

## Identification of Electronic Documents: The Diversification of Document Supply and its Consequences for Document Description

Achim Osswald

Institute of Documentation  
Frankfurt/Main  
Germany

### Abstract

Much effort was necessary to come to an agreement on national and international standards for bibliographic description of printed journal papers. Electronic Publishing will complicate this situation. Not only are there different descriptions of a single physical document, but also different physical supply forms of the master document - no matter whether it is printed or electronic. The possibility to generate documents which have little or no reference to the concept of printed documents complicates the situation even more.

There are several solutions for identifying printed documents or electronic versions of those documents by special numbers like the ADONIS-reference-number or Biblid. In case these identification numbers are not used there is another solution left: a matching code. This, however requires standards for bibliographic description, too. But no agreement exists on the description of electronic documents, especially for those, which are published with little or no relation to the print concept. ISO 10956, which is available as Committee Draft (CD), seems a possible perspective but in spite of the diversification of document supply forms little explication is given in this CD as to the system requirements for using the documents. Solutions to this problem are set forth by the author.

## Identification of Electronic Documents: The Diversification of Document Supply and its Consequences for Document Description

### 0. Introduction

Ladies and Gentlemen,

most of you have a lot of experience in the delivery of documents printed on paper. As a condition of successful document delivery it is essential to give a specific bibliographic description for the identification of the ordered item. It took several hundred years to elaborate adequate rules for the problem of identification and description of printed material and there are still some problems left because the shape and the design of publications is still subject to change.

The standards which have been developed like RAK, the "Regeln für die alphabetische Katalogisierung", AACR2, the "Anglo American Cataloguing Rules 2" following the ISBD(G), the "General International Standard Bibliographic Description" offer different solutions to the problem of bibliographic description. Nevertheless, there has been little concern about the proliferation of "Non-Book Materials" (NBM). Especially the increasing variations and amount of electronic publications will make solutions necessary that offer assistance to those who have to handle electronic documents whether they are scientists who use them or librarians and information specialists who want to prepare them for usage.

My paper analyzes this developing situation and offers proposals for the description and identification of electronic publications. My main thesis is that electronic publications challenge the limits of the traditional bibliographic description of documents. This is because electronic publishing will make different types of documents possible. Such documents cannot be described only by traditional bibliographic means. To manage this problem we will have to go beyond the traditional bibliographic view. This will facilitate different kinds of document supply made possible by those different types of documents.

By searching for such solutions my point of view has been a user oriented one. My proposals focus on the needs of users and pay less attention to sophisticated cataloguing. Clients' needs should and do have priority. In

other words, "standards must reflect the real world of applications beyond libraries and publishing" (Brownrigg/Butler 1990:23).

Let me give you an overview of the points I want to cover:

1. I will describe the changing types of documents made possible by electronic publishing.
2. I will give an overview of existing ways to identify documents by special document identifiers or matching codes. Their usefulness will be reported and their application to electronic publications will be described.
3. There will be a close look at the description and identification of electronic documents offered by ISO 10956.

### 1. Types of documents - their diversification as a consequence of electronic publishing

In the print-on-paper world we have the following situation (see figure 1):

Authors elaborate ideas and findings which are published as documents. That means the original data is represented in a document (e.g. a journal article or the published version of a paper). Several printed issues might exist even of one paper. Furthermore, every issue might be described according to several different bibliographic rules and their variants.

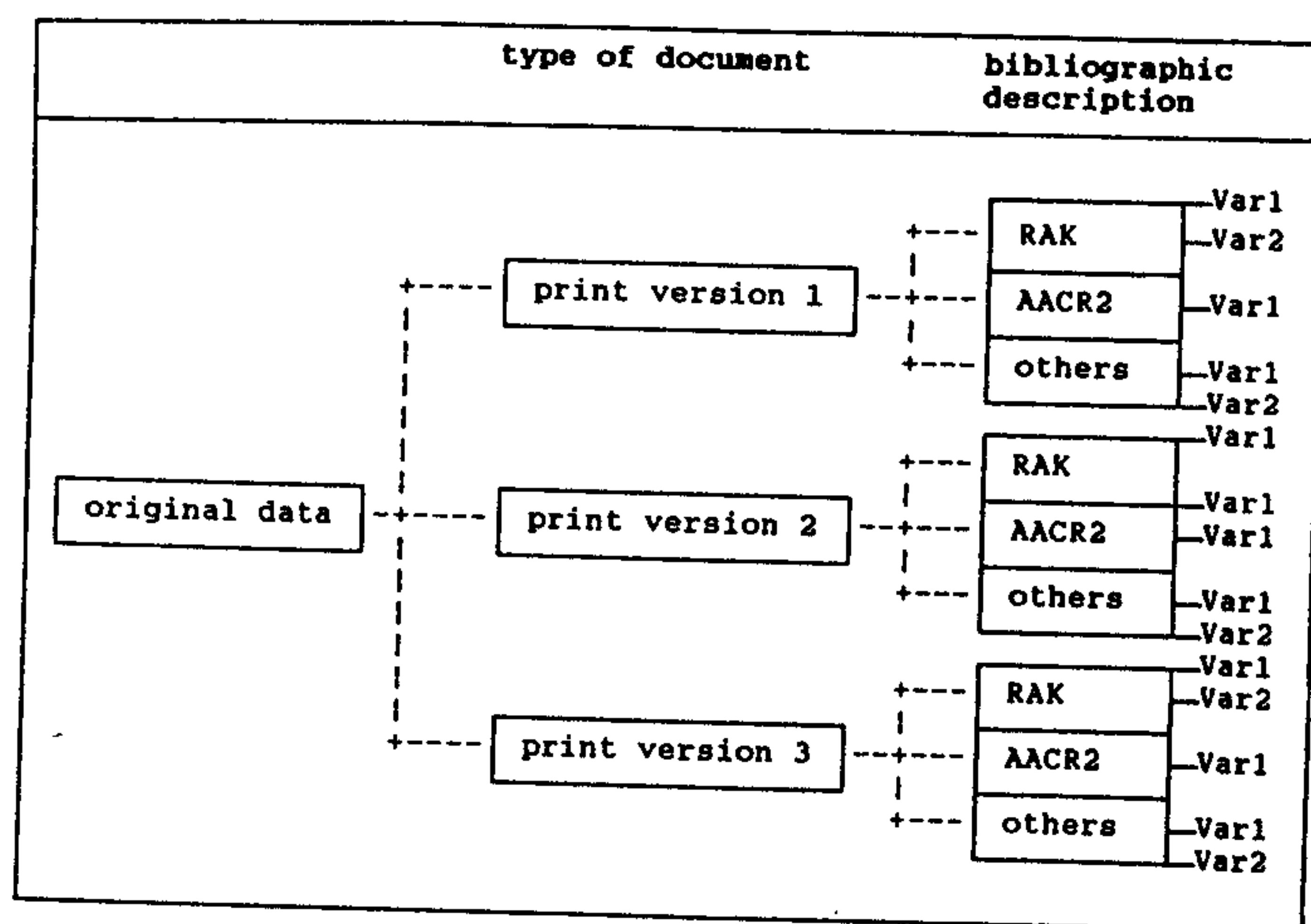


Figure 1: Types of documents and their different bibliographic descriptions

Electronic publishing complicates this situation by offering - at least - three additional types of documents which have implications in document description and identification. They are

- a) electronic parallel publications,
- b) publications which are related to the print-on-paper concept,
- c) publications with little or no relation to that concept.

Let me give you some explanations to these types of documents<sup>1</sup>:

#### *Electronic parallel publications:*

Most of the electronic publications which are available are parallel publications. A lot of them are 1:1 duplicates, which means they have also been published on paper, either as a book, a newspaper or some other kind of publication or parts thereof. Their bibliographic description seems to be no real problem as only the specifications of the electronic type of supply have to be added. But there is no standard for the bibliographic description and ordering of that kind of documents even though there are explanations and recommendations in reference to NBM.

#### *Publications which are related to the print-on-paper concept:*

Electronic newsletters, bulletin boards and other kinds of electronic publications<sup>2</sup> still have some relation to the print-on-paper concept. However, under normal circumstances there is no real equivalent in paper-based form. Some of these publications are dynamic documents, which means the computer environment allows to modify existing parts of the document or add new ones.

Existing NBM-rules are simply inadequate for giving a correct "bibliographic" description of these dynamic documents because these rules do not require data about the system requirements for the use of electronic documents.

#### *Publications with little or no relation to the print-on-paper concepts:*

The slowly but steadily increasing amount of publications based on the hyper-concept calls for the description, identification and supply even of

<sup>1</sup> See for more details about those types of documents Osswald 1991

<sup>2</sup> Whether the term "publication" is justified for this kind of material or not should be discussed in another paper.

those documents. Because their characteristics allow different views on the knowledge basis or the document derived from it there is a need for referencing those views. Additionally, there are some other unsolved issues like the adequate description of the compression ratio of pictures or audio-sequences.

*To summarize:*

In my opinion we have no adequate solution yet to the problem of bibliographic description of electronic publications or parts thereof. Nevertheless the user has somehow to order and get them. Additionally, the user and we as information specialists are supposed to preserve the integrity of the intellectual process. Therefore we as information specialists need to create the conditions for the possibility of clear and unequivocal references to these electronic resources.

So what we are really talking about is the unequivocal identification of electronic publications. This, in turn, depends on the power of special document identifiers.

## 2. Document identifiers

Performing a document order requires not only receiving the right data in the right sequence and with the right punctuation but also implies the need to solve the logical ambiguities typically contained in the clients' order. Doing this electronically is the most rigorous way of solving this problem. There are two different methods of identifying documents electronically:

- a) The creation of a special document identifier like the **ADONIS-reference number** or **BIBLID** (the ISO standard 9115 for "Bibliographic identification of contributions in serials and books") followed by a matching process with an index of identifiers representing documents of a stock;
- b) a **matching process** between the (bibliographic) document description and the data of the document order or between parts of both.

The concept of the **ADONIS-reference number** has been well explained in several papers by members of the ADONIS-staff (Stern/Campbell 1988, Stern/Compier 1990). Therefore I don't need to explain the way it is constituted. But the combined code of the ADONIS-reference number created for the electronic parallel publications gathered on the ADONIS disks only applies to a small sample of journal articles. That may be the reason for the fact that - as mentioned by Mr. Korwitz - only one commercial online database (EMBASE) refers to this code, severely restricting the usefulness of this code.

**BIBLID** is based on a similar concept but is created out of other elements of a document description. It is available as an ISO standard, for books as Book-Biblid and for serials as Serials-Biblid. Publishers are supposed to create the code in reference to the ISO standard and print it on the top of an article or a contribution to a serial publication. The problem of this visually readable code is that - until now - it has not been accepted by publishers. As a consequence BIBLID has not been able to be applied for searches in online databases and therefore it is still no real and practical help for document identification and ordering.

Both the ADONIS-reference number and BIBLID are created out of traditional bibliographic data and rely on the print-on-paper concept. They are approaching their conceptual limits when used for documents which have little or no relation to this concept.

Recently<sup>3</sup> there has been the suggestion of a concept without any relation to the print-on-paper concept, i.e. a transfer of the ISBN-system to electronic information products<sup>4</sup>. If such a system were to be accepted in the same way the ISBN-system<sup>5</sup> was accepted in the print sector this would be a great step toward the identification of electronic documents.

The **second concept** of identifying documents also uses data contained in bibliographic document descriptions. It can be applied to every kind of data. The basic concept is to compare (string by string) the (bibliographic) data of the document description with data of the document order. Automating this process is only possible if an electronic version both of the (bibliographic) document description and of the (bibliographic) data of the order exists.

Most of the existing matching codes for duplicate checking offered by online hosts are based on this concept. Their usefulness and precision differ widely.

The most sophisticated one seems to me the **USBC**, the "**Universal Standard Bibliographic Code**", which was created during the DOC-MATCH projects and tested by the project QUARTET in an ADONIS test bed (Ayres et al. 1990, Yannakoudakis/Ridley 1989). It is constituted of four elements:

"The first is a letter representing the year of publication. The second element is six characters long representing the least frequently occurring characters in the author's name. The third element is taken from the pagination using the last two digits of the page on which the article starts. The fourth element is seven characters long representing the least frequently occurring characters in the title. Any blank spaces are filled with stars." (Yannakoudakis/Ridley 1989:4)

Depending on the quality of the data of document orders the USBC can be qualified to different levels of exactness.

<sup>3</sup> See Password 1991.

<sup>4</sup> There are plans for the further development of the ECHO-database DIANEGUIDE to a product called I'M-GUIDE. The directory covers not only databases and databanks, but also CD-ROM and CD-I products, diskette bases, image banks etc.

<sup>5</sup> It is reported that 75% of the American and 60% of the European publishers apply the ISBN-system (Password 1991:8).

Like other codes USBC refers to the concept of traditional bibliographic description of documents. But the problem we have to face by the increase of electronic documents is that we don't know how those documents can and should be described. Consequently there is no agreement on the data which might be used for document identification regardless of whether the identification should be processed automatically or intellectually.

### 3. Bibliographic references to electronic documents or parts thereof

As explained electronic documents have several specifications in both form and content which originate in the computer environment in which they are created, maintained, and disseminated. As ISO 10956 explains in its introduction: "Although an electronic document may stylistically resemble a print publication, such as a monograph, a serial, or an article or chapter, the physical characteristics inherent in printed publications do not always apply to electronic forms." (ISO 10956:Introduction) To create an adequate description of the electronic material we need to reflect on their special identity which means to "understand not only their physical characteristics ..., but also the type and amount of information that the user will need to identify the material." (Rogers 1982:23)

To give assistance to this process there has been a Working Item (WI) added as Number 117 to the programme of work of the TC (Technical Committee) 46<sup>6</sup> of the ISO. Since 1986 it has been developed by a working group of subcommittee (SC) 9 which concentrates on the "Presentation, identification and description of documents". It is now under study as ISO Standard 10956 Committee Draft (CD). As this document is still subject to change my citations and comments have a preliminary status.

To quote from the chapter "Scope" of ISO 10956 the standard "is intended for use by authors and editors in the compilation of references for inclusion in a bibliography, and in the formulation of citations ... . It

<sup>6</sup> ISO/TC 46 Information and Documentation defines its tasks and working scope like this: Standardization of practices relating to libraries, documentation and information centre, indexing and abstracting services, archives, and information science and publishing.

does not apply to full bibliographic descriptions as required by librarians." ... "It sets out a prescribed order for the elements of the reference and establishes conventions for the transcription and presentation of information derived from the source document." (ISO/CD 10956:1)

As long as no other agreement on bibliographic references of electronic documents exists we should use this chance offered by the (draft) standard to gather experience with the practicability of its rules. Using the standard citation of electronic documents might offer additional possibilities of creating automatically processed document ordering.

The concept of ISO 10956 is very similar to other rules for the bibliographic description of documents like e.g. the AACR2 rules for non-book material. Nevertheless, it tries to keep in mind the special identity of electronic documents.

It is not the purpose of my paper to explain the standard in detail but I want to explain the basic concept and the essential problem which arises out of this concept for the description of electronic documents.

As in other bibliographic rules the data for the description derives from the item itself. This approach seems to me a real problem as - in most times - the users' needs regarding the item go beyond that information.

If we look at the conditions (e.g. the hardware and software requirements) under which programmes or electronic publications are used you can imagine how important the information about "system requirements" is.

Like the "Note Area" given by AACR2 for non-book material, e.g. machine-readable data files, the standard associates essential information about the system requirements to an optional field under the label "Notes".

This association has its origin in the basic concept of "autopsy" which means the use only of data which can be taken out of the source itself<sup>7</sup>. From this point of view data for system requirements doesn't exist or seems to have a marginal status but in reality it is essential. An example given in the standard may explain this inadequate situation:

"Books in print plus: the complete Books in print system on compact laser disk database on CD-ROM. 5th ed. New York: Bowker Electronic Publishing, Jan. 1990-. Updated quarterly. 2

<sup>7</sup> See for example RAK and RAK-WB § 115.

computer disks: colour, 5<sup>1</sup>/<sub>4</sub> in. or 3<sup>1</sup>/<sub>2</sub> in.; 1 compact disk. Accompanied by: 1 manual. (Books in print plus series). *System requirements: IBM PC, XT, AT, PS/2 or full compatible; 640 K memory; DOS 3.1 or higher; hard disk recommended or 2 5<sup>1</sup>/<sub>4</sub> in or 3<sup>1</sup>/<sub>2</sub> in. floppy drives; monochrome or colour display; compatible MS-DOS Extensions device driver required; CD-ROM players supported include Amdek Laserdrive 1000, Hitachi 1502 or 1503 or 2500 or 3500, Philips LMSi 100 or 110 or 121 or 131 or 201 or 210 or 212.*" (ISO/CD 10956:18)

To know what kind of system requirements exist to read and use an electronic publication is not only an essential point for the user but also for all those who are concerned with document delivery. Therefore it is not acceptable that information about the system requirements is called "optional". According to the ISO-standard there are only some specifications as to which components of a hard- and software configuration are to be described. The supplementary field "Information concerning system requirements" is only explanatory. These explanations should "include the specific make and model of computer on which the program is designed to run; the amount of memory required; the name of the operating system and its version; the software requirements; and the kind and characteristics of any required or recommended peripherals. ... To reduce ambiguity, it is suggested that the words 'System requirements:' or an equivalent precede this information." (ISO/CD 10956:10)

There is only a small chance for a librarian or other staff of a document supply centre to get all the relevant data of the technical environment necessary to use the electronic publication. Such information can "only be achieved with the cooperation of those who produce, store and distribute computerized information" (Dodd 1982:48). To reach that objective we have two possibilities:

- a) The system requirements may be added to the source itself which may thus follow the concept of "autopsy". But there is catch: Reading about those "system requirements" presumes that the work, which remains to be done, has already been done. The data has been read in a computer environment and essential system requirements have been fulfilled.
- b) Information about the system requirements should be added on obligatory accompanying material or on an additional label accompanying the storage medium.

The system requirements themselves can be shortened by creating macro descriptions for typical configurations of hard- and software. Such configurations might be agreed on by publishers, user groups and - very essential- by the information industry. But until now no such initiative has been forthcoming. Even the MAJOUR SGML-document type definition for article headers which has been developed by the European Workgroup on SGML (EWS 1991) doesn't offer a solution to the problem yet. Perhaps the conciousness of the problem will increase in the next years as electronic publications seek more importance on the information market. In my opinion the time to think about it has come.

Ladies and gentlemen, thank you for your attention.

#### Literature:

##### **Ayres et al. 1990**

Ayres, F.H.; Huggill, J.A.W.; Ridley, M.J.; Yannakoudakis, E.J.: DOC-MATCH: Automated input to ADONIS. - In: Interlending and Document Supply 18(1990) 3, 92-97

##### **Brownrigg / Butler 1990**

Brownrigg, Edwin; Butler, Brett: An electronic library communications format: a definition and development proposal for MARC III. - In: Library Hi Tech 26(1990) 3, 21-26

##### **Dodd 1982**

Dodd, Sue A.: Cataloging machine-readable data files. An interpretive manual, Chicago 1982

##### **EWS 1991**

European Workgroup on SGML (Hrsg.): MAJOUR. Modular application for journals, DTD for article headers, o.O., 1991

##### **ISO 10 956**

Information and Documentation - Bibliographic references - Electronic documents or parts thereof, Committee Draft ISO/CD 10 956, 1991-01-21

##### **Osswald 1991**

Osswald, Achim: Perspektiven des Elektronischen Publizierens in Deutschland. - In: Nachrichten für Dokumentation, 42(1991), 115-127

##### **Password 1991**

Eine ISBN-Nummer für elektronische Informationsprodukte? - In: Password (1991) 3, 8

##### **Rogers 1982**

Rogers, JoAnn V.: Nonprint cataloging for multimedia collections. A guide based on AACR 2, Littleton, Colorado 1982

##### **Stern / Campbell 1988**

Stern, Barrie T.; Campbell, Robert: ADONIS: the story so far. - In: Oppenheim, Charles (Hrsg.): CD-ROM, Fundamentals to applications, London u.a. 1988, 181-219

##### **Stern / Compier 1990**

Stern, Barrie T.; Compier, Henk C.J.: ADONIS - Document delivery in the CD-ROM age. - In: Interlending and Document Supply 18(1990) 3, 79-87

##### **Yannakoudakis / Ridley 1989**

Yannakoudakis, E.J.; Ridley, M.J.: The DOCMATCH project. Automating document delivery by linking references to full text(sic)databases. - In: Outlook on Research Libraries 11(1989) 9, 3-7

Libraries and electronic publishing : promises and challenges for the 90's ; Festschrift in honour of Richard M. Dougherty / 14th International Essen Symposium 14 Oct. - 17 Oct. 1991. Ed. by Ahmed H. Helal ... Essen, 1992. - 204 S.

(Veröffentlichungen der Universitätsbibliothek Essen ; 14)

ISBN 3-922602-15-0

NE: Essen Symposium <14, 1991>; Helal, Ahmed H. [Hrsg.]; GT; Dougherty, Richard M.: Festschrift

© 1992 by Universitätsbibliothek Essen

Essen University Library

Organization and publication: Doris Pohl

Layout: Annelore Ax

Printed: Postberg Druck GmbH, Bottrop

ISSN 0931-7503

ISBN 3-922602-15-0

Authors are responsible for the choice and the presentation of their manuscripts

## Table of Contents

Preface . . . . .	IX
Agenda of the Seminar . . . . .	XVII
List of Participants . . . . .	XXI
Participating Vendors . . . . .	XXXVII

### **J. Andrew Braid**

Barriers to the Introduction of New Technology . . . . .	1
--	---

### **Richard M. Dougherty**

Nudging a Dinosaur in Order to Avoid Library Extinction . . . . .	17
---	----

### **Johan van Halm**

Are Libraries Ready for Electronic Document Delivery? . . . . .	31
---	----

### **Ulrich Korwitz**

Electronic Journals and Document Delivery: Chances and Risks . . . . .	39
--	----

### **Bonnie Postlethwaite**

The Global Library: The Impact of High-Speed Networks on Libraries . . . . .	55
--	----

### **Christian Heinisch**

Integrated CD-ROM-Retrieval in Networks: How to Automate the Post-Processing of Downloaded Export Files . . . . .	73
---	----

### **Achim Osswald**

Identification of Electronic Documents: The Diversification for Document Supply and its Consequences for Document Description . . . . .	83
---	----

### **Bernard Gallivan**

Full Text Imperatives for the National Library of Scotland . . . . .	97
--	----

<b>Nada Cucnik Majcen</b> Electronic Publishing: Possibilities, Barriers, and Environment in Slovenia and Past Yugoslavia . . . . .	115
<b>Hans-Eric Malmberg and Elisabeth Ehrnström</b> State of Art and Current Trends in Library Automation in Scandinavia . . . . .	123
<b>Donna L. McCool and Maureen Pastine</b> Libraries and Electronic Publishing: Issues for Academic Library Administrators to Address . . . . .	131
<b>Sigrid Reinitzer</b> Use of Library Material in Both Paper and Electronic Media in an University Library . . .	145
<b>Jennifer Walton</b> Information Services for the Public . . . . .	159
<b>Ruth Wüst</b> Optical Disk Based Document Delivery at the Congressional Research Service of the Library of Congress . . . . .	167
<b>Eva Bertha and Erich J. Schwarz</b> Hypermedia in the Library: Design Guidelines Based on Assessment Tests . . . . .	179
<b>Maurice B. Line</b> Conclusions and Comments . . . . .	195

IX

14th International  
Essen Symposium 1991

Essen University Library

**Libraries and Electronic Publishing:  
Promises and Challenges for the 90's**

14 October - 17 October  
1991

**Preface**

---

The 14th International Essen Symposium 1991 was held at Essen University Library during the period from 14 to 17 October, 1991, with full registration of 91 participants from 14 countries. The goal of Essen Symposium is, as in the previous symposia, to bring together internationally recognized librarians and library automation specialists to discuss new developments in electronic publishing. Librarians, publishers and users are the components of this co-existence.

The potential impact of electronic publishing and expected changes in existing library procedures will face us regardless of possible acceptance by librarians or disruptions to existing databases. Electronic publishing will strengthen sharing of resources and indirectly stimulate an increase in usage for all libraries and information centres; the final aim is to provide universal access to information.

This year the symposium was held in honour of Dr. Richard M. Dougherty, an internationally recognized authority on library management and information technology. The theme of the 14th International Essen Symposium was "Libraries and Electronic Publishing : Promises and Challenges for the 90's".