

Guarantee of existence for digital information and its significance for web-based knowledge work

Max Harnoncourt, Paul Meinl and Lotte Krisper-Ullyett

factline Webservices GmbH

Introduction

Web-based knowledge work is gaining ground. According to the authors, the actual potential of the Internet has been used rather poorly so far. The main reason is the uncertainty of whether and how long information will be available on the Internet. This problem is getting the more urgent the more intensive the Internet is being used for collaboration. The following article deals with this problem and suggests, as a solution, a concept for assigning guarantees of existence.

I. Overview

Developments in the area of information and communication technologies create a changing working environment in which so-called „knowledge work“ is playing a more and more central role.

A decisive characteristic of knowledge work is the usage of one's own as well as other people's knowledge resources in order to solve problems. New knowledge is developed by identifying, using and combining existing knowledge. At this point, the immense importance of exchanging knowledge with others should be highlighted. Knowledge work always implies collaboration („distributed knowledge work“).

These elements have always been part of knowledge-based work. However, they have gained significant importance due to the development of the Internet and the subsequent creation of new communication and information infrastructures. In the meantime, web-based collaboration via Internet has become indispensable. More and more often, central Internet communication platforms are used rather than communicating via e-mail and sending files back and forth. The stated reasons are increased efficiency due to central structuring and archiving of information as well as the usage of a common communication infrastructure which can be accessed by all participants.

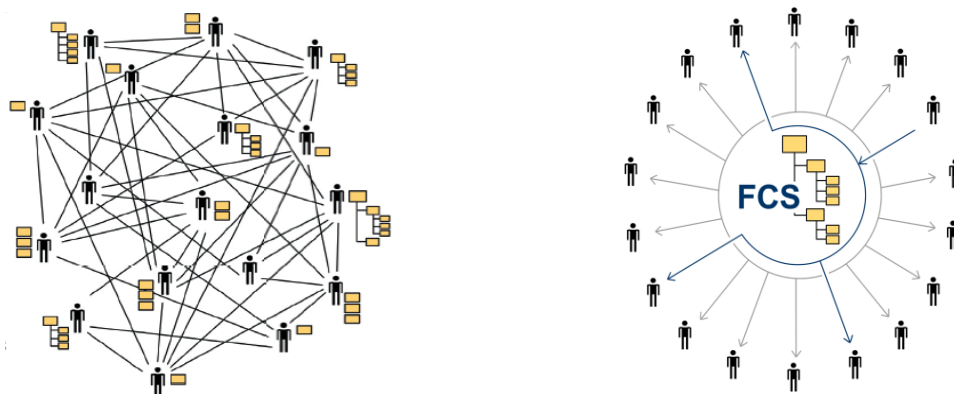


fig. 1: Increase of efficiency via usage of an internet platform (central structuring and archiving) as opposed to e-mail

A further but rarely used strength of internet platforms is the possibility to involve co-workers in the knowledge development process at an early stage. By doing so, knowledge workers do not only use the Internet to transport end results. Instead the expertise of others is included in the thinking process as early as possible. Due to the fact that distributed knowledge workers are usually members of several project groups, expertise of other knowledge areas can be included. The result of such distributed, web-based knowledge work is oftentimes dynamically linked hypertext of a collective of authors which in turn is linked with hypertexts of other collectives. The contextualization and integration of texts in these kinds of „living knowledge repositories¹“ creates added value which cannot be shown in a printable form.

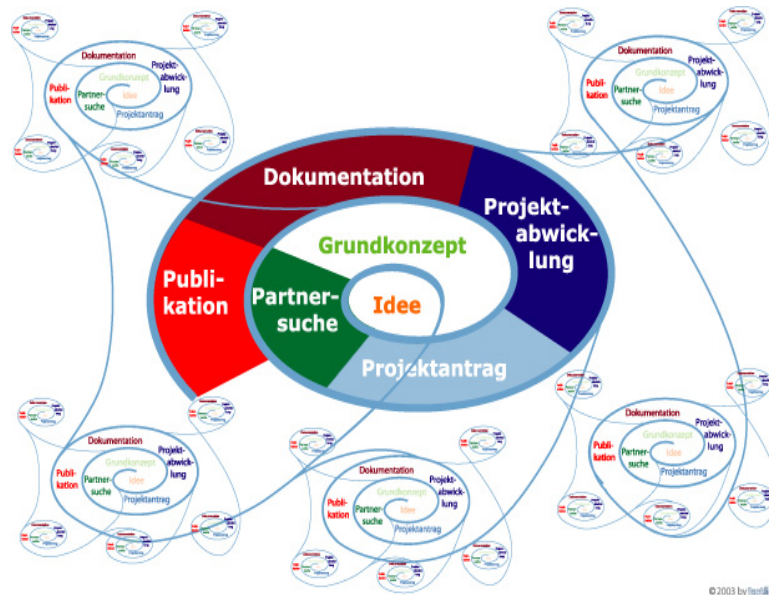


fig. 2: distributed, web-based knowledge work by hypertextually linking to different knowledge areas

However, a fundamental obstacle needs to be overcome: the uncertainty of whether and how long integrated text or rather text modules will be available in the Internet. At present, there is no accepted principle authors can rely on if they want to integrate web-generated text in their hypertext. A typical reference to web-generated text reads basically as follows: „The page looked like that when I visited it that day for the last time.“ At the same time, authors do not have a possibility to signalize other authors how to deal with their texts. As a result, the potential of Internet technologies is not used: texts are printed out, documents are stored on hard disks etc. This problem is the origin of the following contemplation.

First, a short excursion about internet platforms and EU-projects and their typical users will follow.

In chapter II, we will elaborate on the problem and deal with the aspects of referencing and reliable availability.

In chapter III, we will roughly describe a solution with the help of guarantees of existence. We will look at permissions and restrictions for assigning GOEs.

Chapter IV includes future prospects of how the concept will influence our current working methods.

¹ following the concept of communities of practice by Lave & Wenger (1991) and Wenger (2001)

Excursion: internet platform, example EU-projects

A good example for illustrating web-based knowledge work as described above is the organisation of EU-projects. The European Union supports cross-country and cross-organizational research projects in order to further the europe-wide exchange of know-how and the development of professional relationships in the European research community.

Without the Internet, the realization of such projects is hardly imaginable. The usage of internet platforms is already considered during the application process. Sometimes they are even used for composing the project application.

The advantage of internet platforms is that they combine three basic functionalities:

- central information pool („knowledge pool“)
- communication infrastructure („communication“)
- representative website for dissemination („website“)

This way, it is possible to document, edit and link all project related information.

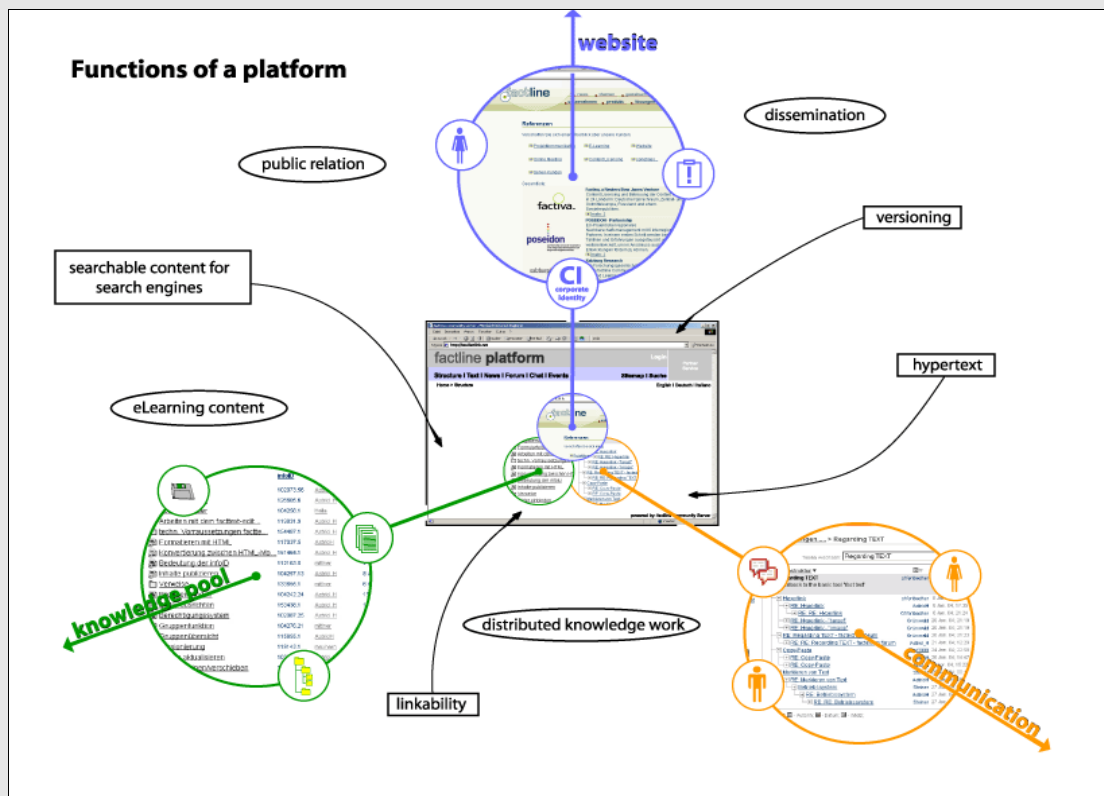


fig. 3: illustration of the three functionalities of an internet platform

II. Working with online information – dynamic hypertext

In order to increase the efficiency of collaboration via Internet, two aspects will be gaining considerable significance: the possibility to reference (link) dynamic information (hypertext) as well as its reliable availability.

II.1. Referencing (linking)

Due to collaboration via Internet, information is often available online and can be connected by referencing (linking) forming so-called hypertexts. This form corresponds to the nature of the Internet, a network of connected servers, as opposed to linear texts known from print media. Complex, dynamically changing content can be connected in a way that redundancies are avoided and associative structures, similar to human thinking, develop.

A decisive factor is the dealing with written information: On paper, written information (letters, articles) is still forwarded at a rather late stage. This is particularly true for research projects. A text in written form usually represents the end of a long thinking and wording process.

When it comes to using digital media, a change can be observed: While writing an e-mail, which corresponds best to a letter on paper, common requirements towards form, spelling, structure or argumentation become less important. This is even more true for communication forms such as SMS, chat and online forums.

On the one hand, this may signify a decrease in value of information in the eyes of the recipient. On the other hand, new opportunities for collaborative knowledge work arise due to the direct, process-oriented approach.

Thoughts, text modules and arguments are made accessible, and thus commentable, extensible and improvable, at a very early stage. This kind of online work, which takes place in written form, integrates information which is usually part of oral communication and allows insightful documentation of the work process at a later time.

At the same time, there is the opportunity of linking single pieces to a meaningful whole. For instance: A comment criticizes a text which includes relevant fragments of an online-discussion which in turn refers to an invitation to a panel discussion about a similar topic as well as to the profiles of the participants etc.

In order to be able to link content to a meaningful whole, the high dynamic of online content, in other words its constant changes, needs to be taken into consideration. When creating a link, authors need to be able to decide: Do I want to direct the reader to the most current version of the project plan or to the initial version of the plan as devised in the course of the project start? The possibility to create such dynamic links provides the precondition for efficiently referencing online content.

II.2. Reliable availability

The second precondition for efficient collaboration via Internet is the reliable availability of information. In order to integrate text into the context of external information, it is essential to make sure that references are maintained and work permanently. It is necessary to be able to trust on the availability of information.

At present, the Internet does not meet this requirement. „Dead links“ (links to false or no longer existing information) are the order of the day. At the moment, information on the Internet resembles a constantly disappearing and changing phantom which can not be caught.

The simplest solution to the problem of reliable availability would be to ensure that online published information remains eternally available under the same internet address. Intuitively,

one will realize that this is impossible. The main argument relates to economic questions: Who should pay for this?

It is imaginable that non-profit institutions such as libraries take over this task. There are already attempts in these areas. However, our ambition is to create the preconditions for efficient web-based collaboration in the course of a business model which is not dependent on subventions.

III. One solution: the guarantee of existence

One solution is to enable authors to make reliable statements about the availability of information during the creation of texts. The point is to not only indicate author, copyright owner, publishing date in the metadata of a piece of information but also how long the information will be available at the very same address. It is a matter of commitment on the part of the publisher who guarantees not to delete the information until a certain deadline has expired.

Author: Max Harnoncourt
Copyright: factline Webservice GmbH
Published by: Max Harnoncourt
InfoID: 222030.2 (...history) published on 24 Nov. 2005; 10:22
Guarantee of existence - GOE: 21 Dec 2005 extend

fig. 4: metadata for online published information with guarantee of existence (GOE)

Two things are important for the concept in order to work properly: Firstly, the assigned guarantee needs to be clearly visible. Secondly, it needs to be obvious which information the guarantee belongs to. Furthermore, the guarantee needs to be trusted. Trust in a guarantee of existence is dependant on the credibility of the organisation or institution behind as well as reasonable, transparent systematics for assigning guarantees of existence.

If credibility and systematics are accepted by authors and readers, the precondition of „reliable availability of information“ mentioned above is fulfilled. Combined with referencing, information can be reliably integrated in a developing environment in order to turn it into a long-lasting knowledge resource. The basis for efficient online collaboration is provided.

III.1. Systematics for assigning a guarantee of existence

When developing systematics for the assigning of guarantees of existence several basic questions need answering. First of all, a decision has to be made who is able to set guarantees of existence.

The following three possibilities seem reasonable:

1. The guarantee of existence is set by the author of the text during the online publishing process.
2. Each piece of information within a pre-defined area (e.g. within a communication platform) is automatically provided with a pre-defined guarantee of existence. Either a period of time (e.g. until six months after publication) or a specific date (e.g. 2005/07/01, end of project) can be defined.
3. A reader, interested in the information, defines a certain point in time until which the

information should be available. He either sets a new guarantee of existence or extends the one set by the author (or another reader). In order to allow the publisher to maintain its authority as an author to withdraw information, he or she has to be able to refuse the extension of a guarantee of existence.

III.2. Restrictions to the assignment of guarantees of existence

Due to several, mainly economic, reasons it can be assumed that the assignment of guarantees of existence needs to be subjected to restrictions. Let us look at the following restrictions:

- duration for which a guarantee can be set
- disk space

The process of deciding on limits is reasonably divided into two steps: First, the system operators need to specify how long and to which extent they can and should provide the technical infrastructure for publishing content. Only then can users be provided with guaranteed assets. Afterwards, users are able to set or extend guarantees of existence within the scope of their assets for their own data as well as for external documents, provided their authors have permitted to do so.

IV. Impact of the guarantee of existence

IV.1. Hypertext as a text format for knowledge workers

We anticipate that a major impact will show itself in the fact that knowledge workers will no longer be kept from using efficient hypertext format, which is usually superior to print format, due to doubtful availability. An important indication could be the future density of links and references in online created texts.

IV.2. Impacts on collaboration

We anticipate that the possibility to set guarantees of existence will influence online collaboration in several ways.

A major effect will concern project managers who will be required to develop a concept dealing with the usage of information created during the communication process even before the project actually starts. Precisely because guaranteed availability is connected with costs, one has to ask questions like „How long should some specific piece of information be available?“, „What should be communicated externally?“, „What different kinds of information are there (e.g. process, topic, legal information)?“.

Even project members will start to reflect on how their contributions should be treated. We have made the experience that these questions are usually discussed during or at the end, in both cases too late, of the project. Due to the possibility to set guarantees of existence, it is necessary to decide how information will be treated at the very beginning of the project.

In addition, project members or authors need to decide for each publication whether it should be possible to delete it or whether the information should be available until a certain point of time. The author indicates to the readers how he or she values the published information, whether he or she wants to dispose freely of the information or issue a guarantee for the content. It is essential to deliberate on the advantage of indicating the value of information via a guarantee of existence and the question whether authority over information should or can be given up. If others are allowed to extend the guarantee of existence authority over the information might be handed over completely. However, it clearly signals how deeply an author is committed to the published information.

Authors can certainly disregard the possibility of assigning guarantees of existence. However, we assume that their articles will be less appreciated as opposed to articles with guarantee of existence.

Then again, it is not possible to provide each piece of information with a guarantee of existence. In the course of project planning a budget is drawn up and a decision is made to which extent guarantees can be set. Both have a direct influence on the limitations of assets of project members.

Readers also benefit from information with guarantee of existence. The fact that a reader can rely on the availability of information, might cause him or her not to store the information. At present, the only possibility – and therefore common practice – to document important information which is available online is to store it on parallel media (e.g. one's own hard drive, server) or even to print it out. This is the only way to make information which should be included or referenced in an article reliably available. As a result, information is stored several times together with all arising problems such as disk space, data security, copyright etc.

IV.3. Contextualization and referencing as a quality indicator

As regards information requirements, quantitative measuring units such as number of pages, words etc. are usually used in the area of knowledge work. You will often hear that an article is supposed to have a certain amount of pages. Obviously, these measuring units poorly reflect the actual, informative value of an article.

In the area of scientific publications, where information is usually dealt with in a highly responsible way, the scientific quality of an article is gathered from the number of references, showing to which extent the author has considered existing concepts in his work. After having proved itself in the area of scientific publications, this indicator could find its way into knowledge work.

It is imaginable that the quality of a text in the Internet will be increasingly measured according to the extent of contextualization. In other words: In a hypertext environment written inputs can be regarded as being of high quality if they show few redundancies, with respect to the common information pool, and high contextualization (referencing).

In order to put concepts into context, an information pool including existing concepts in form of knowledge modules suited for hypertext is needed. The more easily existing concepts can be put into context the more efficient the completion of tasks by a knowledge worker. The more often a module is put into context the higher its value.

IV.4. Value of information

As mentioned above, the assignment or renewal of a guarantee of existence is an indirect statement about the relevance of a document. Documents with a long guarantee of existence seem more important than those with a shorter one. This might not apply to individual cases but on the whole it is absolutely imaginable.

One of the biggest problems within large information systems is to retrieve relevant information or, in other words, to separate the wheat from the chaff. Ranking mechanisms (or ranking algorithms) play an important role. They determine the order in which retrieved information is listed.

The incredible success of Google can be attributed to refined ranking mechanisms. Information is much more worth seeing if it is referenced by other websites which are in turn classified as worth seeing. Here, the realization that the linking of texts could be a relevant indicator for the importance of information was decisive.

In addition, the assured and indicated availability of information could be accepted as a criteria. Hence, it could be imaginable to consider the guaranteed availability of documents in ranking mechanisms. Whether this applies to search engines or to enclosed information systems is not important.

IV.5. Online publishing and public relations

If information with guarantee of existence is linked more often it will be, according to the ranking mechanisms described above, better ranked and thus more likely to be found. The described logic might cause organisations and projects, which want to achieve high public awareness, to develop an interest in providing information with long guarantees of existence.

In this context, it pays to look at the development of international and cross-organizational research projects such as EU-projects. The wish to publish research results for a broad audience causes the topic of „online publishing“ to gain importance. However, it is not sufficient to publish results online. Instead, it must also be possible to find the results. A long-term, beneficial strategy might be to provide online information with long guarantees of existence on a grand scale.

IV.6. More efficient search agents

Provided an accepted mechanism for guarantees of existence exists, search engine operators could use the guarantee of existence indicated in the metadata of a piece of information to efficiently search the web for new information resources.

At present, it is nearly impossible for search agents to determine how long a website will be available or when existing content will be changed. If a website is provided with a guarantee of existence, it will not be necessary to access it until the guarantee has expired. Subsequently, a reduction of page requests by search engines, which burden a system, could be achieved.

Please note: The guarantee of existence is not supposed to prevent a document (a concept) from being updated just because availability of a specific version has been guaranteed.

IV.7. Integration of text modules

Even in an environment where hypertext or dynamic hypertext is primarily used, creation of a linear text might be necessary from time to time. An example for a linear text in a hypertext environment might be a progress report of a research project.

A significant part of the information needed does not necessarily have to be created again. Instead, it can be composed of text modules available in the knowledge pool. However, a simple link to additional information is often not satisfactory. There is a desire for creating a linear text which includes all relevant information elements itself.

If, as mentioned above, the storage of redundant pieces of information in a common information system is refused, the copying of text modules needed will not constitute a satisfactory solution. Hypertext visionaries such as Ted Nelson often expressed a desire to include existing text in one's own text without having to copy it. The integration of external text into one's own text demands even more that the borrowed text is available as long as the new text. A guarantee of existence could create the necessary preconditions and make the integration of external texts common practice.

IV.8. Efficient deleting mechanisms

Permanently working information systems need to consider the whole content life cycle as regards the management of data. Oftentimes, the phase of deleting information is paid little

attention.

The point of the guarantee of existence is not to delete a piece of information as soon as the guaranteed date of existence has been reached. In order to guard against misunderstandings: The date of existence is not an expiration date. Even if the guarantee of existence of a piece of information expires, the information does not necessarily need to be deleted.

However, it can be assumed that intelligent deleting mechanisms will evolve in environments where the concept of the guarantee of existence is used. Unused information can be identified and deleted, contextualized information is maintained. If required, all information with expired guarantee of existence can be displayed. Information is ordered by degree of linking which reflects its relevance (see above). With regard to the ranking the following automatic classifications could be possible:

- (1) highly ranked information:
The guarantee of existence is extended if allowed by the authors.
- (2) moderately ranked information:
Information will be maintained without extending the guarantee of existence.
- (3) lowly ranked information:
Information will be deleted.

References

[1] Lave J., Wenger E. (1991): *Situated Learning: Legitimate Peripheral Participation*. Cambridge University Press.

[2] Wenger E. (2001): *Supporting Communities of Practice. A survey of community-oriented technologies*. Available: www.ewenger.com/tech [2005-02-05].