

# Electronic Publishing in Japan - a review in 1990

FID/CAO Workshop on Electronic Publishing, Seoul, November 1st, 1990<sup>1</sup>

*Japanese characters*

NAITO Eisuke<sup>2</sup>

Electronic publishing, vaguely defined area of activities and vaguely defined scope of products, is consisting of electronic or computerized processing of text, images, graphics for publication in one of electronic forms and also in a traditional printing on paper. As it is vague in scope, topics of this paper are of diverse interests.

Technological developments related to electronic publishing are rapid in Japan. Activities and events by various organizations, both in the private sector and the public sector, are not few, and interlaced.

## 1 JAPANESE DATA FOR ELECTRONIC PUBLISHING<sup>3</sup>

Enormous amount of digitization of Japanese text is progressing by word processing machines that are sold millions every year. However, the digitized data is not yet fully utilized because of the existence of technical problems such as incompatibility among Word processor, personal computers, and mainframe

computers. Another major obstacles against the full utilization of the digitized Japanese text are variety of complex requirements of Japanese printing composition. This problem is characteristics proper to Japanese printing convention, in other word, they are proper to Japanese printing culture with strong relationship to Japanese language. The situation, that is the composition information produces burden on exchange of electronic text data, is common to most languages in many countries.

### 1. 1 Background

The major topic of this presentation is concerning electronic processing of Japanese text and their printing composition. In the last 20 years, computer processing of Japanese scripts has been developed with extraordinary speed and acceptance, and huge amount of Japanese text are being digitized. For example, major Japanese newspapers are now printed fully by computerized type-setting (CTS), and another example, the Toppan, one of the leading printing firms in Japan, took off the movable type made of lead, and about 80% of products is set by CTS. The remaining 20% is set by

---

<sup>1</sup> revised as of May 1991.

<sup>2</sup> Professor, NACSIS, 3-29-1, Otsuka, Bunkyo-ku, Tokyo 112, Japan Fax:+81-3-5395-7064; Tel: 3942-6958  
Opinions expressed in this paper do not necessarily represent those of institutions cited.

<sup>3</sup> Chapter 1 is based on the following papers: OGASAWARA, O & KOMACHI, Y., *Overview of Japanese Composition*. a document submitted to ISO/IEC JTC1/SC18/WG8 in January 1989. and TANAKA Yoichi & FUKAMI Takushi. *Japanese Word Processing and Printing*. a paper presented at the 3rd Forum for Standardization of Information Technology, Singapore, December 9, 1989.

manual photo-typographically. Recently, CTS developments have advanced to printing of comics, weekly magazines that have been excluded from CTS processing.

### 1. 1. 1 Computerized Type Setting (CTS)

Document processing of Japanese text is categorized into three groups: Document processing in printing business; Desk top publishing (DTP) among corporate bodies; and Word processing for personal use.

Technology of the first group, document processing in printing business, is represented by computerized type setting. As predicted, there still exists incompatibilities among the CTS systems made by these manufacturers. Up until now, there is no compatibilities among fonts of various CTS system manufacturers. Then, Japan Federation of Printing Industries established an industry standard of font in 1988, which will be expected to be applied by manufacturers who deal with font.

Most of the CTS systems currently in operation are mini-computer-based, however, newspapers and major printing companies who require a capability of mass processing of data have their own CTS systems based on mainframe computers. The CTS system of main frame machines are used to process the printing composition of dictionaries, encyclopaedias, products catalogs, directories, etc. Although CTS data of this kind are often applied to produce CD-ROMs as a second product, there is little compatibility among CTS data.

### 1. 1. 2 Desk Top Publishing (DTP) among Corporate Bodies

In 1988, there was a great rise of interests in Japan on desk top publishing following to the business promotion of DTPs in the USA. Now,

there are many software and hardware for DTP including turn key products in Japanese market for processing Japanese documents.

DTP technology is also important to printing companies. Japanese printing firms apply DTP technology for printing materials such as weekly magazines of which layout and styling have to be highly flexible and text-editing has to be done in a very limited time. For example, both the Dai Nippon Printing and the Toppan Printing installed DTPs in the offices of publishers, and they are connected to the printing factories by telecommunication links.

There is yet little success to implant the capability of handling Japanese language to DTP systems based on English language. Transformation of English DTPs into Japanese DTPs is not simple as it seems at a glance, because of differences between English and Japanese languages. So that requirements of Japanese printing composition and requirements proper to handling Japanese texts should be assimilated in systems design of Japanese DTPs.

### 1. 1. 3 Word Processing for Personal Use

In contrast to word processors in the USA which are mostly based on personal computers (PCs), Japanese word processing machines, a variation of personal computers but solely devoted to word processing, are more common. "Ichitaro" is the name of the best selling word processing software of sale with millions to be used on personal computers.

The top priority of word processors' function is to provide a quality printing, and there are word processors with special keyboards for input process, 64 dots font, outline fonts, etc. These word processors are relatively cheaper than personal computers, with their rich

capabilities and functions such as special keyboards for efficient Kana-Kanji conversion, etc. Enormous number of word processors have been sold in Japan. Compatibility among word processors is also a problem for exchange the resulting documents in machine-readable form. At present, only character codes can be exchanged among various word processors in MS-DOS file for the minimum convenience.

## 1. 2 Characteristics of Japanese Printing Composition

### 1. 2. 1 Japanese Characters

Scripts used in Japan are: Kana (Hira-kana and Kata-kana) as phonetical scripts that are used in mixture with Kanji (Sino-Japanese). When there is no proper Kanji to express concepts, Hira-kana are used to describe sound. Kata-kana are used in special cases such as imported words as "Television". Kana can be expressed on computer by *JIS X0208 Code of the Japanese Graphic Character Set for Information Interchange* that is conformative to ISO.

However, when character codes of JIS X0208 are used with ASCII, it is so cumbersome to process them that the Shifted JIS code is used as an expedient method. The Shifted JIS is the majority of character code of

personal computers in Japan. In Japanese printing, the width of character/font of Hira-kana and Kata-kana are basically uniform, as same width as Kanji. However, the pitch between fonts may be extended or shortened according to composition requirements.

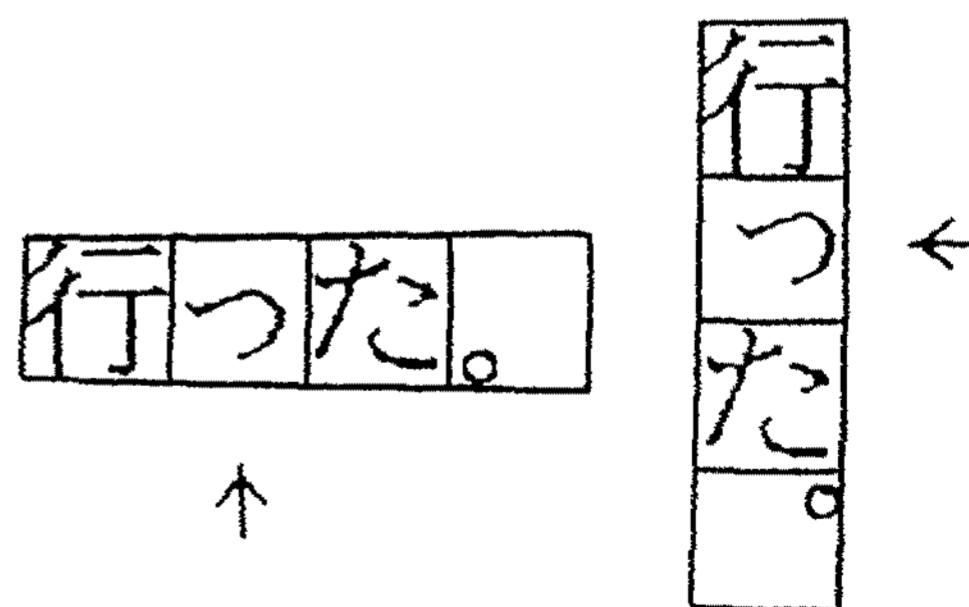
### 1. 2. 2 Vertical and Horizontal Composition

There are two types or methods of composition in Japanese composition. One is a vertical composition, unique to oriental materials, the other is a horizontal composition, similar to Latin materials. These methods are mostly common to Chinese and Korean documents. Most of general documents are by the vertical composition, however, documents that include Latin alphabets or numerics are by the horizontal composition in the case of scientific journals or office printing. The differences between vertical and horizontal composition is not only to rotate the font 90 degree, but also to change the expression of text. There are two methods for a text string of Latin alphabets: one is to rotate, the other does not, in other word, one is the Latin alphabets that used as a part of Japanese text, and the other is the Latin alphabets as used in the horizontal composition.

昭和61年4月11日(金) 14:00-17:30  
 (社)日本事務機械工業会 第一会議室  
 (順不同・敬称略)

昭和61年4月11日(金) 14時-17時30分  
 (社)日本事務機械工業会 第一会議室  
 (順不同・敬称略)

Expression of Time



Double Consonants Characters

In the vertical composition, chapter numbers are in Kanji numerals, and Arabic numerals are rarely used for chapter numbers. In Japanese writing, punctuation marks, and small Kanas such as double consonants and contracted sounds are used to support the description of pronunciation.

These small scripts look so strange when they are rotated 90 degrees for changing composition from vertical to horizontal and vis-a-vis that there are two types of fonts for vertical and horizontal composition. However, there is no distinction in the JIS Kanji Code for vertical and horizontal compositions.

### 1. 2. 3 Ruby (Japanese Ruby)

Ruby (Japanese Ruby) is used to express the pronunciation of a Kanji script or text. They are attached to the right hand side of a vertical text, above of a horizontal line.

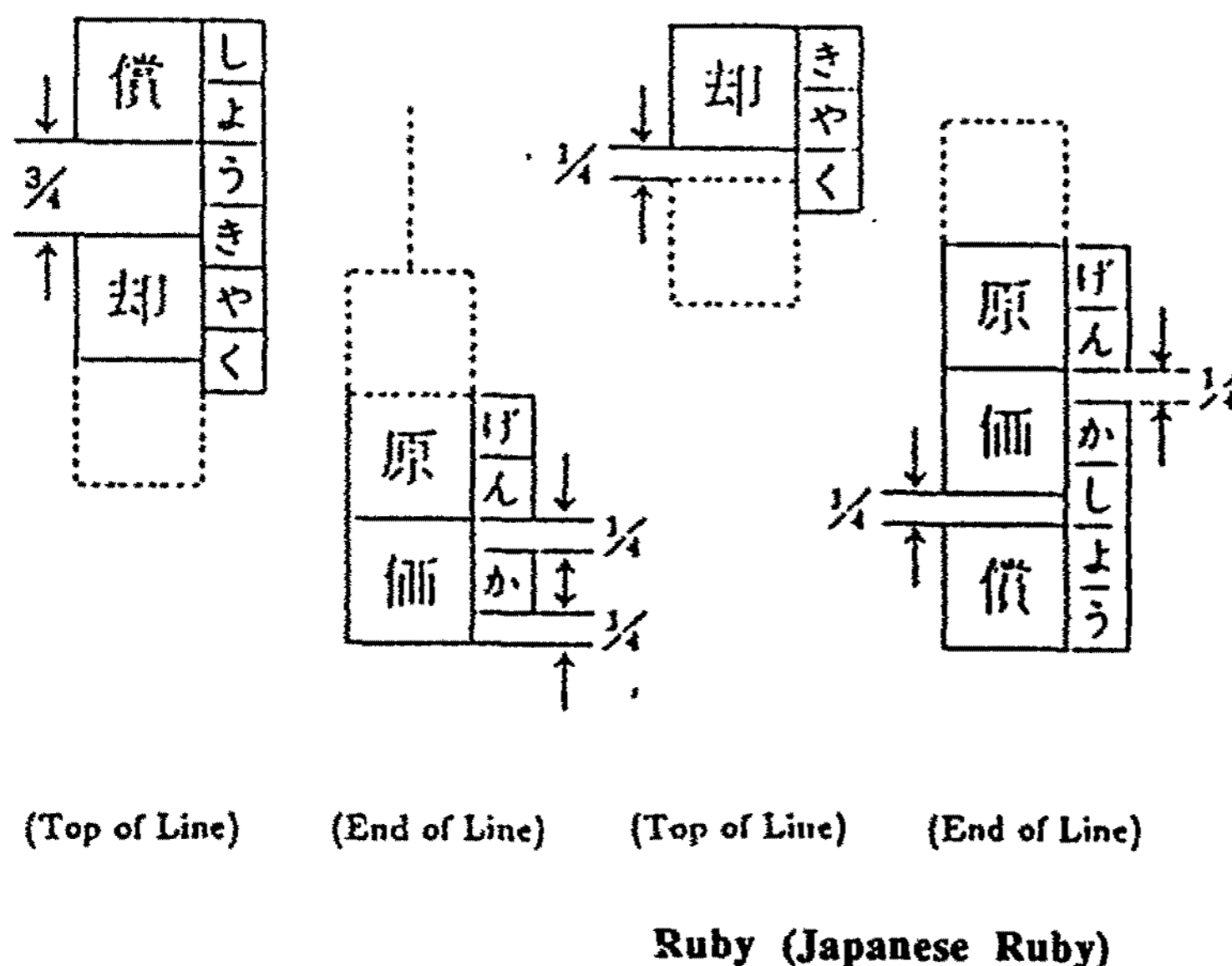
Each Japanese Kanji (Sino-Japanese) has its pronunciation according to its Chinese origin or Japanese origin or unique pronunciation proper to Japanese language. Japanese pronunciation

is relatively simple so that there exists many characters that have same pronunciations. For example, there are more than one hundred characters that have the same pronunciation in the JIS X0208.

Japanese Ruby plays an important role of distinguishing the pronunciation of Kanji in a text string. Ruby in Kata-kana is also attached to Latin character strings as their pronunciation, or Ruby in Latin character is attached to Japanese text to specify the original word spelling.

Font for Ruby is normally a half size to that of the main character, therefore, two Hira-kana in half a width as Ruby are attached to the main character. In some cases, there are three Ruby characters to one main character, then composition becomes very complexed. One overflowed character in a three-character-Ruby causes special attention. At the top of line or in the end of line, the floated Ruby (and the closing bracket) should be taken care of.

At the same time, special care should be taken to accommodate the combination of Ruby



に馬肉屋のある四辻で、辻の向側には曹洞宗東清寺と  
刻した石碑と、玉の井桶の鳥居と公衆電話とが立っ  
ている。わたしはお雪の話からこの桶の縁日は月の二  
日と二十日の両日である事や、縁日の晩は外ばかり賑

遣欧使節

開港開市の延期を使命とす

る遣欧使節(正徳府内保徳・副使  
松平康政・目付京臣)

明<sup>西</sup>は、文久元年十二月二十二日、イギリス軍艦オーディン号で横浜を出帆した。当時幕府内には、使節の海外派遣が攘夷派を刺激するものとして、難色を示すものもあつたが、安藤は「群議を排しこれを断行するに決し」とたう。かれが使節に与えた訓令では、たんに当面の外交交渉のみにとどまらず、各国の政治・学制・軍制を調査させているのによると、西洋の制度・文化を取り入れて、国内の改革に寄与させようとしたことがわかる。

いっぽう、オールコックは賜暇帰国を目前にして、老中久世<sup>ヒサヒサ</sup>広周と会見した(安藤は坂下門外  
の安で自四欠席)。文久二年(一八六二)二月十六日の最後の会見にさいして、久世

#### Inline Notes

to compound words or phonetic equivalents (false substitute characters) to which the length or the width is not equal between the combination. Ruby are allocated with space(s) to match the width of main string. On the contrary, Ruby is floated in the case of more than three Ruby, spaces are allocated in the main character string to match the width of the overflowed Ruby. There is no authoritative convention for allocating Ruby or allocating spaces to the main string, and editors or printers select their options according to their local conventions.

#### 1. 2. 4 Inline Notes (Warichu)

Notes appearing in the text are called "Inline notes", and normally composed in a smaller characters. When they are set in the text, they are set in two lines within the width of the main line. The two lines for the inline notes are kept in the same length as much as possible. When the inline notes run-on to the next line, each two lines of inline notes are set to be the same

length. Examples are descriptions of pronunciation in a dictionary.

#### 1. 2. 5 Area Specific Justification (Jiwari)

There are composition rules for making good appearance for tabular matters. One typical rule is applied in the treatment of headers or titles of tabular matters. The length of header are adjusted to match the whole length of tabular matters. For example, one space is set between each of four character header to be set in a seven character space. This method is often applied to directory entries for making a balanced appearance of surnames and given names that are of various length.

On the contrary to the above case, when the string length is relatively too short to the space of a header, the allocation of characters in a header may vary. There is no particular rule for this case, and treated according to aesthetic sense of editor or printer.

梟 <small>司馬遼太郎著</small> の 城 ¥260 152A	薄 <small>五味康祐著</small> 桜 記 ¥260 151E	柳 <small>五味康祐著</small> 生 武 芸 帳 (全三冊) 上 ¥220 中 ¥220 下 ¥240 151B~D	秘剣・柳生連也斎 <small>五味康祐著</small> ¥180 151A	運 <small>柴田三郎著</small> 命 (全三冊) 前 ¥300 後 ¥280 150K, S	赤 <small>柴田三郎著</small> い 影 法 師 ¥200 150Q	孤 <small>柴田三郎著</small> 剣 は 折 れ ず ¥320 150P
--	---	--	--	--	---	--

その他 (1回1.0%)

120,000~121,000

回数	差引残高	クレジット 手数料	お支払高 お残	返済金		運営費
				第1回目	第2回目以後	
2	120,000	2,400	122,400	61,200	61,200 × 1	3,600
3	—	3,600	123,600	41,200	41,200 × 2	
4	—	4,800	128,400	31,200	31,200 × 3	

Area Specific Justification

1. 2. 6 Mixed Type Setting of Latin and Japanese Characters (Wa-Ou Konshoku)

Latin alphabets are quite occasionally used in Japanese text. It is natural to use Latin characters in textbook of English language, of course. They appear frequently in scholarly papers. Arabic numerals are treated similar to Latin alphabets.

Normally, there is no space in Japanese text, however, space of a quarter em (a unit of length for printing) is set between Japanese string and Latin string. Naturally, Latin alphabets strings such as English text are set according to the composition rules of English text.

One of specific characteristics of Japanese font is that they are of same width, or same height. When Latin characters are used in

Japanese character strings, particular categories of Latin fonts are chosen such as wide-width, similar heights between upper and lower characters, fonts that have similar curves to Japanese characters, etc.

It is elegant to align the centers of Latin alphabets and Japanese Kanji characters in one line. However, there are Latin alphabets that extend their shape below the line or above the line such as "y", "j", "k", "h", etc. When the two centers are matched to the alignment line, many of Latin alphabets go below the base-line of the string, so that the base line of the Latin character string is shifted slightly upward. There are Latin character fonts designed to match the center or the base line of Japanese characters.

オートクチュール haute couture 高級衣装店。本来は高級裁縫の意。創始は一般にナポレオン3世紀ウジェニーの専属ドレスメーカー、ウォルトとされる。1868年に創立され1911年に改組されたパリ高級衣装店組合事務局 La chambre syndicale de la couture parisienne の規約にみられるオートクチュールの条件は、(1)年2回、一定期間内に生きたマヌカンを使って60点以上の作品を発表すること、(2)その店の創作衣装を顧客に売ることなどであり、43年には法的にその著作権が認められた。現在組合加盟店は30余で、最近はこれまでの高級注文服に比し、プレタ・ポルテ（一種の既製服）の比重が増してきている。  
 オートクラシー autocracy 専制政治。国家主権が民意や法律上の制約を受けずに運用される政治。立憲制と区別され、近世初期の君主政治にみられた。

## 社会病理

Social pathology 一般には、犯罪・非行、売春、自殺、スラム、貧困など、いわゆる異常現象をさすと考えられることが多いが、これらは文字どおり現象であって、社会病理の一面を示すにすぎない。社会病理を十分に理解するためには、単にこれら個々の現象を考察するのみならず、現象の基底にあるものにまで注目することが必要である。

すなわち社会病理とは、個人、集団、地域社会、そして全体社会が、その内部的、外部的な条件によって、生活機能に障害（生活障害 social dysfunctioning）を生じ、その結果さまざまな逸脱行動 deviant behavior や、逸脱現象 deviant phenomena を生ずる過程である。なお、個人、集団、地域社会、全体社会を規制しているそれぞれの文化、つまり価値体系や慣習などの異常 abnormality、逸脱 deviance も、社会病理の一面としてとらえることができる。

### Mixed Type Setting of Latin and Japanese Characters

## ● 海と島の博覧会・ひろしま

広島市の市制100周年・築城400年を記念した「海と島の博覧会」は「海と島のグランドデザイナー輝く海と島と人をテーマに開催中である。これは瀬戸内海に面する広島が、海と島すなわち21世紀に向けて、新しい瀬戸内海の多様で豊かな可能性と人間との関わりの中で、世界に向けて大いなるメッセージを発信することを基本理念としている。会場はメイン会場のほかに、主催、共催、個別合わせて、瀬戸内海の島や

### Horizontal-in-Vertical

#### 1. 2. 7 Horizontal-in-Vertical (Tate-Chu-Yoko)

It becomes complicated when a Latin character string appears in a vertical composition. Sentences of Latin characters are left as they are in the vertical string, but symbols in Latin character such as "AM" for time and short character string like this are rotated as done for Kanji string of horizontal composition. In this

case, Arabic numerals in a half size may look so ugly that they are treated as one Kanji character, or changed to Chinese numerals. Two Arabic numerals are set into the space of one Kanji character. The treatment is followed up to three Arabic numerals, but a four Arabic numerals is too long that can not be placed according to this technique.

今、若い人の間ニラッピングがさうつとしたブームですが、このことは同時にギブットの習慣が増えているところを示しています。↑  
若い人は感受性が豊かで、温かさを持った、モノを贈る行為にも大人の感覚とは違つたものを持っているのですが、美の観点から見ると

中野はよく「競輪は楽しくやればいい」と言う。確かに練習は苦しくつらい。それは自分がいちばんよく分かっている。しかし、練習をやらなければ勝てないと知ったときに、それをやるかやれないかは、それぞれの生きている世界によつて差が出るというのだ。↑  
競輪の世界では、勝つことによつて賞金を手にし、自分の生活を豊かにすることができる。中野にとつてはこれ

保障に加えて年金も受け取れる仕組み。生保各社が、今後の有力商品として力を入れていく「保険・年金組み合わせ型」と真っ向から競合するため、生保業界は「言葉の民業圧迫」と警戒を強めている。↑

Line Holding Rules

1. 2. 8 Line Holding Rules (Kin-soku)

There are many line holding rules as they exist in the Latin script composition. Characters that can not be placed in the top of line are the repeater marks, punctuation marks, closing bracket, double consonants, contracted sounds and symbols in general. Characters that can not be place in the end of line are the ending bracket, right arrow, etc. To conform to the line holding rules for the top of the line, the character concerned should be left to the preceding line or put forward to the next line. The converse processes take place for applying the rules for the end of line.

The pitch of character is equal in Japanese character strings, and the length of lines are also the same. However, when the line holding process is taken place, the rule is applied. In this case, the number of characters in a line is changed to keep the length of one line as that of other lines. In other words, spaces are allocated between characters to accommodate the left-off or kept-in characters. Adjustment may applied before or after symbols if any in a line, or if

there is no symbols in a line, pitch of all the characters are affected. Punctuation marks are hung to the last line, and never appear at the top of the line. There is no authoritative rules to cover the line holding rules in Japan, and editors and printers apply their local conventions.

1. 2. 9 Needs for Style Manuals and Standards for Authors

Some conventions and rules are introduced above for the case of Japanese printing composition. There must be similar conventions in Asian languages, and of course, there are totally different conventions and rules in each countries. In the U.S.A., there is "Chicago Manual of Style" to control authors, editors, and printers for English text. It is desirable to have this kind of manual for processing the full text to print, to make them into databases. Each printer have their own style, but their style is not yet described as commonly applicable rules or conventions.

Composition rules have been established to



	Databases Accessible in Japan							
	1982	1983	1984	1985	1986	1987	1988	1989
Registered DBs	604	916	1,242	1,702	1,959	2,440	2,858	3,096
Unique DBs	456	679	924	1,289	1,483	1,795	1,964	2,128
Regstrd Enterprises	42	58	78	100	121	137	194	211

	Number of Unique Databases in Online Search Services in Japan							
	1982	1983	1984	1985	1986	1987	1988	1989
Foreign Database	334	522	725	1,008	1,187	1,370	1,436	1,466
Japanese Database	122	157	199	281	296	425	528	662
TOTAL	456	679	924	1,289	1,483	1,795	1,964	2,128

make the text readable for ordinary people, that are cumulated in a long time of printing histories with or without logical backgrounds, or according to cultural customs, or depending on aesthetic sense. In other words, they represent the printing culture or the culture itself, and changing themselves as the culture evolves. For example, the printing composition rules have changed drastically when the CTS was introduced to the era of the manual movable types. It seems still changing in the age of computer. It is inevitably necessary to establish a common composition rules in a universal framework with respects to each language, or to establish a common conversion procedure between different composition rules for universal access of full text data resource.

### 1. 3 Database Production in Japan

#### 1. 3. 1 MITI Database Directory

MITI, the Japanese Ministry of International Trade and Industries, started to compile the "Database Directory" in 1982. In the 1990 annual volume which includes data for the year of 1989, there is a total of 3,096 databases registered with duplications, which are served by 211 organizations, most of them are private companies.

"Database White Paper, 1991", published by DPC in March 1991, reported the number of

registered databases, unique databases available through online search services and the number of registered IR service enterprises in Japan.

Average annual increase of number of Japanese databases is 127% in the past seven years for 1982 and 1988. Database provision is reported in the Directory, however, the usages of online IR services are not included.

#### 1. 3. 2 Databases Produced within Universities

An annual survey has been conducted by the NACSIS since 1987 on database construction within university community in Japan. In Summer (June-July) 1990, 1990 survey was carried out and found 1,073 databases.

Nearly a half (45.9%) of the 1,073 databases being produced in June-July 1990 is textual database. However, there has been no statistics yet of the data volume measured by the number of records, the length of records or number of bytes per record, etc. Of the 25% databases available is of numeric, and nearly 151 databases (14%) include images and graphics. The number of full text databases being produced is only 78 (7.2%). All the databases are being "publicly" accessed within the university community. Majority of textual data in campus, such as manuscripts for journals are not yet offered for public service.

	January 1989		September 1989		June 1990	
	Institutions	DB	Institutions	DB	Institutions	DB
National university	95	549	96	601	95	621
Municipal university	38	17	38	28	37	28
Private university	358	162	358	226	365	256
Inter-university research institutes	14	6	15	94	15	108
Other institutions	14	54	11	53	14	54
					537	6
<b>TOTAL</b>	<b>518</b>	<b>848</b>	<b>521</b>	<b>1,002</b>	<b>1,063</b>	<b>1,073</b>

<b>Textual Databases</b>		
Full text		63
Full text and bibliographic data		15
Bibliography		224
Abstracts		72
Dictionary, directory		34
other		83
Subtotal (Textual Databases)		491
<b>Combination of Textual and Other forms</b>		
with Numeric data		142
with Numeric and Images		53
with Images or Graphics		27
other		2
Subtotal (Combination of Textual and Other forms)		224
<b>Numeric Databases</b>		264
Numeric & Images		55
Images, Graphics, Voices, and/or Program		16
Other		14
na		9
<b>TOTAL</b>		<b>1,073</b>

### 1. 3. 3 Bibliographic Databases

Although the main stream of information search services is on the financial and business data such as the stock market information, investment, etc., bibliographic information has been one of basic core in database business. For example, most of the JOIS databases (other than JOIS-F), the leading Japanese IR service now provided world-wide, are bibliographic or abstract databases.

The Japan MARC, produced by the National Diet Library (NDL), Tokyo, is a machine-readable national bibliography, and the Japanese equivalent to the Korean MARC or KOR-MARC which is produced by the National Central Library in Seoul. It includes more than 60,000 monographs each year and there are

more than 40 subscribers of the Japan MARC tapes who utilize the bibliographic data either for their own sake, for cooperative purposes, or for redistribution. It is the genuine core of national bibliographic control. Since 1989, NDL started the distribution of the Japan MARC serials.

J-BISC, a CD-ROM version of the Japan MARC, is sold by the Japan Library Association, and the number of subscribers exceeded more than 500 in 1990. A large part of the subscribers are university libraries. It is expected that the public libraries will subscribe J-BISC for their local cataloging task without the network facilities or a substantial computer facilities.

### J-BISC Distribution

University Libraries	44%
College Libraries	12%
High/Secondary School Libraries	4%
Public Libraries	17%
Corporate Libraries	17%
Others	6%

### NACSIS-CAT Union Catalog Database (as of May 1991)

		Bibliography	Holdings
Books	Japanese	506,000	2,850,000
	Foreign	1,005,000	2,085,000
Serials	Japanese	66,000	1,437,000
	Foreign	109,000	886,000
Author Authority			466,000
Title Change Map	Japanese		5,876
Title Change Map	Foreign		12,801

NACSIS-CAT, a union catalog database of holdings of 140 university in Japan, records more than 3.8 million volumes or 1.3 million unique titles of monographs. The Volume of serials record in the union catalog are 160,000 titles and about 2 million holding records.

#### 1. 3. 4 Electronic Manuscripts

Electronic Campus, Electronic Library, Electronic Book, Electronic Journal, Electronic Publishing, all starts from electronic manuscripts. Electronic manuscripts are produced by numerous authors with individual purposes at diverse hardware/software environment with, at present, little exchangeability.

Word processing machines as well as word processing software on personal computers in offices, campuses, and at home have been producing a mass of electronic manuscripts most of which, at present, complete their once-only-mission to print on paper. However, conversion software for text files encoded by *dumb* word processing machines is a strong under current in the PC market, that reflects the demands for exchanging word processing

manuscripts including graphics in electronic form. Also, up till now, the editing information for page layouts, sizes of characters, etc. in a word processor file are not transferred to other word processing machines, so that the receivers have to re-edit the manuscripts.

Book publishers, journal publishers started to accept manuscripts in a machine-readable form such as disket/floppy disk which are transformed by a conversion mechanism to computer type setting data. Small printing firms started to follow the suite. Re-editing of electronic manuscripts require not only labor-intensive but also highly intellectual manipulation to reproduce the text as the author expects.

Technological incompatibilities of character sets and graphic information, both physical and logical file structures, layout and style information prohibit the free-flow of the electronic manuscripts even within the same language as well as in a single organization. Standards related to flexible disks have been established for physical and logical specification.

Lack of social conventions also prohibits the exchanging information as a commercial products, as an academic products, as a commonly shared products for public domain. The traditional link are to be reformed by the introduction of electronic manuscripts among authors-editors-publishers-book traders and bookshops-readers and libraries as well as information distributors.

A process from writing of manuscripts till publishing is not a single track. For example, after the reading by referees, authors have to re-write the text, re-illustrate graphs and tables. Designing of layout may be carried out by a professional editor and/or author him/herself.

Here is an important area to be standardized or set guidelines for preparation and/or encoding data and their exchange.

## 2 ORGANIZATIONS PROMOTING ELECTRONIC PUBLISHING

### 2.1 JEP A

JEP A (Japan Electronic Publishing Association)<sup>4</sup>, founded in September 1986, has promoted CD-ROM publishing. Its mission is to *resolve the problems faced by those seeking to plan, edit and create electronic publishing, electronic publications and all kinds of information media. Its goal is to contribute to the healthy development of the publishing and information industries.*

JEP A is composed of corporate membership, and its membership has increased from 43 at the inception in 1986, 55 in the end of 1986, 105 in 1987, 134 in 1988, 163 in 1989, and 171 in September 1990. Members are publishers,

printing companies, newspapers, computer hardware manufacturers, software houses, and some abstracting and indexing businesses.

As early as 1987, JEP A started to develop a standard logic format of CD-ROM for Japanese language that is compatible to ISO 9660:1988. Then, in 1989, a test disk was developed with the collaboration of 22 member companies and seven CD-ROM press factories. JEP A determined a standard for multi-media data format in 1989, and a standard for CD-ROM accommodating documents (texts), voice, images in 1990 under a grant awarded by the Database Promotion Center (DPC).

The first standard CD-ROM, named "Wado Kaiho" after the very first coin used in Japan, has the following four features:

- 1 adaptation of ISO 9660:1988 for Japanese language and Japanese language can be used for file names and the directory. Thus a multi-lingual file structure is achieved;
- 2 the test disk can be driven on a variety of personal computers such as NEC/PC98 of NEC, Fujitsu/FMR, IBM/AX, and Hitachi/B16, etc. to achieve a wider market;
- 3 Joint publishing of 22 software (data) is made into the test disk. "Comprehensive Database Directory of the Ministry of International Trade and Industry" is one of the test data. Joint publishing may reduce the cost of publishing.
- 4 A registration law for CD-ROM copyright, equivalent to the title page, the contents, and the colophons, is being proposed;

Activities of JEP A in the first four years concentrated on the development of CD-ROM of textual data. As shown above, JEP A aims at other forms of data such as voices, images and will extend its scope towards multi-media electronic publishing. It is important to secure the way of distribution, and JEP A is apparently in a better position in this aspect, since major publishers are members who are part of the traditional distribution system of printed media.

---

4 JEP A, Shuppan Kenkyu Senta, 2-23 Kanda Jinbo-cho, Chiyoda-ku, Tokyo 101

## 2. 2 DPC

Database Promotion Center, Japan (DPC)<sup>5</sup> was established in 1984. with 126 corporate membership. Organizational characteristics of DPC may be of government promotion with strong support of the commercial sector. Its activities for promoting database production and utilization, backed by the Government, are of four areas such as:

- 1 Financial assistance for database production to help offset the high costs of production of, and long period required to realize return on investment in databases deemed essential for the national economy;
- 2 Publication of "Database White Papers" since 1986 and "Database Directory" in printed version since 1983 as well as in a floppy disc version since 1987.
- 3 International activities such as holding conferences; overseas missions; participating in organizing conferences.
- 4 Survey, investigation and publication of manuals. Recommendations and reports have been made by the committees such as "Opening of governmental data to the private sector", "Promotion of informatization of rural areas", "Advanced utilization and the promotion of distribution of statistical data", "Establishment of synthetic networks for databases", "Charge system and regulations for utilization of databases", etc.

Financial supports by DPC with two main purposes a) to promote the construction of databases with social, economical, international importance; b) to activate R&D of database related activities. Publishing activity of DPC, featured by "Database White Paper" and "Database Directory", is becoming popular and playing the leading role of business promotion. "Database in Japan 19XX" is an English summary of the annual white paper. In 1990 edition, five problems confronting the Japanese database industry are pointed out:

- 1 Low independence as a business with low profit;
- 2 Structural immaturities of the industry;
- 3 Intensification and diversification of information contents;
- 4 technological gaps with the latest developments;
- 5 Lack of database specialists

---

5 DPC, World Trade Center Building 7F, 2-4-1, Hamamatsu-cho, Minato-ku, Tokyo 105. This section is based on "Outline of Database Promotion Center, Japan" 1988. 8p. in English.

6 DINA, 2-7, Kagurazaka, Shinjuku-ku, Tokyo 162 This section is based on "Introduction of DINA : Japan Database Industry association" [1980]. 4p. in Japanese.

7 SGML Forum, c/o INSTAC, 4-1-24 Akasaka, Minato-ku, Tokyo 105

## 2. 3 DINA

Japan Database Industry Association (DINA)<sup>6</sup> was founded in 1979 by 19 commercial companies with the aims of establishing the database industry in Japan and contributing to business of members. DINA plays its role as the voice of the industry. By the end of 1989 there are 109 corporate members. Activities are defined by its rule:

- 1 promotion of database;
- 2 investigation on promotion of the database service industry;
- 3 recommendations to government administration policies on the database service industry;
- 4 seminars for the database service industry;
- 5 international exchanges on the database service industry;
- 6 establishment of codes of business for the database service industry; etc.

DINA runs more than ten committees such as "media committee", "PC communication committee", "database regulations committee", etc. An activity of DINA that attracts the members of the Association and other professional people is a series of seminars such as [DINA Database Utilization Seminar].

Database Fair was initiated in 1989, by the joint organization of DPC and DINA. *Database '90*, the second in the series, attracted 68 exhibitors, mostly Japanese vendors but some foreign, and about 30,000 audience for three days.

## 2. 4 SGML-Forum

The SGML Forum<sup>7</sup> was created in May 1989 at the INSTAC with support of MITI. Members of the Forum are mostly from mainframe computer manufacturers, software houses, and printing firms. The majority of publishing firms are waiting to the coming period of

fruitful standardization of SGML.

The Forum is composed of a Planning Working Group, a Technical Working Group, and an Application Development Working Group. The Forum holds an open seminars twice a year in Spring and Autumn, and attracted many audiences in 1989/90.

## 2. 5 ELICON

The Electronic Library Consortium (ELICON), Japan<sup>8</sup> was established in early 1991 with strong support by every type of libraries in Japan which anticipated the coming of the age of electronic publishing. Its activities, although very young, are to cover:

- 1) lecture seminar, training on electronic library, electronic information media;
- 2) research on electronic library, databases, electronic media;
- 3) publishing;
- 4) international exchange;
- 5) consultation; exhibitions.

SIGs are being organized for applications of electronic publishing among museums, art museum, libraries; multimedia; intellectual property; and electronic library.

## 2. 6 JICST

The Japan Information Center of Science and Technology (JICST)<sup>9</sup> is founded in 1957 by the Japanese Government (Science and Technology Agency) under a law. Its mission is to collect, process and store national and international information in the field of science and technology for open public access. With ten branches, two deposit centers, and two overseas offices in Paris and Washington, D.C., JICST provides a comprehensive information services for science and technology.

Databases are produced by JICST and there are a few non-bibliographic databases in JOIS-F services:

DNA Databases;  
JICST/NRIM Materials Strength Database for Engineering Steels and Alloys;  
Chemical Dictionary;  
Thermophysical and Thermochemical Property Database;  
Mass Spectral Database  
Crystal Structure Database;  
Chemical Substance Safety Regulation Database.

The last database, Safety Regulation, is a full text database and JICST will extend its service toward full text database with its English-Japanese machine-translation facilities. JICST has lead a development and implementation of machine-translation for its vast amount of databases for overseas access.

## 2. 7 NACSIS

The National Center for Science Information System is one of the National Inter-University Research Institutes, whose purpose is "to gather, organize and provide scholarly information, as well as to carry out research and development (R&D) of scholarly information and a science information system (SIS)," as defined by Article 7 of the National University Chartering Law.

At NACSIS, *Scientific Papers*, a conventional full text database of professional journals in Japan are now in service with the expected annual increase of 300 articles. At present, images such as tables and graphs are stored as printed in optical disk storage, and retrieved by reference code embedded in the database, and send out by conventional telefacsimile (G III).

NACSIS is experimenting to apply SGML in collaboration with academic societies:

<sup>8</sup>ELICON, 3-1-401, Higashiyama, Meguro-ku, Tokyo 153

<sup>9</sup>JICST, 2-5-2, Nagata-cho, Chiyoda-ku, Tokyo 100, Japan

- 1) for inputting (writing) articles into a machine-readable form through word processing machine (WPs or PCs) by authors;
- 2) for editing these articles into a bulletin of an academic society;
- 3) for publishing the bulletin in a traditional printed form as it is or in a form of CD-ROM for local access;
- 4) for loading the database to provide online access.

### 3 ACTIONS PROMOTING ELECTRONIC PUBLISHING

#### 3.1 CD-ROMs

CD-ROM, Compact Disk Read Only Memory, is a new product in the database industry. It includes 540 Mega bytes, or 5.4 million characters in one byte code, or 280,000 pages of 2,000 words/A4 pages in English.

CD-ROM as a database has specialities. It is handled independent to the central computer, network, and other resources required to online network search. Update cycle is not hours or days that are common to online database but normally months. It can be down-loaded in contrast to the usual limitation to the online usage. CD-ROMs are handled on a personal computer with accompanied software dependent to a particular personal computer. Therefore, users have to handle it on a limited options, and

when there is no matching personal computer to a CD-ROM, which becomes only a plastic disk.

*CD-ROM Directory of the World (Sekai CD-ROM Souran)* is compiled in Japanese language by Kyodo Keikaku Shuppan, Co. Ltd<sup>10</sup>. and distributed by Kinokuniya since 1988. The latest edition, Vol.4 for 1990, includes 567 entries of CD-ROMs. The Directory estimates the world products of CD-ROMs in 1990 as follows:

Mr. MURAMATSU Hiroshi, the editor of the Directory indicates that the CD-ROM marketing is shifting from the Selected Few, in which the volume of sale can be estimated but small sale, to the Mass Market with emphasis towards entertainment in which the volume of sale is unpredictable but in the order of more than 10,000. CD-ROM hardware manufacturers expect the growth of the market as a part of home use / family use business.

CD-ROM developments in Japan is not slow, and JEPA has been working on developing CD-ROM standards based on the ISO standards. There are examples to show the variety from bibliographic to dictionary database:

	Number of Entries in <i>Sekai CD-ROM Souran</i>				Total	New Title
	Japanese CD-ROM	New Title	Foreign CD-ROM	New Title		
1988 Vol.1	55	(22)	154	(11)	209	(23)
1989 Vol.2	77	(26)	211	(7)	288	(33)
1990 Vol.3	173	(25)	257	(5)	430	(30)
1990 Vol.4	273	(9)	294	(-)	567	(9)

#### World Products of CD-ROMs in 1990 (*Sekai CD-ROM Souran*)

Japan	300	18%
USA	1,200	70%
Europe	150	9%
Other	50	3%
Total	1,700	100%

<sup>10</sup> Kyodo Keikaku Shuppan. Futaba-Building #202, 3-4-18 Mita, Minato-ku, Tokyo 108

"Koujien" is a dictionary of Japanese language published by Iwanami Shoten. The CD-ROM version covers the 3rd edition of the Koujien and 200,000 entries of Japanese words can be searched. Figures, images are also included. "CD-WORD 8" is a dictionary of seven languages (Japanese, English, French, German, Spanish, Italian, and Dutch) published by the Sansyu-sha.

NACSIS also produced a union list of periodicals in a CD-ROM that includes 40,000 titles of journals in Japanese languages held by university libraries of about 633 locations. The number of holdings is one million. The date of the coverage is as of October 1985. Another CD-ROM of the union list of periodicals for Western languages is under preparation to be published in 1990.

At the moment, CD-ROMs produced in Japan conform to JIS character sets. That means they can be accessible only by personal computers that apply JISs. Standardization for hardware and software is required for user to gain flexibility on personal computer and CD drivers to use. And establishment of standard search procedures for beginners or professional users will be necessary. Durability of CD is also another aspect for user to look at.

### 3. 2 Full Text Databases

Full text database or electronic manuscripts is the next generation of form of database, and has been a focus of database research and industries. Examples are : legal information; newspaper articles; journal articles, full text of literary works, etc. Legal information seems to have established one of the major areas of information market in the USA

Potentials of full text database are: quicker input by authors; complete set of information;

direct source for printing and publishing; direct source for online access; etc.

Methods of updating, retrieval, and other aspects of processing full text databases are comparatively different from bibliographic databases. For example, newspaper full text is required to update quite in a short time, and has little or no validation. In contrast to the newspaper, input of monographic full text database is almost once only and supposed to be complete so that validation takes laborious checking.

The most strong feature of the full text databases is to provide users the complete information. Conversely, bibliographic databases only provide a guide to the original documents even the record includes an abstract and require the collection of original materials such as libraries or data archives. Access methods to full text databases are provided either online or offline such as CD-ROM. Here is an enhanced options for users to access the complete information.

A variety of full text databases are already accessible online in Japan through JOIS, NACSIS-IR, NIKKEI, etc. Most of IR software currently in operation are based on functions for retrieving bibliographic information so that it is expected to develop a new generation IR software that accommodate fully the demands of searchers.

Full text applications of patent information in Japan, a topic that will be discussed elsewhere, has been intensive with variety of mechanism and products.

## 4 CONCLUDING REMARKS

Electronic publishing, with variety of products



and potentiality, is being developed in Japan such as CD-ROMs, Full text database, and Printing on paper, etc. Developments with substantial investment will be needed in both private sector and public sector, since academic publishing is a part of frontier in development. Developments needed for promoting the electronic publishing are diversified, as listed below, from technical, social and international for achieving national and international compatibility for achieving the equality of access of information.

**1) Technical developments (standardization)**

Input/Data capturing  
Terminal/Workstation  
Editing/Markup  
Version control  
Data quality control/Authority control  
Delivery mechanisms  
Resolution (reproduction) quality  
Retrieval software

**2) Social developments**

Copyright  
Conventions between  
Author/Editor/Printer/Publisher/Distributor/Reader  
Charging mechanism and conventions  
Training of editors and/or authors  
Style manual

**3) International developments (both technical and social)**

Character sets  
Common command language for full text, images  
Style manual

**Acknowledgement**

Thanks are due to the KIET and FID/CAO for attending this good opportunity to exchange information and opinion with colleagues in Korea. Thanks are also due to Dr. KOMACHI Yushi of Matsushita Graphic Communication Systems, Inc. and TANAKA Yoichi of the Toppan Printing Co. Ltd. Thanks are also due to Mr. Lee Chang Kyo, Ms. Kim Sukyong and Mr. Chun Hak Byoung of KIET.