). [...] a new meme appears in the recipient brain which is causally connected to the source meme through the information provided by the message-catalyst. (Aunger 2000 s. 220-221)

I will base my suggestions for an ecological view of documentation on this somewhat expanded view of artefacts as potential spreaders of memes.

Documents as carriers of memes

Hull (2001) has generalized the central concepts of evolution and labelled the three levels that we recognize from biological evolution as "replicators, interactors and lineages". In human cultural evolution the basic unit or replicator is Dawkins' meme, as outlined above. In my first essay I argued that documents, as defined by Lund (2001), can be seen as carriers of memes, their shape determined by an interplay between their memetic payload and their psychological, social, cultural, and physical environment. Over time the succession of documents will form lineages corresponding to "forms of documentation":

General concepts:	Replicators	Interactors	Lineages
Genetics:		Individuals	Species
Science of documentation + memetics:	Memes	Documents	Forms of documentation

² Dawkins (1989) uses the term "vehicle" instead of "interactor", but I will stick to Hull's terminology, since it gives a clearer impression of what actually goes on at this level: "Interactors are those entities that interact as cohesive wholes with their environments in such a way as to make replication differential." (Hull 2001 s. 22) We will see that this interaction lies at the heart of the ecological view of documentation.

In this essay I will focus on documents as interactors, explore what constitutes their environment and how they interact with it; in ways that give rise to a differentiated survival of the memes that they carry and that helped determine their shape. In other words, I will propose that we can study documentation and its environment in a way similar to the way ecologists study living organisms and *their* environment.

The ecology of scholarly documentation

What is ecology?

Chapman and Reiss define ecology as:

[...] the study of organisms in relation to the surroundings in which they live. These surroundings are called the environment of the organism. This environment is made up of many different components, including other living organisms and their effects, and purely physical features such as the climate and soil type. (Chapman and Reiss 1999, p. 2)

I will propose that documents can be substituted for organisms to yield a definition like this:

The ecology of documentation is the study of documents in relation to the surroundings in which they are created and used. These surroundings are called the environment of the documents. This environment is made up of many different components, including other documents and their effects, and non-documentary (e.g. cultural, social, psychological and physical) factors.

In this essay I will elaborate on some of the main points of this definition, and also try to sketch out some of the implications of this way of looking at documentation.

Ecosystems

I will suggest that the human activity of documentation can meaningfully be divided into separate ecosystems that involve a limited set of interrelated forms of documentation, groups of actors and relevant environmental factors. Examples of such ecosystems might be scholarly documentation, the making and distribution of children's books, the performance of traditional Balinese dance and so on and so forth.

When studying ecological relationships in the physical world it would be impossible, and

probably quite meaningless, to study the whole biosphere as one interrelated "system". As a consequence of this, most studies are done on some part of the world that make up a meaningful whole. Such a limited system is called an "ecosystem": "it appl[ies] to a whole community of organisms and its environment as one unit" (Chapman and Reiss 1999, p. 187). Since "no ecosystem is an island", but to some degree is dependent on factors outside itself (e.g. energy from the sun, at the very least), it will always be possible to question the limits that have been set for an ecosystem under study. On the other hand a limit has to be drawn somewhere, in order to yield a scale that can be meaningfully studied with the resources that are available to any given project.

Nardi and O'Day (1999) use the ecological metaphor in a way somewhat similar to what I am doing here, however, I think their use of the metaphor is problematic, because they use the term ecology on the wrong level. Their definition of an information ecology looks like this:

We define an information ecology to be a system of people, practices, values, and technologies in a particular *local* environment. (p. 49, my emphasis)

This definition gets into trouble when the authors consider the Internet:

There is no single Internet information ecology. Information ecologies are local habitations with recognizable participants and practices. Nothing as huge as the Internet can be seen in that light. We view the Internet as a set of environmental conditions that provide a substrate for the growth of ecologies that span traditional geographic or social boundaries. The Internet can serve as connective tissue between and within local information ecologies.

In biology there is no concept of "an ecology" - ecology is simply "the study of organisms in relation to the surroundings in which they live" (Chapman and Reiss 1999, p. 2). To make this meaningful one has to concentrate on some "local" part of the world, and this is termed an "ecosystem", not "an ecology". Below I will discuss terms that can be used to describe even finer divisions of a field of study, i.e. "habitat" and "niche".

Three genera of scholarly documentation

A natural point of departure for an ecological look at documentation is to identify some of the "organisms" that inhabit this ecology. In the area of scholarly documentation I think we can identify at least three different main groups of documentary forms. Instead of calling these groups species (which would result in an unwieldy plethora of "sub-species") I will borrow the term "genus" (plural form "genera") from the biological nomenclature, to designate a closely related group of species (forms of documentation).³ The three genera and some of their main characteristics are as follows:

Primary documents

This is what we would normally think of as "scholarly documentation": journal articles, books, chapters in books, reports and so on. These are created by scholars working alone or in teams, usually subjected to some form of quality control (e.g. peer review) and then published or in some other way made available to the world at large. These documents can be acquired directly by other scholars, but very often libraries (both as economic actors and as repositories) are positioned between producers and users/reproducers of such documents.

Secondary documents ("metadata")

Metadata is often described as "data about data", and here I am thinking of documents that have as their purpose the description (or documentation) of other documents. A classical example of this would be the catalogue-card from the old paper-based library catalogue, which describes (documents) the properties of one primary document or document-complex, with the aim of representing this document in some systematic context and facilitating finding it on the shelves of the library. Roughly the same way of describing primary documents will be found in modern, automated and database-based systems, where the catalogue-card is replaced by a "record" in the system. Another form of documentation with the same purpose is the abstract, which usually describes a primary document in a less formal and more content-oriented fashion.

In the electronic realm URLs (or more generally: URIs (Berners-Lee 2000, p. 67-68)) can be said to perform some of the same functions, in that it is a reference to the physical location of a document in a network. Variants of this have been created in e.g. the DOI and URN systems (Powell, 1998), which substitutes more logical "locations" for the physical locations of URLs.

^{3 &}quot;**genus** *n*. Taxonomic group of closely related species, similar and related genera being grouped into families." (Henderson 1989, p. 203)

Secondary documentation can be found as docemes in primary documents, e.g. as introductory abstracts or as parts of a bibliography/list of references.⁴ They also play a crucial role as docemes in the third genus:

Systemic documents

Systemic documents can be thought of as collections of secondary documents that are organized in such a way as to make retrieval easy, through some combination of searching and/or browsing.⁵ One example of such a system could be a bibliographic database that contains descriptions of and/or pointers to primary documents. Another example might be a scholarly "portal" to the Internet, such as the Norwegian "BIBSYS emneportal" (http://emneportal.bibsys.no/), which contains descriptions of websites, along with the URLs of these sites.

One important aspect of these documents is that they are not random collections of secondary documents, the constituent parts must be organized in some way that takes into account the fact that it should be easy to search through them in one way or another. Some sort of alphabetical ordering might be one way of accomplishing this (think of paper cards in an old library catalogue), the possibility of searching within all the fields of a bibliographic description in an automated catalogue might be another.

With the advent of so called "full text databases" (e.g. ProQuest, EBSCO EJS) we are also increasingly seeing primary documentation taking on the role of docemes in systemic documents. This has led to a change in what is actually searched, from just the descriptions of documents to the contents of the documents themselves (in addition to the descriptions).⁶

What can be seen as a third step in the evolution of systemic documents is the rise of the fully automated, purely full text-based "search engine", which can most prominently be found in the services that indexes and makes searchable parts of the publicly available Internet, such as Google (http://www.google.com/) or AlltheWeb.com (http://www.alltheweb.com/).

⁴ Lund (2001, p. 117) defines a docem as "part of a document" (my translation).

⁵ The term "systemic document" is borrowed from Austrheim (2003).

⁶ At least this is true for collections of text-based documents. Searching through the contents of images or sound-recordings has so far not been a viable option.

Relationships

Although these are just three genera of documentation described in a broad and general fashion, it is evident that there exists a complex set of relationships between them, the main outline of which will look something like this:

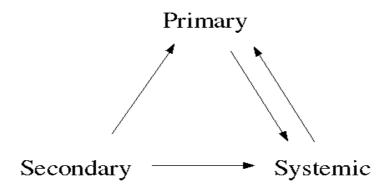


Figure 1: Relationships between the three genera of scholarly documentation

- Primary documents can contain:
 - Secondary documents, in the form of an introductory abstract or a bibliographic reference to the article or book itself.
 - Systemic documents, in the form of a bibliography. Such a list of references is of
 course built up of secondary documents. This phenomenon has been studied with great
 detail within the field of citation analysis.
- Systemic documents can contain:
 - Secondary documents, e.g. in library catalogues and bibliographic on-line services.
 - Primary documents. Over the last years commercially available bibliographic databases
 have increasingly made a transition to contain "full text", which means that the primary
 documents described by the secondary documents are available from within the
 service, at the click of a mouse-button.

Actors

The ecosystem that contains the forms of documentation that we are interested in studying will contain one or more actors, non-documentary entities that play an active role in the ecosystem, by providing input of resources into the system or by consuming resources produced by others, or a combination of these. These actors will also influence the environment through the values that they bring to the tasks they fulfil. We can easily identify several actors within the ecology of scholarly documentation (this list is of course not intended to be exhaustive):

- Scholars-as-authors. These are producers of the primary documentation. They can work
 independently of other scholars when they construct a document, or they can work in
 teams of two or more.
- Publishers. These can be big, commercial, profit-maximising publishing-houses with a portfolio of hundreds or thousands of titles, or it can be scholarly associations that run just one journal on a not-for-profit basis. No matter what their size or motivation is, these actors have traditionally provided the essential function of quality control, through editorial screening and administration of peer review systems, as well as copy-editing and proofreading. As computers have become ubiquitous in academia several of the functions traditionally filled by publishers (e.g. typesetting) have shifted to the authors.
- Libraries. A lot of scholarly journals are too expensive for individuals, and libraries have traditionally taken on the role as (one of) the major sources of income for publishers. Libraries have also provided functions such as selection, organization and long term archiving of primary documentation. Inter library loan has also been an important function of libraries, since this to a certain degree has helped alleviate the fact that it is (financially as well as physically) impossible for any library to acquire all the material that might be of interest to its users.

• Scholars-as-readers. These are the ultimate endpoints for the movement of documents that starts with the author. The interests of scholars-as-authors and scholars-as-readers are not necessarily the same. In the first capacity they want to publish as much as possible, while in the second capacity they express a wish for reading as little as possible, or at least only material of as high quality as possible (Mabe and Amin 2002).

The documentary environment

As stated in the definition above, other documents are important factors in the environment that helps give shape to any given document. But there are also other factors that come into play, factors that can not be characterized as documentary. Lund (2001) gives an idea of what these factors might be, when he stresses the importance of physical, social, cultural and mental factors. To these I would like to add economic factors, since these have a central role in many (but, significantly, not all) of the social transactions that are involved in scholarly documentation.

Energy transfer and food chains

When documentary species (or genera) have been identified, along with actors and environmental factors, it might be instructive to draw up a "map" of the different entities that make up the ecosystem, and how they interact and rely on each other, what values they produce and how the output of one actor is taken as the input of another. Such visualisations can be done with a high degree of detail, something along the lines of the work reported in Björk and Hedlund (2004), which contains "64 activity boxes and around 200 arrows" (p. 10). But visualisations that only take account of the most important entities and their functions and relationships can also be of value, depending on what is to be investigated.

The relationships in such a visualisation of an ecosystem should not be conceived as static. An ecosystem will tend to strive towards some sort of balance or equilibrium, but the numbers or functions of one of the actors may be altered as a result of internal or external factors, and the resulting change in output from this actor will have consequences for those actors that rely on this output. A map of the "energy transfer" in the ecosystem may then be used to predict how the changes will affect the other actors as well as the system as a whole.

It is important to note, however, that ecosystems do not necessarily attain equilibrium. There

are systems that show cyclic variations in populations. The best known example of this is probably the lemming-populations that grow dramatically every fourth year, leading to increases in the numbers of carnivores such as birds of prey and foxes, that utilise the abundant lemmings as an input of energy. Although this ecosystem never attains one balanced situation it is still useful to have a "map" of the participants if one wants to have some idea of how they influence each other.

I think the ecosystem of scholarly documentation has been in a state of relative equilibrium for quite a long time. Authors have submitted their manuscripts to publishers, who run it through a peer review-process, bundle the accepted articles into journal issues and ship these off to libraries, where scholars have consume them. But this situation has been disturbed by several factors over the last 15-20 years:

- The advent of the Internet, which makes it possible to disseminate scholarly
 documentation in new and very cost-effective ways. The Internet has also made it possible
 to question the role of the middlemen, i.e. publishers and libraries, and to envision a
 situation where scholars are completely in control of the documentation-processes, from
 creation to dissemination.
- The "journal crisis". There is some discussion about the realities of this crisis, or whether it should be called by some other name, but the basic dynamic of the situation seems to be that
 - library budgets are cut, resulting in cancellations of subscriptions
 - journals become ever more expensive in an attempt by the publishers to recover some of their dwindling revenue, due to library-cancellations
 - libraries have to cut more subscriptions because of the increases in subscription prices
 - and so on, in a vicious circle
- A political movement claiming that publicly funded research should be publicly available
 without having to pay for expensive subscriptions, and that developing countries should
 not be excluded from the communication of research results based on a lack of funding.

These factors are combining to destabilise the ecosystem, and only time will tell how and if

the system (including the different forms of documentation involved, the actors and the relationships between them) will change to regain a state of equilibrium under the new environmental conditions.

Habitats, niches and niche-construction

According to Chapman and Reiss (1999, p. 106) the relationship between a habitat and a niche is as follows:

In ecology [niche] came to stand for the precise way in which a species fits into its environment. A habitat is a description of where an organism is found, but its niche is a complete description of how the organism relates to its physical and biological environment.

Odling-Smee, Laland and Feldman (1996) add to this the important notion that organisms play a crucial role in determining the exact make-up of their own niches:

Organisms, through their metabolism, their activities, and their choices, define, partly create, and partly destroy their own niches. We refer to these phenomena as "niche construction". (p. 641)

One way in which this construction of niches manifests itself is the fact that new generations of organisms inhabit environments that have been altered by previous generations, and this results in its turn in altered evolutionary pressures. It is also possible to think of this as a complement to genetic inheritance, in that succeeding generations are shaped not only by the genetic material they have inherited, but also by the altered environment that they "inherit", as a result of the niche constructing activities of preceding generations.

One way of viewing niche-construction in connection with scholarly documentation might be in connection with "seminal works", works that help lay the foundations of new academic disciplines or that give rise to a literature of interpretation and explication. *On the origin of species* by Charles Darwin or *Understanding media* by Marshall McLuhan might be seen as examples of this. Because of the appearance of these documents (along with other factors) new academic departments and positions are created, and thus provide the material basis for the creation of new documents that would not have been possible without them. Such works can also form the basis for new documentary traditions, new "rules" for how and in what form documents that are to be accepted can be created. An indication that this is actually

happening in connection with the works of the German media-scholar Friedrich Kittler is provided by Winthrop-Young and Wutz (1999, p. xxxii):

[...] a growing number of younger scholars have modelled their writing on Kittler's very personal style: to the delight of connoisseurs of German academese, *Kittlerdeutsch* is already as distinct an idiom as the equally unmistakable *Adornodeutsch*.

This hints at a situation where the environment of new documents has been altered by those documents that went before it. An interesting question in this regard might be what factors influence the environment-altering capabilities of a document. Why are some documents excluded or not taken seriously because they do not conform to the generally accepted standards, while others actually alter the standards?

Co-evolution

One consequence of the fact that other documents are an important part of the environment of any given document, is that there is a mutual influence in how forms of documentation evolve over time, their shape turns out to be a different one from what it would have been if it had evolved without the presence of other forms of documentation.

Such effects would be expected in documentary forms that have evolved in close contact with each other, such as the three genera of scholarly documentation outlined above. One example of this might be seen in the changing looks of on-line library catalogues. These systemic documents grow out of the paper based catalogues, and in the early stages of the transition to an on-line format there were attempts at mimicking the "look and feel" of the traditional catalogue as closely as possible, e.g. in how individual records were presented. In time several factors have changed, bandwidth and storage-capacities have increased dramatically and patrons have acquired a taste for rich book-descriptions from commercial sites such as Amazon.com. As a result of the evolution of systemic documents into the electronic, on-line domain, and the changing expectations and demands from the environment, secondary documentation is now being enriched with tables of contents, descriptions and pictures of books, which is putting some strain on their MARC-based format.

Succession

Over time it is not unusual to see different evolutionary lineages occupy the same niche. If "dissemination of written texts between scholars" is seen as a niche, there has been (to make a long story short) a succession from handwritten letters to printed journals to electronic collections of articles made available through the Internet.

As new forms of documentation begin to make its way into niches that have been occupied by stable and recognized forms there can be quite a lot of commotion and doubt as to whether the new form will actually be able to perform the functions of the old in a manner that is as good as the old one. Such discussions are rife today, when we are witnessing the transition from articles in printed journals to electronic dissemination of articles in forms that resemble traditional journals or in completely new forms (e.g. pre-print servers). One outcome might be that electronic forms completely replace the old, printed forms, but it might also turn out that these forms differ in essential ways, resulting in a coexistence of the forms, where they perform different functions within the scholarly ecosystem that had previously been performed by just the old form.

One very interesting question in this area is what are the criteria of success? What makes one form of documentation replace another form, or take on just some of its functions? The answer will probably be found partly in the qualities of the form itself, and how well these are (or can be) adapted to the task at hand; but also partly in other circumstances, such as who has the strongest marketing budget or which form manages to attract key, trend setting or policy-determining actors.

Why?

Why go through the exercise of constructing a theory of an ecology of documentation? I think the ecological perspective can bring with it several benefits to the study of documentation:

- First of all there is the view of forms of documentation as evolutionary lineages, that have been shaped over time to fit into an environment with certain characteristics.
- Then there is the incentive to construct models of how documents interact with each other

and their environment, and not least what actually constitutes that environment. There is also the added benefit of a realization that such interactions are not static and given once and for all, but rather that changes will and do occur, and that a change in one environmental factor will have consequences for the other factors that are interacting to form any given ecosystem.

Based on this I think ecological thinking might be a fruitful perspective where changes in the factors that help determine the shape of documents is the object of study. However it should be noted that the present sketch of an ecological approach to the study of documentation is just that, a sketch in need of proper research that can explore the usefulness and applicability of its concepts.

Abbreviations and acronyms

DOI	Digital object identifier
MARC	Machine Readable Cataloguing
URI	Universal resource identifier
URL	Uniform resource locator
URN	Uniform resource name

References

- Aunger, Robert (ed) (2000). *Darwinizing culture : The status of memetics as a science*. Oxford: Oxford University Press.
- Austrheim, Gunhild (2003). *Theory and methodology in documentation science exemplified by key word indexing*. Tromsø: Universitetet i Tromsø.
- Berners-Lee, Tim (2000). Weaving the Web: the past, present and future of the World Wide Web by its inventor. London: Texere.
- Björk, Bo-Christer and Hedlund, Turid (2004). "A formalised model of the scientific publication process". *Online Information Review*, vol. 28, nr. 1, p. 8-21.
- Chapman, J.L. and Reiss, M.J. (1999). *Ecology: principles and applications*. Cambridge: Cambridge University Press.
- Dawkins, Richard (1989). The selfish gene. 2nd ed. Oxford: Oxford University Press.
- Enger, Magnus (2003). Dokumenter som bærere av memer: et fruktbart perspektiv?

 (Online) Tromsø: Universitetet i Tromsø. URL:

 http://thedocumentacademy.hum.uit.no/students/master/magnus.enger/for-alle/teomet.html (Accessed: 2004-05-23)
- Henderson, I. F. (1989). *Henderson's dictionary of biological terms*. 10th ed. Harlow: Longman.
- Hull, David L. (2001). *Science and selection : essays in biological evolution and the philosophy of science*. Cambridge: Cambridge University Press.
- Lund, Niels Windfeld (2001). "Omrids af en dokumentationsvidenskab: anno 2003". *Norsk tidsskrift for bibliotekforskning*, nr. 16, p. 92-127.
- Mabe, Michael A. and Amin, Mayur (2002). "Dr Jekyll and Dr Hyde: author-reader asymmetries in scholarly publishing". *Aslib Proceedings*, vol. 54, nr. 3, p. 149-157.
- Nardi, Bonnie A. and O'Day, Vicki L. (1999). *Information ecologies : using technology with heart*. Cambridge, Mass.: MIT Press.

- Odling-Smee, F. John; Laland, Kevin N. and Feldman, Marcus W. (1996). "Niche construction". *The American Naturalist*, vol. 147, nr. 4, p. 641-648.
- Powell, Andy (1998). "Resolving DOI Based URNs Using Squid: An Experimental System at UKOLN". (Online) *D-Lib Magazine*, vol. 4, nr. 6. URL: http://www.dlib.org/dlib/june98/06powell.html (Accessed: 2004-05-16)
- Winthrop-Young, Geoffrey and Wutz, Michael (1999). "Translator's introduction". In Kittler, Friedrich A. (1999). *Gramophone, film, typewriter*. Stanford, Calif.: Stanford University Press.