

Digital document and interpretation: re-thinking “text” and scholarship in electronic settings

Stefan Gradmann · Jan Christoph Meister

Published online: 17 January 2008
© Springer-Verlag 2008

Abstract The contribution starts from outlining the evolution of the scholarly production flow from the print based paradigm to the digital age and in this context it explores the opposition of digital versus analog representation modes. It then develops on the triple paradigm shift caused by genuine digital publishing and its specific consequences for the social sciences and humanities (SSH) which in turn results in re-constituting basic scholarly notions such as ‘text’ and ‘document’. The paper concludes with discussing the specific value that could be added in systematically using digital text resources as a basis for scholarly work and also states some of the necessary conditions for such a ‘digital turn’ to be successful in the SSH.

Zusammenfassung Der Beitrag beginnt mit einem Überblick zur Evolution des wissenschaftlichen Informationskontinuums auf dem Weg vom druckbasierten Paradigma in das digitale Zeitalter und geht in diesem Zusammenhang näher auf die Unterscheidung ‘digitaler’ und ‘analoger’ Repräsentationsmodi ein. Anschließend behandeln wir den als Folge des Übergangs zu genuin digitalen Publikationsformen erwartbaren dreifachen Paradigmenwechsel und dessen spezifische Konsequenzen für die Geistes- und Sozialwissenschaften sowie als deren Folge wiederum die Re-Konstitution elementarer Kernbegriffe geisteswissenschaftlichen Arbeitens wie ‘Text’ und ‘Dokument’. Der Beitrag schließt mit einer Betrachtung des spezifischen Mehrwerts, der sich aus dem systematischen Rekurs auf digitale Textressourcen in den Geisteswissenschaften ergeben könnte und geht dabei auch auf die erforderlichen Vorbedingungen eines solcherart erfolgreichen ‘digital turn’ in den Geistes- und Sozialwissenschaften ein.

S. Gradmann (✉)

Regionales Rechenzentrum der Universität Hamburg, Schlüterstr. 70, 20146 Hamburg, Germany
e-mail: stefan.gradmann@rrz.uni-hamburg.de

J. C. Meister

Universität Hamburg, Institut für Germanistik II, Von-Melle-Park 6, 20146 Hamburg, Germany
e-mail: jan-c-meister@uni-hamburg.de

Résumé Nous partons d'un bref aperçu de l'évolution de la chaîne d'information scientifique à partir du paradigme de l'impression vers l'âge numérique, évolution qui peut être conçue comme un passage d'un mode analogue vers un mode binaire de représentation. Ensuite, nous traitons du triple changement de paradigmes entraîné par des futures modalités de publication numérique et ses conséquences spécifiques pour les Sciences Sociales et Humanités (SSH). Ces conséquences à leur tour entraînent la nécessité de re-constituer des notions de base des SSH telles que 'texte' et 'document'. Pour conclure notre contribution nous considérons la valeur ajoutée par une pratique philologique systématiquement basée sur des ressources textuelles numériques avant de considérer les conditions requises pour le succès d'un tel tournant numérique dans les SSH.

1 Introduction

Even the most ardent defendants of traditional print media have by now come to admit the pragmatic advantages offered by digital documents: ease of access and distribution, ease of revision, ease of information retrieval, to name just a few. However, we believe that understanding the full impact and potential of electronic publishing in the Social Sciences and Humanities (SSH) requires reflection upon broader methodological issues. Several vectors or primary oppositions constitute this complex context:

- the scholarly information continuum as a whole and its evolution from print-based to electronic working paradigms,
- the underlying shift from analog to digital representation modes,
- the revolutionary changes that can be foreseen as a consequence of the combined first two vectors,
- the specific difference of the SSH as opposed to the Science-Technology-Medicine (STM) culture of organizing information and the specific impact of the digital revolution resulting from this specific difference.

These context vectors constitute a multi-dimensional continuum which we ought to explore in order to measure the innovative potential of genuine digital approaches within the SSH and to identify the conditions required for realizing this potential.

Our contribution is an attempt to outline this context and to give some indications as to the potential re-thinking of basic scholarly notions such as 'document' or 'text' in future digital settings.

2 Evolution of the scholarly information continuum from print to XML

As W. McCarty (2003)¹ has put it, "Academic publishing is one part of a system of highly interdependent components. Change one component [...] and system-wide effects follow. Hence if we want to be practical we have to consider how to deal

¹ http://lists.village.virginia.edu/lists_archive/Humanist/v17/0336.html.

with the whole system.” Thus, in order to understand the coming paradigm shifts it is useful to first consider the evolution of the print based scholarly information continuum which has been stable and basically unchanged for centuries. This continuum can be conceived as a circular work flow centered around basically monolithic printed information objects and is sketched in Fig. 1 below. In this traditional view of the scholarly information chain typical stages such as ‘authoring’, ‘reviewing’, ‘publishing’, ‘managing’, ‘apprehension’, ‘quotation’ and ‘annotation’ of scholarly information objects were implemented using very few and very stable cultural techniques (basically reading and writing). Furthermore, these stages were organized in linear, circular workflows with no or at most marginal modifications in sequence and centered on well understood, monolithic entities (documents). With the advent of digital media and working instruments this functional sequence remained practically unchanged in a first phase, during which the individual steps were simply electrified using digital means to emulate what had been done using traditional cultural techniques before as indicated below in Fig. 2.

This scholarly value chain in emulation mode is somewhat similar to incunabulae in early print age: just as the latter have been preserving major characteristics of medieval folios the former kept (and partially still keeps) typical characteristics of the traditional value chain. Not only is the circular sequence preserved, but also its individual stages remain functionally unchanged and the use of well known cultural techniques remains constitutive. The same is true for the information object at the center of the circle which uses print-analog formats such as PDF to emulate basic characteristics of the ‘bookish’ information support.

The first real qualitative change within this functional continuum happens with its transition to a third phase which is illustrated in Fig. 3 below including some of the questions related to this process. In this third phase individual stages in the still basically unchanged linear function paradigm are now remodeled digitally and thus

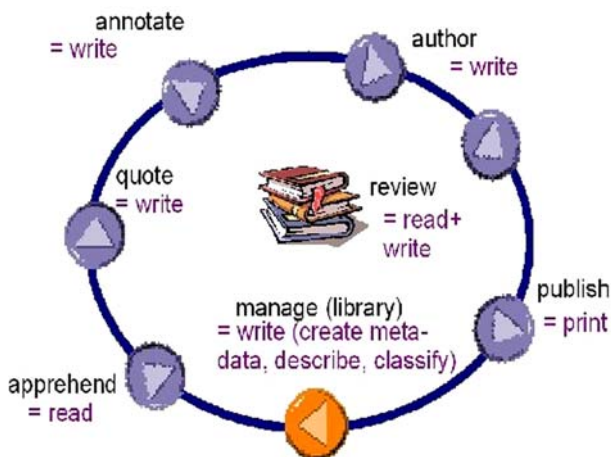


Fig. 1 The traditional scholarly information continuum

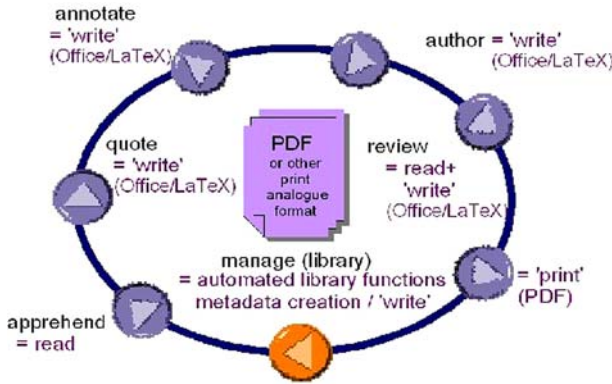


Fig. 2 The traditional continuum in emulation mode

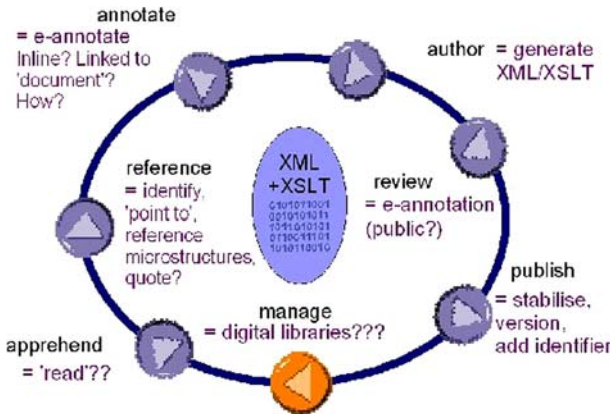


Fig. 3 Scholarly information continuum ... going digital!

undergo substantial changes. Transition to this phase is currently under way and more or less advanced depending on the different scientific cultures.

Authoring of scholarly documents, for instance, turns into generating of content using some XML syntax and appropriate presentation modes using XSLT or similar processing techniques. The reviewing stage turns into a more or less public and open procedure of digital annotation. ‘Publishing’ in this context may be equivalent of stabilizing document content, applying version information and a unique identifier. ‘Quotation’ instead of replicating parts of external documents more and more turns into identifying external information objects and referencing to its internal micro-structure. It remains unclear, to which extent the term ‘reading’ can still be applied to the related acts of apprehension. And it becomes more and more evident that the ‘library’ metaphor is increasingly inappropriate for the fundamentally changed management methods for digital information objects.

3 Digital versus analog

Clearly, going digital has changed our practices and methods—but has it also changed our underlying methodology and our conceptualization of the objects that we are dealing in?

The slogan of the so-called ‘digital revolution’ is hard to avoid in this context, and juxtaposing ‘the digital’ and ‘the analog’ in a somewhat metaphorical sense seems compelling. Early reflections on the technological paradigm shift—viz. Dieter Balkhausen’s “*Die dritte industrielle Revolution. Wie die Mikroelektronik unser Leben verändert*” of 1978—have perhaps unintentionally contributed to the mystification of ‘the digital’ by affording it ‘revolutionary’ status. As such ‘the digital’ is metonymically elevated to the status of one of the driving forces behind the change from a tangible goods oriented industrial society to a post-industrial society that deals in intangibles such as knowledge, information and services. While these intangibles would seem to be more akin to the abstract objects that traditional humanities focus on, concepts such as ‘information’ and ‘knowledge’ are at the same time reductive—even in combination they cover only a small part of what makes up the phenomenology of the mental. But this is not the only conceptual incompatibility between traditional Humanities and the propagated ‘information society’. Even if we accept knowledge and information to constitute the outcome of cognitive and mental processes, this very perspective onto ‘outcome’ as a finite ‘product’ is what makes them problematic to process oriented thought. Humanities and Social Sciences conceptualize their objects as historical and dynamic—always in transition, and always contingent on historical contexts which are in flux. Seen from this perspective, mere information is trivial, because it lacks context, and the fact that digital media make even more information available will only increase the problem. Digital texts, if we merely conceive of them as delimited containers that carry a certain amount of information, will not help us to solve this problem either. Could it be that something is ‘wrong’ with the mode in which that information has been sampled? Is the digital *modus operandi* perhaps per se incompatible with the Humanities’ endeavor?

It might help to clarify what it means for a bit of information to be ‘digital’. In terms of signal theory, a digital signal is one that is made up of a series of discrete measurements that indicate the value in some parameter at different points in time. A digital signal can be represented in tabular format, or as a matrix. Analog signals by contrast are non-discrete—we tend to visualize them as amplitudes which may, perhaps, even be expressed in terms of a mathematical formula, but which in reality (i.e., as sensual phenomena) cannot be broken down into a string of individual bits and pieces with empty spaces in between.

From a humanists point of view, this idea of the digital is indeed hard to accept; it is almost anathema to what Humanists and Social Scientists study—the historical continuum of emotional, mental and behavioral responses of human beings who find themselves embedded in a world that is not just constituted by physical objects and empirical events, but to a large degree by just that—mental and behavioral responses of (other) human beings. But what exactly is so problematic about ‘the

digital', and what exactly makes it incompatible with the human experience of the world?

The core issue seems to be that of discreteness. Digital information processing and digital representation are based on the idea of the world as something that is experienced in terms of (if not even made up of) discrete, and hence measurable states. In order to be discrete, a phenomenon has to be clearly delineated and individuated. The pragmatic advantage of taking this approach is obvious: it makes phenomena measurable, thus rendering them suitable for a type of exchange where nothing is lost, or added in the course of the process. However, the metaphysical consequences of this mode of conceptualizing the world have troubled philosophers from the very beginning. Zeno's well-known paradox of Achilles and the turtle which he can never overtake comes to mind. The little we know of Zeno (450 BC) as a person is owed to the first few pages of Plato's *Parmenidis*. More important and widely discussed ever since Aristotle are the over 40 paradoxes which Zeno made up in order to defend his teacher Parmenides who had attempted to dispute the thesis of so-called 'ontological pluralism': that is, the idea that the world is made up of discrete entities. With his paradoxes of plurality and movement (of which Achilles and the turtle is the most famous) Zeno tried to demonstrate that this premise leads to logical contradictions. Accordingly, the gist of Achilles' never ending race with the turtle was to prove that a description of the world in terms of discrete states—that is, as a series of measurements taken at individual positions along an indefinitely shrinking time line—will not be able to grasp what is evident to everyone: the fact that Achilles overtakes the turtle. Zeno took this to prove that the world is indeed just one entity, and not many individual ones.

Clearly, phenomenology and metaphysics do not go hand in hand in Zeno's paradox—and neither do physiology and epistemology in the paradoxical situation which the human mind finds itself in. There is no paradox here either, for our own sensory apparatus performs just like the iPod: it registers discrete signals. This holds true for our sense of sight, our sense of hearing, our sense of touch: they all have a certain threshold below which they cannot distinguish discrete impulses *as* discrete, but rather begin to merge the individual signals into one. The threshold level is different in every sense, our sense of hearing being the one with the highest capacity for resolution since we can distinguish variances in pitch of 0.3% only (i.e., a 1,000 Hz signal from a 1,003 Hz signal) and down to a 30 ms difference in extension over time. But what turns all of this into music is—our brain. So where is the 'digital revolution' in a CD, other than in the brute sense of the technological apparatus? And even there the dividing line between analog and digital media gets blurred on closer inspection. For example, was there ever a truly *analog* photography? Photographic film is made up of crystals which, in terms of their density, account for the film's physical properties, such as granularity, sensitivity to light, etc. Physiologically speaking, the fact that our eyes did not register this merely had to do with the size of the crystals. Epistemologically speaking, registering individual crystals simply does not make sense—we want the picture, not the pixel. In this perspective the technological dimension of the digital is rather trivial; it is by

no means as new, foreign or revolutionary to us as its proponents would like to make us believe.

What is the conclusion to be drawn from this? For the humanities the potential benefit of the digital paradigm cannot reside in the technological ability to measure the finer grain and transmit that bit of information without distortion. As soon as our brain gets involved, we always deal—and will continue to do so—not in ‘the real thing’, but in our own arte facts: sensory information integrated into *Gestalt* like phenomena in as much as abstract ideas integrated into discourses embodied as ‘texts’ and ‘documents’.

4 The triple paradigm shift

Even if the formative power of traditional cultural techniques rapidly decreases within the individual stages as part of the transition from analog to digital representation modes as indicated above, other basic characteristics of the traditional continuum remain unchanged in this stage: the scholarly value chain remains linear-circular and is focused around a well understood monolithic information object, the ‘document’.

However, these two remaining characteristics in turn may be subject to de-construction in a next phase that is already casting its shadows and which is likely to influence the continuum as indicated below in Fig. 4.

Two tendencies can already be outlined regarding this future phase: the stages that used to be organized in a sequential-circular will increasingly relate to each other in almost any networked order and the central information object, the ‘document’ loses its monolithic character, itself becomes a networked cluster of information entities with increasingly dynamic and diffuse borders.

We will thus be facing a triple paradigm shift but which has specific consequences with respect to the different scholarly/scientific cultures.

If one accepts—at least as a working hypothesis—the distinction established by C. P. Snow in his Rede lecture on “The Two Cultures” and considers the respective

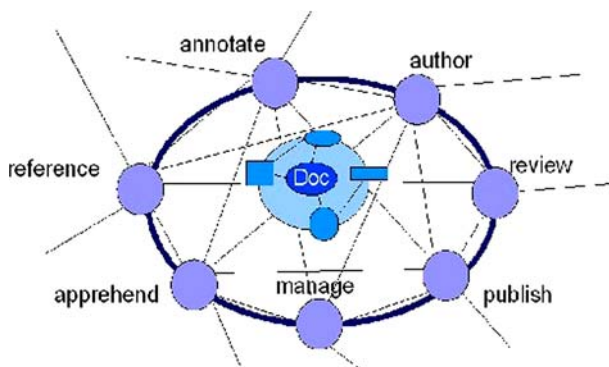


Fig. 4 A de-constructionist scholarly information continuum

consequences of the triple paradigm shift for the sciences (henceforth STM) and the humanities (SSH) striking differences are almost evident.

In such a perspective, the *erosion of the linear/circular function paradigm* only marginally affects the way ‘publication’ is conceived in the SSH because of the prevalent ‘monolithic’ publication practice in this culture:

- journal articles and related peer reviewing procedures are still rather marginal,
- authors still tend to work in ‘splendid isolation’ in the SSH with collaborative authoring still being an exception (such as the present contribution!).

The *declining formative power of traditional cultural techniques* certainly affects the SSH (and probably much more than the sciences), but this does not specifically affect the publishing function.

However, the *de-construction of the ‘document’ notion in digital, networked settings* vitally affects the SSH in a very specific way. This process fundamentally changes the conditions of production and publication as well as the conditions of apprehension and reuse of scholarly documents. The consequences touch the very core of scholarly work which in both of its main strands of work is fundamentally concerned with documents both as objects and as instruments of scholarly activity. As shown in Fig. 5 below, both the ‘aggregation’ (arrows pointing down) and the ‘modeling’ strands have their point of origin in digital corpora (and thus most of the time in document clusters) and produce new documents in turn!

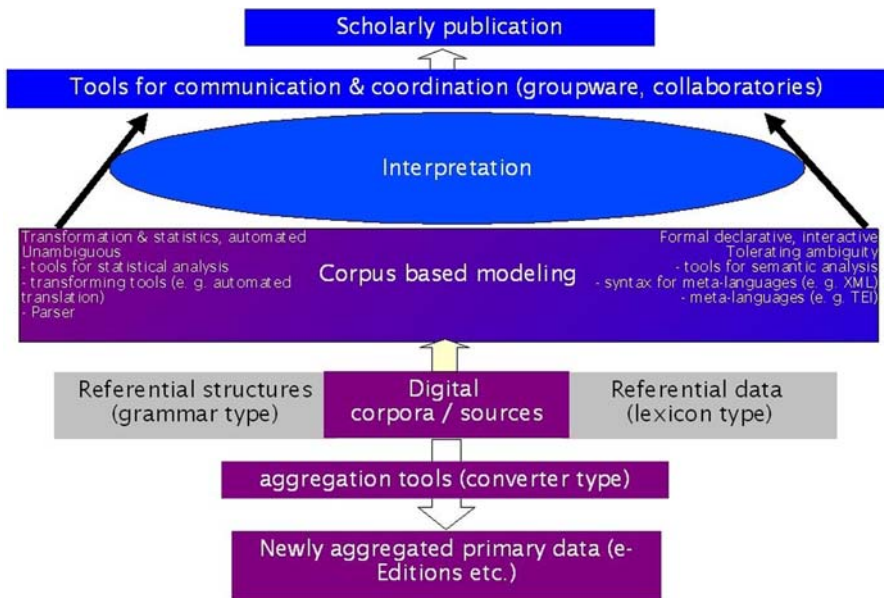


Fig. 5 Digital corpus-based modeling and aggregation in the SSH

And this observation organically leads to a closer investigation of the specific relation between the SSH (especially the hermeneutic disciplines) and the constituent representation modes of documents as complex signs.

5 The Pandora's box of semiotics ...

When considering this issue in more detail it becomes clear that signification and document modeling in all discussion related to electronic publishing up to now have basically been coined on the information model prevailing in the empirical sciences. In this model, scientific research as the core activity is completely dissociated from the publication process. Only once 'research' has yielded 'results' these in turn are 'packaged' in discourse and published (typically as a journal article): in this extremely robust and not very complex 'container' model of scientific publishing it is perfectly sufficient to remain on 'emulation level' as outline above, since the publishing stage is not at the core of scientific work, anyway.

However, scholarly publishing in the SSH takes place in a substantially different information model: scholarly research and discursive 'packaging' cannot be separated in this perspective and the published results of the core scholarly activity are documents. This accordingly results in complex document models and publishing formats heavily intertwined with core research operations. In such a view, the 'container' models used in 'hard sciences' publishing are over-reductionistic and inappropriate, and complex relations between signifiers and signified subjects are constitutive.

Clearly, behind the different information models underlying the respective publication cultures of the STM and the SSH lurks another, even more fundamental semiological difference. In fact, dominant discourse in electronic STM publishing communities (mostly emanating from computer science) uses terms such as 'document', 'sign' or 'name' quite naively and without referring to their inherent semiological complexity. This results in a (technically) high-level nominalist regression: the 'Pointer → Object'-Model, in which 'words' point to 'real' things as in Fig. 6 below.

The perfect incarnation of such thinking represent the 'ontologies' of the semantic web!²

Fig. 6 Words pointing to 'things'



² The paper of Benel et al. (2001) gives a very valuable discussion of the profound inappropriateness of positivistic ontology based approaches in the SSH.

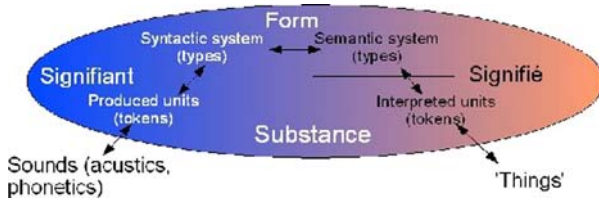


Fig. 7 A simplified model of the semiological space

As opposed to this very simple mode of conceiving the relation between words and things it is useful to consider the linguistic model of significance that has developed in the twentieth century starting from De Saussure's theory of the sign and considerably refined by Hjelmslev, Eco and others³ as indicated in the (much simplified) Fig. 7.

Signifiers and signified subjects cannot be dissociated in this vision as it is impossible to consider form and substance of constituents independently: produced and interpreted individual units always have to be seen as part of their respective systemic context. And both sounds and real 'things' are *not* part of the representational space in such a view.

Such thinking has once been declared by a senior computer scientist as "opening the Pandora's box of semiotics"—but the fact is that exactly such thinking is required to understand the way the SSH relate to documents, which in turn must be conceived as complex significant units and themselves are part of a system made up of such units (vulgo 'litarature').

It then becomes clear that (electronic) text is not just a transcription of speech acts (parole) and at the same time it must be noted that the notion of 'text' basically remains a blank spot in linguistics and still is subject to fundamental research as a complex, semiological digital object. In such an approach the model used above might tentatively translate to electronic documents as in Fig. 8 below.

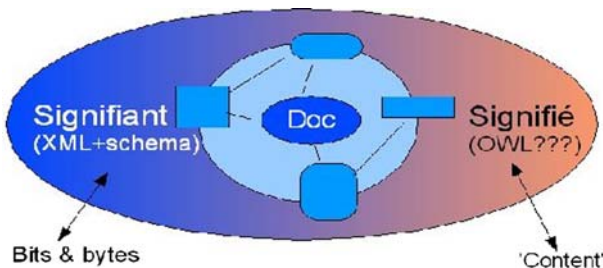


Fig. 8 A tentative representational model for electronic documents

³ Probably the best introduction to this semiological approach still gives Eco (1968, 1976).

6 ... and a way to re-think the ‘document’ notion

The heart of the issue thus seems to better understand the metamorphosis of the ‘document’ notion in the digital context—and a very competent attempt in this sense has recently been made by the French research group RTP-DOC (CNRS) that has used the pseudonym Roger T. Pédaque to publish fundamental work relating to the de-construction of the ‘document’ notion currently under way in the digital, networked context.⁴

RTP-DOC presents the evolution of the ‘document’ notion in the passage from printed to digital documents along three paradigms:

- Form (vu = ‘Look at’, morphosyntax), as material or non-material structured object, the corresponding chapter is *forme, signe et médium, les re-formulations du numérique*;
- Sign (lu = ‘read’, semantics), as meaningful instance and thus both intentional and part of a sign system, the corresponding chapter is *Le texte en jeu: permanence et transformations du document*;
- Medium (su = ‘Knowledge, Interpretation, Apprehension’, Pragmatics) as a vector of communication, part of a social reality with constituting temporal and spatial processes of mediation, the corresponding chapter is *Document et modernités*.

In each of the three conceptional paradigms one of the aspects is used as a dominant, yet non-exclusive vector for developing equations that distinguish traditional, electronic and future web-based document notions with each of these equation triples resulting in a definition of the respective nature of the ‘electronic document’.

Thus, the ‘form’ vector, in which object nature is constitutive, can be summed up in these three equations:

1. Traditional document = medium + inscription
2. Electronic document = structures + data
3. XML-document = structured data + style sheet

And these in turn result in a first definition: “An electronic document is a data set organized in a stable structure associated with formatting rules to allow it to be read both by its designer and its readers”.

Likewise: the ‘sign’ focused on the meaningful nature of documents yields the following three equations

1. Traditional document = inscription + meaning
2. Electronic document = informed text + knowledge
3. Semantic Web document = informed text + ontologies

And the resulting definition reads: “An electronic document is a text whose elements can potentially be analyzed by a knowledge system in view of its exploitation by a competent reader”.

⁴ Two publications are of interest here: Pédaque (2006, 2007); see also the web presence of its group at <http://rtp-doc.enssib.fr>.

Finally, the ‘medium’ vector organized around the ‘document’ as social phenomenon has these three equations:

1. Traditional document = inscription + legitimacy
2. Electronic document = text + procedure
3. Web-Documents = publication + measured usage/access

With the following definition associated: “An electronic document is a trace of social relations reconstructed by computer systems.”

Without referring in more detail to the rich discussions within RTP-DOC it should be evident that the conceptual framework proposed by this group could serve as an excellent basis for re-building consensus regarding the ‘document’ notion and for a better understanding of the nature of digital, networked document resources. Such an understanding in turn is required in order to better understand the specific impact of digital publication techniques in the SSH, as the ‘document’ notion is at the semiological heart of hermeneutics oriented scholarly work.

7 The added value of digital text

Hermeneutic disciplines study the formation, attribution, extraction, exploration or generation of ‘meaning’, procedures which complement one another in a process commonly referred to as ‘interpretation’. What exactly could the relevance of our new notion of ‘document’ be in this regard? Is there a potential benefit over and above the pragmatic and social dimension?

What sets a mere character string apart from a text is the semantic surplus value of the latter, the fact that base-level signification is aggregated into complex constructs such as de-notational ‘meaning’ and in most cases thereafter interpreted beyond what has been encoded at the surface level of representation. This interpretation takes place where the text-transcendent dimension of ‘sense’, in which document and culture interface dynamically, comes into play. Various questions would need to be explored here, including the following:

- Do digital texts intrinsically *carry* an additional and specific semantic surplus value over and above what traditional print media can present us with?
- Or do digital texts rather *enable* us to construct and then exploit such new surplus value?
- Do they perhaps even put me into the position to generate not only new, but richer constructs of meaning and sense?

One possibility to address these questions is to analyze the functional add-ons which digital texts offer in contrast to traditional print texts, and to position them within the two-dimensional continuum of complexity and level of interpretation involved. Figure 9 presents an attempt to this effect. As one can see, the bottom left quadrant of ‘low complexity—low level of interpretation’ consists of a number of operations all of which can more or less be performed on a digital text or texts in a context free approach. Most of these are by now fairly common search and retrieval operations. The level of semantic surplus value to be derived here will hardly reach the threshold

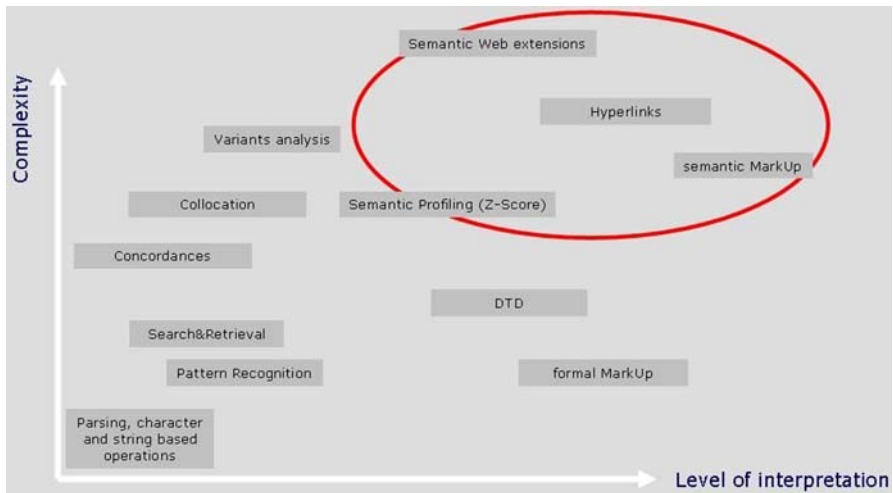


Fig. 9 Current digital text-based operations

that sets de-notational meaning constructs apart from contextually bound sense constructs. These are the operations found in the upper right quadrant indicated by the red circle. Unfortunately, this is still un-chartered territory for the vast majority of digital document users. But not only end users tend to shy away from anything that smacks of high-level Mark-Up; systems and standards developers too tend to find these practices too time consuming, or too idiosyncratic. The bold TEI initiative has certainly gone more than half the distance in this regard—however, when paging through the 23 chapter strong guidelines it is hard not to associate an eighteenth century encyclopedia striving to systematize and capture it all. The problem is that with texts, as with the world, ‘it’ is changing all the time, and relative to the questions we want to ask, neither of which can be predicted.

What seems to be needed, then, is an approach that empowers the user to explore digital documents with respect to the complex interplay of empirical regularity in the base material (from strings upward to higher level formal segmentations), normatively assigned qualifiers (from low level tags to high level semantic markup), and dynamic re-configurations of the digital document which are triggered by user interaction, as well as by inter textual processes which connect the document at hand within the ever expanding universe of digital documents at large. Figure 10 sketches out such a dynamic, multi-dimensional notion of the digital document and its expansion into a functional aggregate of procedures that would turn the traditional ‘text’ into what one might call a ‘heuristic machine’. In essence, an advanced notion of ‘digital text’ could in fact be defined from the perspective of such a virtual machine: in order to qualify as a fully realized ‘digital text’ a given document would have to prove functional within it. This is the stage where ‘going digital’ re-arrives at the cognitive *modus operandi* particular to humans and their societies: synthesis.

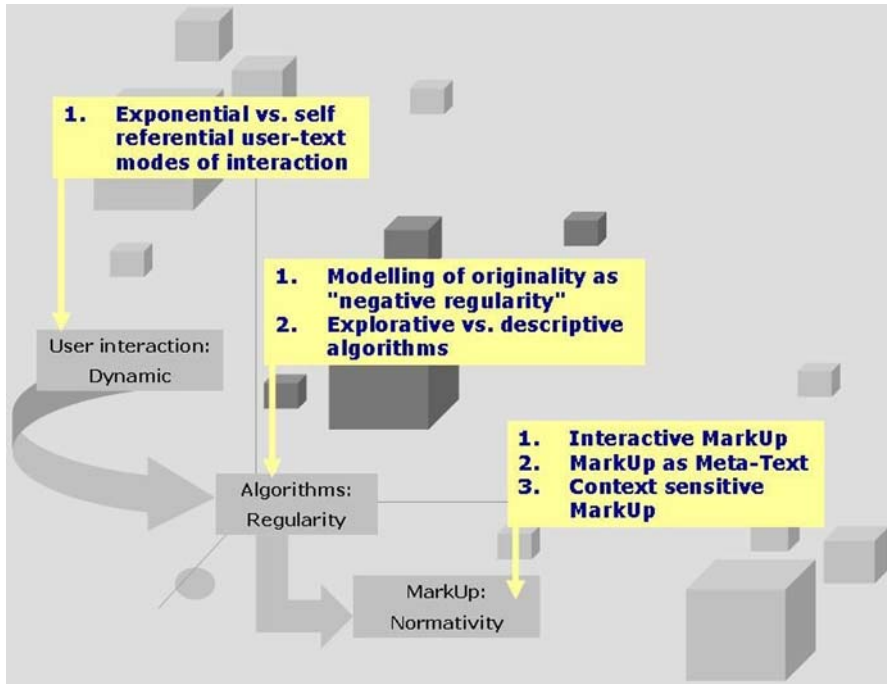


Fig. 10 A three-dimensional approach toward the exploration of the universe of digital documents

8 Barriers to overcome on the way to a digital turn

However, the scholarly reality of the SSH is still quite far away from such a digital turn in general and from realizing the true potential of electronic publishing in particular.

One of the reasons for this state of things is the fact that as long as electronic publishing simply digitally emulates traditional analog publication modes it remains of little specific scholarly interest; it requires a very complex technical machinery for modeling the complex scholarly publication formats without yielding sufficient added value.

On the other hand it should be evident from this contribution that any serious attempt to integrate digital publication formats (both as instruments and objects!) in scholarly discourse and its processing modes would turn out to be a very ambitious and complex undertaking.

The consciousness of the major challenges associated to such a step might explain most of the more or less conscious reluctance widespread in the SSH communities to truly adopt novel digital publication techniques. In the end, SSH scholars may simply be afraid of the “system-wide effects” McCarty referred to in his statement quoted at the beginning of this contribution.

In order to overcome these mental and intellectual barriers, a number of elements are clearly required, and some of these have been touched at in this paper.

We have tried to make clear that a newly established consensus regarding the ‘document’ notion and its constitutive aspects in a networked, digital context is required by the SSH as precondition for operational and persistent models for digital documents.

Furthermore, we need appropriate methods of semantic processing of digital document content clearly beyond the high-tech nominalistic regression of semantic web-ontologies.

We did not discuss the need for a scholarly pragmatic agenda with respect to digital publishing—but it should be evident from this section that both, the culture of appropriately using truly digital resources and a clear vision of the associated added value are not yet as entrenched in the SSH as would be required for leaving the emulation mode.

In a vision that ultimately renders obsolete Snow’s simplistic dichotomy of the ‘two cultures’ one could conclude that for digital publishing to truly work both in the STM and SSH communities we need a broader vision of ‘E-Science’ and ‘E-Scholarship’ alike which then *includes* digital publishing as one of its constituents.

The present contribution should have made clear some of the specific conditions within the SSH for integration in such a broader picture.

References

- Benel A et al (2001) Truth in the digital library: from ontological to hermeneutical systems. In: Proceedings of the fifth European conference on research and advanced technology for digital libraries (Lecture notes in computer science), vol 2163. Heidelberg, pp 366–377
- Eco U (1968) *La struttura assente*. Milan
- Eco U (1976) *A theory of semiotics*. Bloomington
- Pédaque RT (2006) *Le document à la lumière du numérique*. Toulouse
- Pédaque RT (2007) *La redocumentarisation du monde*. Toulouse