

Form and Content in the Electronic Age

Frédérique-Anne P. Harmsze¹ and Joost G. Kircz^{1,2}

¹van der Waals-Zeeman Instituut, Universiteit van Amsterdam
Valckenierstraat 65, 1018 XE Amsterdam, The Netherlands

²Elsevier Science, PO Box 103, 1000 AC Amsterdam, The Netherlands

www.phys.uva.nl/fnsis/onderzoek/comm/home.htm

Abstract- The emerging culture of fully electronic publishing will considerably change the scientific communication process. At the moment, the emphasis is put mainly on the logistics and the storage of electronic documents as such. In this contribution we address the issue of changes in the representation of scientific communications due to the intrinsic possibilities of the electronic medium. With the societal needs of authors and readers as a starting point we present first results of a new, modular, model of scientific articles.

I. INTRODUCTION

Too often the storm troops of technological innovation take us by surprise. However, technology is there to help us and not to hinder us. So, in order to set out a programme for electronic publications and publication management by publishers and libraries, it is important to properly establish the scientists' needs with respect to the form and content of the article, as well as the societal, added-value, functions that publishers and libraries fulfill for scientists (authors and readers). Based on that knowledge the intrinsic capabilities of the medium have to be explored in order to better accommodate the dynamics of the scientific process. The societal functions are based on what authors and readers want using a text, and how these demands change over time, depending on the different roles of the readers in the various stages in the research project. From these needs, and the fact that in an electronic environment textual as well as non-textual information is freely manipulatable, we propose and discuss a new, modular, model for scientific documents. This model fulfills the standard demands with respect to a classical document and it makes use of the intrinsic capabilities electronic publishing is going to provide us. Based on the different cognitive roles various kinds of information play, we propose to cast the scientific communication in modular form; the document ceases to be a linear text, written as if it is to be read from top to bottom. In a modular form, the reader reads those modules of information he/she is most interested in, just as a trained reader, browsing through a journal, starts reading those parts of the article that hit the eye haphazardly, or that are typographically indicated as containing

a particular type of information (e.g. experimental data, conclusions). We suggest that such a model will help to define test environments for electronic publishing experiments, in which the novel technological capabilities will be matched with the technological independent roles of the scientific article.

In this paper we discuss briefly the role of scientific documents and the author's and reader's needs with respect to the scientific article. We then present our model for a modular build-up of scientific articles. Subsequently we present a first evaluation and discuss the findings.

II. ROLE AND USAGE OF SCIENTIFIC ARTICLES

The role of scientific articles is crucial in the process of scientific information and knowledge interchange [1] and plays an important societal role in the scientific community [2]. It is the object around which the whole fabric of writing, publishing and reading is centred. Scientific articles are the prime representation forms for scientific information; they are closed entities, easily portable and well structured as a result of a century long tradition of scientific publishing [3]. In order to determine the form and content of the scientific article that allows for improved, electronic scientific communication, we have to analyse the requirements of the readers and authors independent of technology, but with a keen eye on what technology has in store.

Considering scientific communication via articles as a special case of goal-oriented, rational communication [4], we can derive clear requirements of which the managerial consequences are now vested in publishing houses and libraries. The objectives of the author include the advancement of science, recognition and feedback. The reader's objectives are the desire to understand and use the work. This results in requirements with respect to the clarity, efficiency and consistency of the message.

Taking a more precise look at the authors' and readers' requirements we can make a check list that, must be kept in mind in any serious electronic publication model, although there are variations per scientific discipline. Quoting from a recent overview, Kircz and Roosendaal [5] come to the following points.

On the acquisition (readers') side, we list: a) *Reliability*, which finds its expression in the authentication, validation and certification process. b) *Relevance*, related to subject, scope and

This work is part of the 'Communication in Physics' project of the Foundation Physica, and financially supported by the Foundation Physica, the Shell Research and Technology Centre Amsterdam, the Royal Dutch Academy of Sciences, the Royal Library, and Elsevier Science NL.

level of research. Relevance can only be judged by the individual researcher. c) *Timeliness*, the desired time to access information. This depends very much on the dynamics of the field. Note that it holds not only for current new research but also for old relevant work that rests in an archive. d) *Presentation*, allowing for efficiency of communication and convenience. e) *Storage*, an important issue closely related to the capacity of easy retrieval (local subscription, inter-library loan, web-version, etc.). Within this point we can further differentiate into e.g.: *standards* (although outside the realm of the individual researcher, a most important issue as it guides all kinds of unconscious communication patterns), *transparency* of the archive or database, *integration* of different sources and world-wide *access*.

On the dissemination (author's) side we distinguish *convenience* of the process, *visibility* and *retrievability*, which are important factors needed for *recognition*, *feedback*, and *time* to reach a reader.

Putting it in another way we come to the following functions to be fulfilled: a) The work has to be *authenticated*. The author deserves due credit and priority for his/her work, whilst the reader must be certain about the source. b) The work has to be *validated*. In order to provide credits to the author and to allow the reader to select a source, each article needs a quality check with regard to its correctness. This is normally done by an editorial board based on the concept of peer review. c) The article gets a *certificate* after validation, by acceptance in a journal. The journal name, publishers imprint, or the fact that a library decides to have it are important quality indicators. d) The articles are *collected* in field specific repositories (called journals), which is a precondition for consistent document management. e) The work has to be *indexed* and augmented with all necessary meta-data to make it retrievable. f) The work must be part of a *dissemination* system (presently by a journal subscription). Finally, g) the work must be *archived* for any possible future use. We stress the fact that all these functions are technology independent. Hence, in any novel electronic publishing pursuit, one way or the other, these functions have to be fulfilled, using available technology.

The above discussion indicates the starting point, as well as the constraints for novel electronic publishing endeavours. These are clearly reflected in the way new electronic journals are set up. In order to keep the integrity of the scientific process, all scholarly electronic journals mimic their paper ancestors. Obviously demands such as a clear editorial policy, including peer review, are kept in an electronic environment. More interesting is that the format of the articles is almost indistinguishable from paper versions. The only big change is the usage of hyper-links (mostly to bibliographic information, figures and tables) within the, still linear, text and to other documents or information sources. This means that mainly the logistics and storage capacities of computer networks are called upon and only partially the intrinsic capabilities of sharing information, linking and non-linearity. In the next sections we will further elaborate on these aspects which will bring us to our new concept of modularity which addresses the restructuring of articles in such a way that linking is more than just surfing but can become a tool in dedicated information

management.

III. FORM AND CONTENT

In an earlier publication [3] the idea of the dependence and changes of the structure of the scientific article dependent on the available technology was worked out. It was argued that following the historical development from orality to written texts, culminating in the printing press which enabled modern science, we are now entering a new phase in which again a medium with superior capacities will change the form of the knowledge representations. In this section we review briefly the main line of reasoning and will further develop its consequences, based on the concrete example of articles in molecular physics.

The present-day document with its linear, essayistic form, is a typical product of print-on paper. The central idea of the linear print-on-paper article is that it is structured like an oral report. In our normal parlance we also speak about "reading a paper" at a conference. However, talking and reading are two different things. Oral communication reflects certain pattern of discourse which an interlocutor must be able to follow for the duration of the speech. In the written article this pattern is still followed to a large extent, although nobody needs to (and in practise most readers don't) read a scientific article fully from top to bottom. The work is written as if the reader is indeed interested in the consumption of the complete content. The print-on-paper article is a closed information unit, which can be torn out of the journal and taken away, without losing its integrity. It is obvious that because the reader is unknown to the author, the author should play safe and make the unique article as comprehensive as possible. In reality however regular technical articles are not comprehensive but represent slices of a continuing research endeavour.

The general requirements with respect to the article as discussed in the previous section can be summarised as quality, clarity, relevance and efficiency. However, the question whether an individual reader considers the requirements to be fulfilled depends on his/her background and the particular goals in the actual state of the research process. Different readers read the same article in different ways. Kircz [6] made a distinction between various kinds of readers in order to make this fact more operational. The purpose of the reader may for example be to find some numerical data or some factual data, at another moment the desire is to understand a complicated part of the theory, or to be introduced in an entirely new field. Reading is goal oriented in the sense that the reader is looking for a particular kind of, not necessarily fully articulated, information. But in all cases the requirements, as discussed above, have to be fulfilled.

The modular form we propose takes into account the diversity of the readers' information needs. The obvious next step therefore is to investigate if we can identify the various kinds of information and knowledge represented in a scientific article, in order to restructure the presentation in such a form that different readers can read the same work in a different way. The intrinsic non-linear

form of electronic storage, explicated in hypertext approaches, is then the most prominent feature. So we try and define discrete modules of information which together comprise the content of a scientific article. A first outline of a modular structure and its comparison with the linear printed text was given in Harmsze et al. [7] and Kircz [3].

The goal is, if we have a collection of such, modular, articles in an electronic database environment, to provide the inquiring reader with an extra dimension for searching and specific reading. For instance: after a regular search on author's name and/or keywords in which a number of documents is identified, the reader can restrict the retrieved information to only one (or more) type(s) of module(s). For example, if the reader is only interested in the specifications of some apparatus, only the module describing the apparatus is supplied and not other modules dealing with e.g. the measured data, the theory, or the comparison with other works.

Our task is now twofold. Firstly, we have to develop a heuristic model that shows that the information in a regular scientific article can be cast in a modular form. Secondly, we have to ensure that every module can also meet the requirements mentioned in section II of this paper vis-à-vis certification, validation, etc. A new and exciting aspect is that in a modular model, the logistics between modules in one article as well as between modules in different articles is new territory. Obviously it is not just a matter of linking indiscriminately one module to another; we will see differences depending on the type of link.

IV. TOWARDS MODULARITY

As a new model of representation is best developed in terms of proven practice, in order to be understandable and acceptable, we analyse a set of regular, linear, physics journal articles. This test set is a coherent collection of regular journal articles in experimental molecular physics. The articles are all from the same research group and with the same principal author, who assists us in understanding the physics behind the articles. The research can be identified as mainstream high-level work, published in well-established journals. We can claim that they are prototypical for experimental physics articles.

In reading the articles we tried to follow and understand the physics in the articles as well as the discourse of the scientific argument. As the set of articles represent the output of an ongoing research programme, it immediately became clear that such a series of regular articles have many aspects in common. It is necessary for example to introduce briefly in each article the problem at stake in the context of the research programme as a whole. Naturally this is quite normal, as every classical article is considered an independent, comprehensive, entity in document space. We also see a lot of cross-referencing as items dealt with in an earlier article are used or referenced to in a later article. In an electronic environment where all articles are part of the same memory structure, there is no reason to repeat things unnecessarily, as all stored information is available on the same platform and can be retrieved when needed. This notion

immediately induces the structuring of information into three categories expressing the range of the information: micro, meso and macro.

We define the *microscopic* information as the specific information of a particular article. It entails those elements which really make up for the work presented. It is that information which warrants the publication as a new work.

The information which is shared by a series of articles within the same research programme, we call *mesoscopic* information. An example is the description of instrumentation what is used over a (long) period of time and/or is used for the measurement of features of a wide range of species.

The *macroscopic* information we consider the information that plays a role in a wider context of the scientific quest, e.g. information which can be found in textbooks.

This first division of information proves very useful, as a lot of the repetition of information can now be avoided. If something is already described in article one, in article two we only have to refer to that information plus a possible addition on how some aspects are changed. Of course, this demands that the original information is written in such a form that reuse is possible. This emphasizes the point that a modular model is not aimed at a simple recasting of existing articles, but at writing new articles in a fully electronic context. The demand of reusability induces specific demands on writing which in their turn must adhere to the general notions of validatability and certifyability mentioned in section II.

A standard linear article is normally built-up following a consecutive listing of section headings e.g.: Introduction, Methods (theoretical or experimental), Results, Discussion and Conclusions. This already reflects a structuring and suggests that simply cutting the article in its sections would be a good start. Immediately it becomes clear that this is insufficient. In a linear article the author has a style of argumentation geared to readers who are supposed to read the whole article. Hence, the argumentation and the different ingredients in the discourse are used throughout the whole body of the text. Normally a cut-out section of a well written article is not self-contained. In a modular model however, we demand the self-containment of each module. A typical example of the problems is that in an experimental article, technical parameters are not fully given in the section describing the measuring apparatus but can also appear in the sections on data acquisition and even in the final conclusions. So, a modular model is intended for a different structuring of the information contained in a linear article rather than for a direct translation of the linear article itself. Of course, in the analysis we can not step into the mind of the scientists who wrote the articles and start a new representation of the research results from scratch. The analysis we perform is forced to be based on published articles. This allows us to compare two representations of the same science.

The route we take is that we analyse the corpus of test articles according to different types of characterisation of information. We then cast this information in a first heuristic modular structure. In doing so we identify overlapping information and lacking

information that we have to add in order to make the modules self-contained and easy to read. After we have recast the articles in a modular form, we analyse the modules thus obtained. This way we reach an understanding of the intrinsic possibilities of modularisation, of the way modules can be defined best and of the demands on the writing of modules (which leads to explicit instructions to authors). We also get a good understanding of the linking structure of modules and the different kinds of links therein. In the following section we present the heuristic model, in the section thereafter we discuss our first findings.

V. THE MODULAR MODEL

In this section we first give a condensed overview of the set of modules we suggest as basis for our modular model. We then touch on the different types of links between modules. In the next section we discuss some first results.

A. Modules

The modular model we develop for the representation of scientific information in electronic articles, endows articles with a modular structure: a set of well-defined modules and their mutual relations. Modules are defined as units of uniquely characterised, self-contained representations of conceptual information. They can be separately located, consulted and read. For logistic purposes we define a special meta module, summarizing the other modules and their mutual relations. The structure has two classes of modules with some hierarchy between them. We distinguish at the lowest level, *elementary* modules, that are the smallest self-contained parts carrying an explicit characterisation. From there we define *compound* modules that consist of elementary modules or a number of (more detailed) smaller compound modules plus a summary of their components.

The definition of the different types of modules that make up a modular article depends on the characterization of the information. In our model we base the distinction of information on four complementary types of characterisation, which correspond to different aspects of the information, in the scientific domain at stake.

Firstly, and in our model most importantly, the information is grouped into units characterised by their so called *conceptual function*. These information modules express the role the information plays in the scientific problem-solving process reflected in the article. In first approximation they resemble typical sections of linear articles. Secondly, we classify and group the information by its *range*, as introduced in section IV, in microscopic, mesoscopic or macroscopic modules. Thirdly we classify the so obtained modules according to their *physics content* by a domain-oriented characterisation. In our analysis we use a rudimentary thesaurus of keywords, leaving this aspect of classification to standard indexing and information retrieval techniques. Finally, the modules are further characterised by a set of specified *bibliographical data* associated to them. As the

development on bibliographic meta-data is ongoing, we restrict ourselves to an minimal set for illustration only.

In our model the conceptual role serves as leading modularisation principle. Below we discuss matters in more detail, listing the main compound modules, reflecting phases in the problem solving process, as well as their constituent modules.

The start of the problem-solving process corresponds to the compound module *Situated problem*, which contains two elementary modules: *Situation*, providing the embedding or context of the article, and *Central problem*, stating the definition of the problem or the goal addressed in the article. The *Situation* may be a gentle introduction to the subject for non-specialist readers. Here we see immediately how the repetition of the presentation of the context in various articles, can be avoided by creating a mesoscopic *Situation* module, thus increasing the required efficiency. On the other hand the description of the *Central problem* is required to be concise and well-articulated, as it explains the reason for publication.

The large compound module *Methods* contains a description of the different methods used to solve the problem, as well as a discussion of the reliability and applicability of these methods. The theories and models used in the article are dealt with in the module *Theoretical methods*. In the same way *Experimental methods* are about the experimental setup and the measurements. The hardware and software, as far as they are used for computations or simulations and not for data-acquisition find their place in the module *Numerical methods*. In all cases the restrictions of the methods in connection with the reliability and applicability, e.g. pertaining to experimental precision or theoretical assumptions, have to be explicitly present. Also with respect to the demand of completeness necessary for self-containment, the level of detail required in this modules is such that the reader must be able to obtain sufficient information to use the method, although the full details may be made available via a link to a (external) mesoscopic *Methods* module.

The description of the results and a discussion about their reliability (such as the error of measurement) form the *Results* compound module. Within *Results* we distinguish the elementary modules *Raw data* and *Treated results*. *Raw data* contains the direct output from the measurements or calculations, which often cannot be included in articles published on paper. The presentation of the raw data in a machine readable form allows the reader to manipulate them, and plot them for example in combination and comparison to other data. In the elementary module *Treated results* or fitted data, an account is given of the data analysis and presentation, often in graphical smoothed form. *Results* can harbour various similar, but separate, modules when, for example, the same type of measurements have been performed on different samples or different instrumentation or settings are used for the same sample. The different results are then distinguished by their physics content and represented accordingly in different elementary modules. This fulfills the efficiency requirement for those readers who are only interested in specific results. The authors' demands for writing these modules are obviously given by a clear need for transparent data reduction and

well-defined error bars.

A difficult compound module is *Interpretation*, where the author seeks to explain the observations described in detail in the *Results* module, using theories or models explicated in the one of the modules of the *Methods* compound. This interpretation - both the process and its outcome - are both described and discussed, in separate *Qualitative interpretation* and *Quantitative interpretation* modules when desired.

The final compound module *Outcome* contains the elementary module *Findings*, which briefly recapitulates what has been achieved, including either an explicit answer to the central question or an explanation why it cannot be (entirely) answered based on this work. It also includes a module *New problems* in which a description is given of the problems that surfaced during the research presented. It might include old, yet unsolved, problems.

In order to display and clarify the structure and relationships of these conceptual modules, reflecting the problem-solving process, we have added an "information switchboard module", called *Meta-information*, which serves as a "linchpin" around which all modules are grouped. It comprises all standard meta-data like the author's name and address, publication dates and keywords. In this module we include an abstract summarizing the line of reasoning of the article as a whole, as well as a "road-map" of the modular article.

B. Links

Although modules are self-contained, they are not independent of their context, just like "traditional" linear articles. This interdependence of the information finds its expression in links representing the relations between modules, both inside and outside the article. Thus the modular article represents a network of information that is embedded in the network of all electronic scientific documents, in which readers can choose a path to suit their particular information needs.

The relationships between modules in a modular structure differ in function and in structure. In our analysis we have so far encountered and made explicit links of various different kinds. We have organised these in a preliminary taxonomy which distinguishes types of links that are 1) *organisational*, and types of links that are 2) *referential* in nature. A link between two modules is not necessarily unique and the full characterisation of a link may consist of complementary types.

The *organisational* link types include: a) a *hierarchical* type of links between compound modules and their components, b) a link type indicating whether the link is *external* (between different works) or *internal* (connecting modules in the same article), c) a type of link between information and the associated *meta-information*, and d) a *sequential* link type associated to a "sequential path" through an article. We have specified a sequential path running through all modules, based on the idea that some readers want a sequential report of all available

information. Other specialised sequential paths can be defined as well

The *referential* link types include: a) a type of links expressing *problem-solving dependency* relations. Results for example thus depend on the method used to generate them. A special case of such a dependency relation is the input-output relation between simple facts (such as numbers) represented in one module and their usage in another one. Other types are: b) a type of links expressing the difference in *range* between two modules, indicating which one is more general, c) a link type to indicate if a cited target module contains an *elaboration*, i.e. if it provides a more detailed explanation or argumentation, or more background, and d) links that mark a *comparison*, with the special cases of agreement and disagreement. Thus we can for example express in a link that the results in the cited module agree with the ones in the module at hand.

All these link types are invertible, meaning that they can be followed in the opposite direction. The link types indicating comparison and the internal-external links are symmetric. The other link types are asymmetric. The inverse, for example, of an external link leading to a more detailed and general (mesoscopic) module providing more background, is an external link leading back to the more specific (microscopic) module with a focussed summary.

With the hierarchical ordering of articles, compound modules and elementary modules mentioned above, and the representation of the structure of an article shown in its *Meta-information*, this explicit characterisation of the links should provide the reader with an insight in the structure of the information network, in order to meet the requirement of clarity. Furthermore, the characterisation can be used in complex search operations, allowing readers for instance to locate all results which agree with some specified ones.

VI. EVALUATION OF THE MODEL

After analysing only apart of the corpus, we are already able to articulate some findings. In the first place it becomes clear which parts of an article can be modularised relatively easily and effectively. "Common knowledge" of the researchers in the fields, i.e. information that is well understood, can easily be explicated in a separate module. In general, this type of information will be presented in an elaborate mesoscopic (or even macroscopic) module, which can then be cited from various microscopic modules only containing the specific information particular to the article at hand and a brief summary. Examples are the microscopic and mesoscopic modules about existing theory (*Theoretical methods*), and modules about methods that are usual in experimental practice (*Experimental methods*), but also modules formulating the central problem (*Central problem*) and explaining the general background of the scientific endeavour in a certain domain (*Situation*). The effect of the creation of such modules is that the author can conveniently recycle information, by citing an existing mesoscopic module, and that different kind

of readers can consult the modules efficiently: those fully aware of or not interested in the issue ignore the modules, those who do not need the full details restrict themselves to the microscopic module, whereas those interested in an elaborate account consult the complete mesoscopic module.

Also factual information, such as purely numerical or instrumental information, can easily be split off to a separate module as well. Examples of modules with a core of factual information are the *Results* modules, with the *Raw data* and *Treated results*. These results heavily depend on the methods used to obtain them, but they can be selectively searched for and used (cautiously) as simple facts to be inserted as input in a calculation or compared to new results. If such selective facts are to be looked up and referred to efficiently, they have to be presented in focussed modules, devoid of other information that happened to appear in conjunction with it in the article. Therefore the model distinguishes separate modules for only slightly differing results, via the criterion of the physics-based characterisation. For instance, if the same instrument is used for the measurement of a series of different sample we choose to present the results per sample in a separate module. These modules then share a lot of information (for example about the data acquisition and analysis techniques that were commonly used). However, in an electronic environment there is no problem with storing the extra volume caused by any overlap and, if well indicated, the software interface can hide it to the reader who consults more than one module of a similar type.

In the second place, we can identify information that cannot easily be presented in separate modules: information pertaining to the authors' struggle to explain their results and determine the validity of that explanation, in order to draw conclusions on the behaviour of some physical or chemical system. In the *Interpretation* module, descriptions and discussions of one or more candidate interpretations of the different results in the light of some (variants of) a theory, details on these theories, qualitative considerations, quantitative calculations, comparisons to other authors' results or findings, may all be intertwined. Thus arguments and explanations issuing from many different modules, in the same article or other articles, are brought together in a complex line of reasoning. Although structuring the Interpretation module is difficult, it is also important to try to do so as clearly as possible, in order to enhance the understandability which tends to be problematic for this type of information. Intertwined issues cannot be separated, but sometimes the qualitative interpretation can be isolated from the quantitative interpretation. The quantitative interpretation can contain long and involved calculations of intermediate results that have meaning in other contexts. These may be presented in a separate module, which can be consulted apart or be avoided by readers who wish only to follow the line of reasoning of the *Interpretation* as a whole, without being sidetracked by the computation of some values that are needed as input in the next step. The details of a mathematical digression that are not self-contained, and therefore do not form a separate module, may in the actual presentation of the module

still be hidden from first view, in order to preserve the clarity of the main line of reasoning. When different issues can be separated in different modules, the covering compound module should provide an overview summarizing the principle line of reasoning and clarifying how these issues fit together. To avoiding the risk of "drowning" in a complex *Interpretation* module, the essentials of the final interpretation are also summarized in the *Findings* module, which does not have to introduce new information, but rounds off the reasoning of the article.

In the third place, comparing the modularised versions to the original ones, we find that the articles we modularised tended to be somewhat larger in size than the original versions, in spite of the presentation of common information in external, mesoscopic modules. This is caused by the requirement of self-containment of modules, we saw some *Results* modules that report different measurements with the same instrument have identical parts. The addition of brief overviews in the compound modules summarizing their components also gives rise to some extra text. As all the modules are stored electronically the total size of the collection of modules is unimportant; only the module itself counts. The modular version also tends to be clearer, due to these extra overviews and the explicitness of the structure, which facilitates the access to that general information and which resolves (apparent) loops in the line of reasoning. Readers who wish to read selectively parts of the original paper article can generally do so with respect to the results, most of the experimental methods and the findings sections. However, the different types of information are always somewhat intertwined, such that selective reading is more effective in the modular version. The relevance to a reader can be determined more easily in the modular version, because of the detailed characterisation made explicit for the modules and the links between them. Locating and retrieving specific information should be easier in a modular, electronic environment in which more focussed entities with more precise characterisations are archived.

The requirements related to the roles publishers and libraries play can be fulfilled independently of the structure of the publications. We should mention that the process of peer review can even be facilitated by the availability of the modular model with a clear guided structure. Each module will have its own specific demands in relation to the various validation criteria. The criteria for data acquisition reports are, for example, distinctly different from the criteria for discussion of the interpretation and visionary outlooks.

VII. CONCLUSIONS

In this paper we presented the first results of a heuristic model for a new structure for scientific articles in an electronic environment. Instead of a classical linear representation, we proposed a modular representation, in which each module has a well-defined meaning, as well as a well-defined place in a modular web. Although the model is still in development, our

work indicates that it seems possible to rewrite normal articles in modular form, which is a precondition for developing a new way of writing scientific articles. The definition of the different types of modules and relations between them can be used to formulate a "writers guide" for the creation of modular articles right from the very beginning. Secondly, it becomes clear that modules grounded in "common information", and modules with a core of factual information, are easier to create than modules containing the more argumentative novel synthesis of theoretical and experimental insights. A consequence is that, as a first step towards the implementation of a modular system, the easy parts can be isolated from the main line of reasoning of regular articles and presented as self-consistent "digital appendices" with their own intrinsic validation and certification. We can think of appendices about instrumentation, raw data sets, and more or less standardised computational or theoretical methods. In fact, *Interpretation* modules contain the core of the line of reasoning of the article, that is supported by detailed explanation and justification in separate modules of the *Situation*, *Methods* and *Results* involved, and which is summarised in the *Findings*. This way, the evolution from linear paper based documents to non-linear electronically based sets of modules, can become a natural development perspective, for the enhancement of the effectiveness and efficiency of scientific communication.

ACKNOWLEDGEMENT

We would like to thank Jonathan Clark, David Greaves, Hans Roosendaal and Maarten van der Tol for helpful comments.

REFERENCES

- [1] William D. Garvey. *Communication: The essence of science*. Oxford: Pergamon Press, 1979.
- [2] Hans E. Roosendaal and Peter A.Th.M. Geurts. "Forces and functions in scientific communication: an analysis of their interplay." unpublished. Preprint on: www.physik.uni-oldenburg.de/conferences/crisp97.
- [3] Joost G. Kircz. "Modularity: the next form of scientific information presentation?" *Journal of Documentation* vol. 54. no. 2. pp. 210-235, March 1998.
- [4] H.P. Grice. "Logic and conversation.." In: *Syntax and semantics*, vol.3. *Speech acts*. P. Cole and J.L. Morgan, Eds. New York: Academic Press, 1975, pp.41-58.
- [5] Joost G. Kircz and Hans E. Roosendaal. "Understanding and shaping scientific information transfer." Invited talk. In: *Electronic publishing in science. proceedings of the ICSU Press / UNESCO expert conference February 1996*. Dennis Shaw and Howard Moore, Eds. Paris: Unesco 1996. pp. 106-116.
- [6] Joost G. Kircz. "Rhetorical structure of scientific articles: the case for argumentational analysis in information retrieval". *Journal of Documentation*, vol. 47, no. 4, pp. 354-372, December 1991.
- [7] Frédérique Harmsze, Maarten van der Tol, and Joost Kircz, "Naar een modulair model voor natuurwetenschappelijke informatie in elektronische artikelen. In: *Informatiewetenschap 1996. Wetenschappelijke bijdragen aan de Vierde Interdisciplinaire Conferentie Informatiewetenschap*. K. van der Meer, Ed. Delft: Stinfon , 1996, pp.53-72.