



STANDARD ST.40

RECOMMENDATION CONCERNING MAKING FACSIMILE IMAGES OF PATENT DOCUMENTS AVAILABLE ON CD-ROM

INTRODUCTION

1. This Recommendation provides for making patent documents available as facsimile images on a Compact Disc Read-Only Memory (CD-ROM), particularly for exchange purposes between patent offices to assist the future use of CD-ROM discs in place of paper. It is intended that discs published in compliance with this Recommendation will function, without modification, under a variety of operating systems.

2. For the purposes of this Recommendation, the expression "patent documents" includes patents for invention, plant patents, inventors' certificates, design patents, utility certificates, utility models, documents of addition thereto and published applications therefor.

3. This Recommendation gives guidance to offices in their production of facsimile CD-ROM discs. The current or proposed practices of offices in their production of facsimile CD-ROM discs are analyzed in respect of that guidance.

REFERENCES

4. The following references are of importance to this Recommendation:
- CCITT Group IV:1984 T.6, Facsimile Coding Schemes and Control Functions for Group 4 Facsimile Apparatus.
 - ISO 639:1988, Code for the Representation of Names of Languages.
 - ISO 646:1991, Information Technology – ISO 7-bit Coded Character Sets for Information Exchange.
 - ISO 2022:1986, Information Processing – ISO 7-bit and 8-bit-Coded Character Sets – Coded Extension Techniques.
 - ISO 2375:1985, Data processing – Procedure for Registration of Escape Sequences.
 - ISO 8601:1988, Data Elements and Interchange Formats – Information Interchange – Representation of Dates and Times.
 - ISO 8879:1986, Information processing – Text and Office Systems – Standard Generalized Markup Language (SGML).
 - ISO 9660:1988, Information Processing – Volume and File Structure of CD-ROM for Information Interchange.
 - ISO/IEC 10149:1989, Information Technology – Data Interchange on Read-Only 120 mm Optical Data Disks (CD-ROM).
 - WIPO [ST.3](#), Recommended Standard Two-Letter Code for the Representation of Countries, and of Other Entities and International Organizations Issuing or Registering Industrial Property Titles.
 - WIPO [ST.9](#), Recommendation Concerning Bibliographic Data on and Relating to Patent Documents.
 - WIPO [ST.16](#), Standard Code for Identification of Different Kinds of Patent Documents.
 - WIPO [ST.30](#), Recommendation Concerning a Standard Magnetic Tape Format for the Exchange in Machine-Readable Form of Bibliographic Data, Abstracts and Full Texts of Patent Documents.
 - WIPO [ST.31](#), Recommended Coded Character Sets for the Exchange of Machine-readable Records of Patent Documents.
 - WIPO [ST.32](#), Recommended Standard Format for the Generic Coding of the Text of Patent Documents Exchanged on a Machine-Readable Carrier.



HANDBOOK ON INDUSTRIAL PROPERTY INFORMATION AND DOCUMENTATION

Ref.: Standards – ST.40

page: 3.40.2

- WIPO [ST.33](#), Recommended Standard Format for Data Exchange of Facsimile Information of Patent Documents.
- TIFF, Tagged Image File Format, Version 4.2 (from MS Windows Development Kit, Version 2.0).

ANNEXES

5. The annexes to this Recommendation contain the following:

- (i) Technical information specific to CD-ROM:
 - Annex A CD-ROM Structure
 - Annex B Technical Recommendations
- (ii) The current or proposed practices of various offices involved in the production of facsimile CD-ROM products:
 - Annex C Current or Proposed Practices in the Coding of Character Sets
 - Annex D Current or Proposed Practices in Respect of Directory Structure and Documentary Files
 - Annex E Current or Proposed Practices in Respect of Image File Formats
 - Annex F Current or Proposed Practices in Respect of Indexes
 - Annex G Current or Proposed Practices in Respect of Client/Server Models
- (iii) Guidance in the selection and naming of bibliographic data for inclusion on a disc:
 - Annex H Bibliographic Data: Field Names and Field Formats

CD-ROM STRUCTURE

6. The physical dimensions and recording structure of a CD-ROM shall comply fully with ISO/IEC Standard 10149:1989. Details of the physical layout of data on a CD-ROM disc are given for purposes of information in Annex A.

7. The volume and file structure of a CD-ROM shall comply with International Standard ISO 9660:1988. The interchange level shall be set to 1 (as defined in International Standard ISO 9660:1988, paragraph 10.1).

8. Certain specific technical recommendations, particularly providing an implementation of International Standard ISO 9660:1988 for the purposes of this Recommendation, are given in Annex B.

CHARACTER SETS

9. The coding of characters in *descriptor* records, indexes and file tags shall be as specified in International Standard ISO 9660:1988, viz., according to International Standard ISO 646:1991. The coding of character sets used in *data* records will need to take account of the language of the patent document recorded on the CD-ROM. The existing practices in respect of coding of characters, used in the implementation of current or proposed products, are recorded in Annex C.

DISC IDENTIFIER

10. Each CD-ROM shall be given a unique identification by the publisher having the following three parts:

- | | |
|-------|--|
| Title | E.g., PatFacs, U.S. Patent Facsimile Images of the Week, up to a maximum of eight characters |
| Year | Four-digit calendar year of issue (the bibliographic volume number) |
| Issue | The sequence number of the issue within the year, first issue numbered 1, up to a maximum of four characters |

Examples: PatFacs 1921 4
PatFacs 1991 65
PatFacs 2021 234.



HANDBOOK ON INDUSTRIAL PROPERTY INFORMATION AND DOCUMENTATION

Ref.: Standards – ST.40

page: 3.40.3

DIRECTORY STRUCTURE AND DOCUMENTARY FILES

11. A directory name shall consist of no more than eight characters and shall not contain an extension.
12. All patent images shall be recorded in subdirectories subordinate to a single subdirectory which is subordinate to the root directory. The name of the subdirectory at the top of the image subdirectory hierarchy shall be IMAGES.
13. Any software provided by the publisher on the CD-ROM disc for viewing, printing, or searching images or text shall be recorded in subdirectories subordinate to a single subdirectory which is subordinate to the root directory. The name of the subdirectory at the top of the software subdirectory hierarchy shall be SOFTWARE.
14. Where a facsimile CD-ROM disc contains additional bibliographic or full text databases, an additional subdirectory named OTHERAPL shall be present in the root directory and shall contain dependent application subdirectories.
15. Files shall be available in the root directory of the CD-ROM to include some or all of the following information:
 - (a) The full names of owners of proprietary rights
 - (b) Disc identification information
 - (c) A brief description of the contents of the CD-ROM, e.g., the text of the disc label (see paragraphs 44 and 45, below)
 - (d) The file format of the image data on the CD-ROM
 - (e) Issuing office(s)
 - (f) Office or agent preparing the data
 - (g) Software identifier: identifying the name and version of the software initially intended for use with the CD-ROM.
16. Current and proposed practices with respect to directories and documentary files are recorded in Annex D.

STORAGE AND FILE FORMAT

17. Each patent document shall be recorded as raster images (facsimile mode), using CCITT Group 4 T.6 compression.
18. The images of the patent document may be stored in one of three file formats:
 - 18.1 Each A4 page (frame) may be stored as a single file. The image file header will closely follow that specified in WIPO Standard [ST.33](#). Each file recorded in this way should have a file name extension, e.g., S33. For pages larger than A4, more than one frame, each corresponding to an A4-sized area of the page, should be created so that the total area of the page is included. An additional frame may be created which corresponds to a physical reduction of a page larger than A4-sized to a size that can be accommodated within the dimensions of an A4-sized page.
 - 18.2 Each patent document may be recorded as a single file with the appropriate header and “directories” as specified in the Tagged Image File Format (TIFF) specification. Each file recorded in this way should be named with the patent document number and have a file name extension, e.g., TIF.
 - 18.3 Each page of a patent document may each be recorded as a single file with the appropriate header and “directories” as specified in the TIFF specification. Each file recorded in this way should be named with the patent document and page number and have a file name extension, e.g., TIF.
 - 18.4 Current or proposed practices in respect of image file formats are recorded in Annex E.



HANDBOOK ON INDUSTRIAL PROPERTY INFORMATION AND DOCUMENTATION

Ref.: Standards – ST.40

page: 3.40.4

INDEXES

19. An open index shall be included in the root directory, which will provide the physical location of each image stored on the disc. An open index will enable a user to gain direct access to the documents on the CD-ROM when using other retrieval software. It will provide for each patent document on the CD-ROM the following minimum information in records:

1. Two-letter office code (INID 19)
2. Kind-of-document code (INID 13)
3. Document number (INID 11)
4. Image location (absolute starting byte number)
5. Image length
6. Full path
7. File name
8. Page and frame number
9. Subdocument identification code as recommended in WIPO Standard [ST.32](#), paragraph 11.

20. Additionally, for each CD-ROM product, a document location index shall be included on each issue published, and shall give disc location information for all documents published to date in the current year, including the present issue. The following minimum information shall be recorded in the document location index:

1. Two-letter office code (INID 19)
2. Kind-of-document code (INID 13)
3. Document number (INID 11)
4. Disc identifier (see paragraph 10).

21. For each CD-ROM product, a cumulative document location index across all completed years shall be produced, with, as a minimum, the same information contained in the document location index described in paragraph 20. The cumulative document location index shall be published on a regular basis, preferably at the same time as the last issue of a year.

22. The current or proposed practice of offices with respect to indexes is recorded in Annex F.

USER REQUIREMENTS

23. This section of this Recommendation gives details of certain user requirements that have been established as desirable goals in the light of users' experience.

BIBLIOGRAPHIC DATA

General Matters

24. A facsimile CD-ROM will contain the following three distinct kinds of information:

- (a) *image data* which consists of facsimile information normally relating to complete pages of patent documents;
- (b) *bibliographic data* which consists of character-coded information which accompanies the publication of a patent document and is ordinarily given on its front page;
- (c) *index data* which consists of character-coded information organized as search files.

Minimum Items

25. For the purposes of this Recommendation, the following bibliographic data, if available in coded form, are considered as minimum items that should be given for each patent document whose images are recorded on a facsimile CD-ROM disc. INID codes of WIPO Standard [ST.9](#) are given as an aid to identifying each item.



HANDBOOK ON INDUSTRIAL PROPERTY INFORMATION AND DOCUMENTATION

Ref.: Standards – ST.40

page: 3.40.5

<i>INID Code</i>	<i>Data Item</i>
(ST.9)	
(11)	Number of the document
(13)	Kind-of-document code according to WIPO Standard ST.16
(19)	WIPO Standard ST.3 Code, identifying the office publishing the document
(21)	Number(s) assigned to the application(s)
(22)	Date(s) of filing application(s)
(31)	Number(s) assigned to priority application(s)
(32)	Date(s) of filing of priority application(s)
(33)	WIPO Standard ST.3 Code identifying the national patent office allotting the priority application number or the organization allotting the regional priority application number; for international applications filed under the PCT, the code "WO" is to be used
(40)	Date(s) of making available to the public
(51)	International Patent Classification
(54)	Title of the invention
(71), (75), (76)	Name(s) of applicant(s)
(72)	Name(s) of inventor(s)

In addition, the following bibliographic data shall also be regarded as minimum items in the circumstances that they are relevant to the office producing the CD-ROM disc:

(26)	Language in which the application is published
(52)	Domestic or national classification
(57)	Abstract
(74)	Name of attorney or agent
(81), (84)	Designated States according to the PCT, Designated Contracting States under regional patent conventions

RETRIEVAL SOFTWARE

General Matters

26. This Recommendation specifies the user requirements concerning retrieval software developed for use with CD-ROM discs containing facsimile images of patent documents and aims to facilitate the search for, display of, and printing from CD-ROM discs received from several sources. This Recommendation does not specify the use of specific software, but rather lays down certain minimum features and user requirements so as to enable an uninitiated user to become efficient in the use of CD-ROM-based patent information retrieval systems in the shortest time possible.

27. It is the responsibility of the producers of CD-ROMs to make readily available the software necessary to retrieve the information recorded thereon in a convenient and efficient manner. It is acknowledged that the making readily available of proprietary software raises problems in connection with copyright ownership and it is expected that certain restrictions on recipients of such software would be imposed in order to maintain legitimate control of copyright interests.

28. It is recommended that the retrieval software conform to the client/server model of separating the user interface from the retrieval engine. This will allow each end user to build or acquire a user interface (client) that can be used for all patent discs that comply with a particular client/server model, thus enabling consistency of access for that user without constraining retrieval engine (server) designers in any way. Annex G is reserved for future recording of the practices of disc producers who follow a client/server model and includes a more detailed description of the client/server concept.



HANDBOOK ON INDUSTRIAL PROPERTY INFORMATION AND DOCUMENTATION

Ref.: Standards – ST.40

page: 3.40.6

Presentation

29. The retrieval software shall be supplied on 130 mm (5¼ inch) or on 90 mm (3½ inch) diskettes or on the CD-ROM disc itself.

30. This Recommendation does not specify the documentation that users would require to enable them to use CD-ROM retrieval systems. It is expected that offices producing CD-ROM based retrieval systems will make detailed documentation available to users. Documentation should contain sufficient information, written in a clear style, preferably in the language of the user, and illustrated with many examples.

General Requirements

31. The retrieval software shall enable a user to perform the following functions:

- (a) search through data provided on the CD-ROM disc;
- (b) display information and images;
- (c) print selected images on a printing device;
- (d) downloading.

Each of these aspects is dealt with below.

Search Requirements

32. The retrieval software shall enable a user to search the bibliographic data present on a CD-ROM disc. The result of the search should be the display of the hit-list of documents, whose images are recorded on the CD-ROM, which respond to the search question. The retrieval software shall also enable a user to temporarily store a search question in a file and to retrieve previously stored search questions.

33. Annex H defines the input format that should be commonly available as the default, used in respect of bibliographic data items. The software should desirably accept other formats. The recommended field names given are those to be used by the user when operating the software under the so-called "expert mode."

34. In general, each search field will have an index associated with it. If such indexes are present, their contents should be viewable by the end user.

35. The retrieval software shall enable a user to search any of the above items of bibliographic data in combination using the Boolean operators AND, OR and NOT. Other search functions in common use, e.g., truncation, should be available where appropriate.

Display Requirements

36. The software shall permit the following image display options to be chosen by a user in the simplest way possible:

- (a) For a selected document number, the display of the image of the front page, or the display of the image that contains the first page of the description, or of the drawings, or of the claims, or of the search report;
- (b) For a selected document number, the bibliographic data;
- (c) For a selected document number, successive pages of the document starting with the front page;
- (d) For a number of selected documents, e.g., following a bibliographic search, the front pages of each in sequence, or the drawings of each in sequence, or the descriptions of each in sequence, or the claims of each in sequence;
- (e) For any displayed document, the user should be able to select the next page of the document, or the previous page of the document, or any page of the document.

37. For any displayed document, the software shall permit the user to enlarge the image displayed on the screen, e.g., by use of a zoom function, and to select portions of an image to be displayed by moving the viewing area up or down, or left or right, e.g., by use of a pan function.



HANDBOOK ON INDUSTRIAL PROPERTY INFORMATION AND DOCUMENTATION

Ref.: Standards – ST.40

page: 3.40.7

38. The software shall permit a user to rotate the displayed image by +90°, or by -90°, so as to provide comfortable viewing of drawings that are printed 90° to the vertical axis of the page.
39. For any displayed document the user shall be able to mark the document for subsequent display, downloading or printing of the marked document at a later time.
40. Any of the above options or selections should be capable of being stored for repeated use, e.g., recorded as a default setting.

Print Requirements

41. The software shall permit the user to specify at least the following print functions:
- (a) to print all images of one, or a selected number, of documents;
 - (b) to print those images containing the front pages of one, or a selected number, of documents;
 - (c) to print those images containing the drawings of one, or a selected number, of documents;
 - (d) to print those images containing the claims of one, or a selected number, of documents;
 - (e) to print those images containing the search reports of one, or a selected number, of documents;
 - (f) to print the bibliographic data or parts thereof of one, or a selected number, of documents;
 - (g) to print any combination of images referred to in subparagraphs (b) to (e), above.
42. The software shall give the user the option of immediate prints, or prints at a later time. The software shall also permit a user to print a copy of a document, or a page thereof, without having first to display the document on the screen, and to print on both sides of a sheet.

Download Requirements

43. The structure of a CD-ROM disc shall be such as to permit a user to download in electronic form and as a print output:
- (a) the character coded information that is displayable on a screen as the result of a search operation;
 - (b) the bibliographic data recorded on the CD-ROM disc (see paragraph 25, above);
 - (c) the compressed or decompressed facsimile image data;
 - (d) the open index (see paragraph 19, above);
 - (e) the document location index (see paragraph 20, above).

The download operation will be provided under the control of the retrieval software provided by the disc producer and, where appropriate, under the control of the operating system. In either case, the disc producer may so structure the CD-ROM disc that any of the above download functions are prevented, except to authorized users.

LABELLING

44. Each disc shall have a label permanently fixed thereto containing the following minimum information:
- (a) the disc identification code (see paragraph 10, above);
 - (b) identification of the patent documents whose images are to be found on the disc, viz., publishing office(s), kind(s) of document code(s), and document numbers.



HANDBOOK ON INDUSTRIAL PROPERTY INFORMATION AND DOCUMENTATION

Ref.: Standards – ST.40

page: 3.40.8

45. With respect to the requirements of paragraph 44(b), above:

(a) if the images relate to a continuously numbered series of patent documents, it shall be sufficient to give the first and last number of the patent documents whose images are to be found on the disc, and

(b) if the images relate to other than a continuously numbered series of patent documents, sufficient information should be given on the label so as to enable a user to determine the general contents of images recorded on the disc.

PACKAGING

46. CD-ROM discs produced for exchange purposes should be packaged in a strong, protective box, or provided in a so-called "caddy" (which facilitates the loading and unloading of a CD-ROM disc into a reader). The spine of the protective box should be labelled with the information specified in paragraphs 44 and 45.

[Annexes follow]

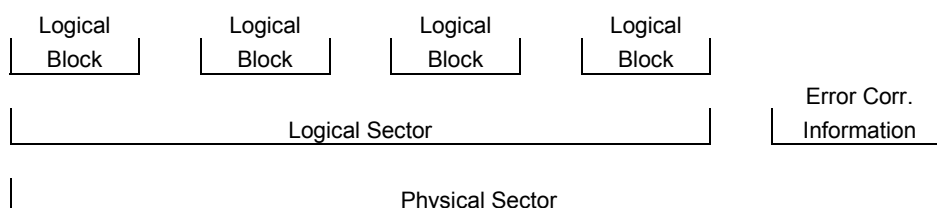


ANNEX A

CD-ROM STRUCTURE

1. A Compact Disc Read-Only Memory (CD-ROM) is a 120 mm diameter optical disc. It has a single spiral track of binary information recorded as pits on one surface of the disc. The track is read by a finely focused laser beam past which the track moves at a constant linear velocity. This requires the rotational speed of the disc to vary, the rotational speed being faster the closer the read head is to the center of the disc.

2. The smallest part of the information area on the CD-ROM that can be addressed independently is a physical sector comprising 2,352 8-bit bytes. The following diagram illustrates the relationship between the physical sector, logical sector and logical blocks:



3. Each physical sector on a CD-ROM is addressed by its playing time. Each minute on the CD-ROM is divided into sixty seconds, and each second into 75 blocks. The sector address thus takes the following form:

M:S:B where

- M = number of minutes of the sector from the start of the track;
- S = number of seconds (from 0-59) within the minute;
- B = number of the block (from 0-74) within the second.

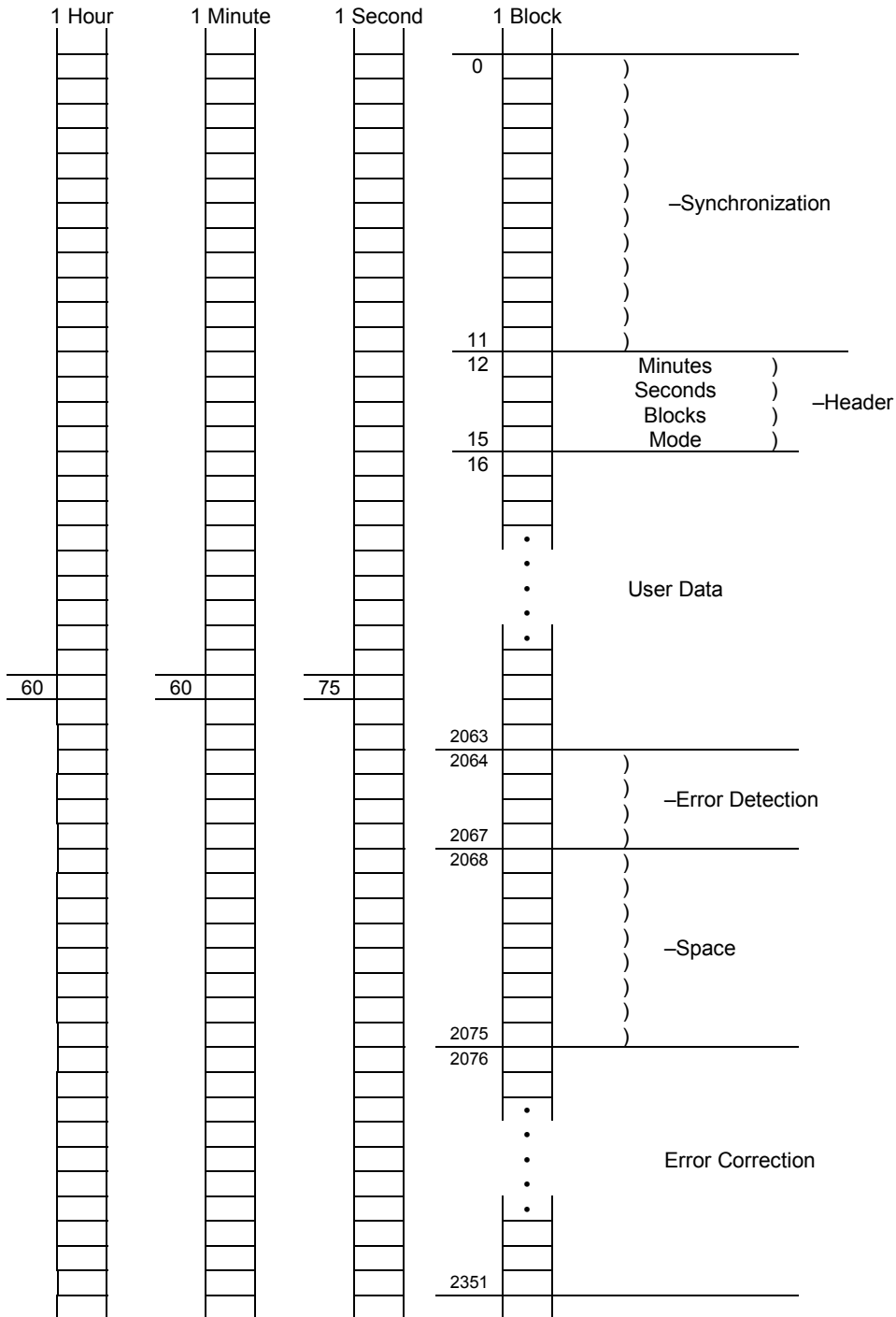
4. Each sector consists of 2,352 bytes, but only 2048 of these are available to the user, the remaining bytes being used for error detection and correction, addresses, headers and synchronization information. The physical layout of data on a CD-ROM is illustrated below.

5. The Sector Mode byte in byte position 15 of the header field of each Sector shall be set to (01) as provided for in International Standard ISO/IEC 10149:1989, paragraph 14.2 (b). This shall mean that all bytes in positions 16 to 2063 are user data bytes and that bytes in positions 2064 to 2351 are set according to paragraphs 14.3 to 14.6 of International Standard ISO/IEC 10149:1989. This requirement permits the use of byte positions 2064 to 2351 for error detection and correction and is a fundamental requirement for data CD-ROMs. Audio CD-ROMs do not have this requirement.



HANDBOOK ON INDUSTRIAL PROPERTY INFORMATION AND DOCUMENTATION

Annex A, page 2



For 1 Hour, User Data = $60 \times 60 \times 75 \times 2048 = 552,960,000$ Bytes
 For 1 Hour, Total Data = $60 \times 60 \times 75 \times 2352 = 635,040,000$ Bytes

[Annex B follows]



ANNEX B

TECHNICAL RECOMMENDATIONS

Given below are certain aspects of the technical matters dealt with in International Standard ISO 9660:1988, and should be regarded as an implementation of International Standard ISO 9660:1988 for making patent documents available as facsimile images on CD-ROM discs for exchange purposes between patent offices. Attention is particularly drawn to International Standard ISO 9660:1988 Section three: Requirements for systems. Technical matters fully defined in International Standard ISO 9660:1988 are not repeated here; this Recommendation deals only with technical matters for which International Standard ISO 9660:1988 permits user decision for implementation.

Volume Structure

System Area

1. This Recommendation does not define the contents of the system area portion of the disc (see International Standard ISO 9660:1988, paragraph 6.2.1).

Note: This Recommendation is only concerned with the data area of the disc. Producers of discs are free to use the system area (viz., the first 16 logical sectors of the disc) if they so wish.

Logical sector size

2. The preferred logical sector size (as defined in International Standard ISO 9660:1988, paragraph 6.1.2) shall be 2048 bytes.

Note: This will enable the disc to be read by most operating systems.

Logical block size

3. The preferred logical block size (as defined in International Standard ISO 9660:1988, paragraph 6.2.2) shall be 512 bytes.

Note: This will enable the disc to be read by most operating systems.

Interleave factor

4. The non-interleaved mode of recording shall be used (as defined in International Standard ISO 9660:1988, paragraph 6.4.4).

Note: This ensures the greatest compatibility of discs among various operating systems.

Associated Files

5. No data which is required to access and download patent data from a CD-ROM disc (i.e., patent data, images, indexes) shall be stored in associated files (as defined in International Standard ISO 9660:1988, paragraph 6.5.4).

Note: This requirement is made since some operating systems require the use of associated files. However, data required by some other operating systems to enable the downloading of images and bibliographic data and the display of images cannot be accessed if stored in associated files.

Level of Implementation

6. The level of implementation for receiving systems (as defined in International Standard ISO 9660:1988, paragraph 13.5) shall be according to Level 2.

Note: This requirement implies that no Level 1 restrictions will be applied.



HANDBOOK ON INDUSTRIAL PROPERTY INFORMATION AND DOCUMENTATION

Ref.: Standards – ST.40

page: 3.40.12

Annex B, page 2

Boot Records

7. Boot records (as described in International Standard ISO 9660:1988, paragraph 8.2) are not obligatory.

Note: Boot records allow different operating systems to boot from the same disc. Due to lack of standardization between different operating systems and the lack of a user requirement to make use of a boot record, no need exists to make a boot record available on a CD-ROM disc.

Primary Volume Descriptor

8. The primary volume descriptors shall be as defined in International Standard ISO 9660:1988, paragraph 8.4 and Table 4. Elements of the primary volume descriptor requiring definition are listed below:

(a) *the system identifier* (as described in International Standard ISO 9660:1988, paragraph 8.4.5) shall be set to spaces;

Note: This indicates that no system identifier is specified. No requirement exists to identify the first 16 logical sectors on a CD-ROM containing images of patent documents.

(b) *The volume identifier* shall be the disc identifier as defined in paragraph 10 of this Recommendation;

Note: This requirement will ensure that the volume identifier is unique.

(c) *the maximum volume set size* (as defined in International Standard ISO 9660:1988, paragraph 8.4.10) shall be set to 1;

Note: This implies that each CD-ROM disc is treated as a volume separate from any other volume.

(d) *the field giving the volume set identifier* (as defined in International Standard ISO 9660:1988, paragraph 8.4.19) should be filled;

Note: This field must be specified and cannot be blanks. The generic name of CD-ROM discs produced by a supplier is given in this field according to this standard.

(e) *the volume creation date and time* (as defined in International Standard ISO 9660:1988, paragraph 8.4.26) shall set the year, month and day to the date of creation of the CD-ROM and set the remaining bytes to zeros;

Note: International Standard ISO 9660:1988 permits the setting of time down to a hundredth of a second. Only the setting of time to the date of creation of the CD-ROM disc is necessary.

(f) *the volume modification date and time* (as defined in International Standard ISO 9660:1988, paragraph 8.4.27) shall be set to the value of the volume creation date and time (as defined in International Standard ISO 9660:1988, paragraph 8.4.26);

Note: The same date as the date of creation of the original disc is implicit in the above statement.

(g) a value for *the volume expiration date and time* as defined in International Standard ISO 9660:1988, paragraph 8.4.28 shall not be specified (i.e., this field should be set to the "not recorded" data format).

Note: An expiration date and time is not appropriate for patent data.

Volume descriptor set terminator

9. The primary volume descriptor shall be followed by a volume descriptor set terminator (as defined in International Standard ISO 9660:1988, paragraph 8.3) with the volume descriptor version field (as defined in International Standard ISO 9660:1988, paragraph 8.3.3) set to 1.

Note: The volume descriptor set terminator indicates that it is the last volume descriptor in the volume descriptor set.



HANDBOOK ON INDUSTRIAL PROPERTY INFORMATION AND DOCUMENTATION

Ref.: Standards – ST.40

page: 3.40.13

Annex B, page 3

Bibliographic file identifier

10. The file identified by the bibliographic file identifier in the primary volume descriptor (as defined in International Standard ISO 9660:1988, paragraph 8.4.25) shall not be used.

Note: Due to the nature of patent data, a separate bibliographic file is not needed, and a bibliographic file as understood in International Standard ISO 9660:1988 should not be used.

Supplementary volume descriptor

11. Supplementary volume descriptors (as described in International Standard ISO 9660:1988, paragraph 8.5) shall not be used.

Note: This will ensure that CD-ROM discs will be supported by multiple disc operating systems, and will ensure that only those characters specified in International Standard ISO 646:1991 can be used for file identifiers and f or directory identifiers.

Volume partition descriptor

12. Volume partition descriptors (as described in International Standard ISO 9660:1988, paragraph 8.6) shall not be used.

Note: A volume partition descriptor is used to identify a partition on a CD-ROM disc. Since no user need to so identify partitions on a CD-ROM disc has yet been identified, volume partition descriptors should not be used.

Directory Records

13. In the formatting of directory records, extended attribute records shall not be used. The extended attribute record length field (as described in International Standard ISO 9660:1988, paragraph 9.1.2) shall be zero and bits 3 and 4 of the file flags field (as described in International Standard ISO 9660:1988, paragraph 9.1.6) shall be zero.

Note: The above requirement will facilitate the use of various operating systems.

Path Tables

14. Either a Type L or a Type M path table, or both, (as described in International Standard ISO 9660:1988, paragraphs 6.9 and 9.4) may be recorded on each CD-ROM.

Note: The recording of both a Type L and a Type M path table will allow applications developed on different disc operating systems to make efficient use of path tables.

15. Optional Path Tables may be used if desired.

Note: The use of optional path tables is designed to reduce the seek time and is recommended as a desirable feature.

[Annex C follows]



ANNEX C

CURRENT OR PROPOSED PRACTICES IN THE CODING OF CHARACTER SETS

1. *Used by the European Patent Office*

The character set used in ESPACE CDs conforms to the PC8 norm. This is because of the display and printing requirements of IBM compatible PCs. PC8 is an extended DOS character set for PCs. It contains the “normal” 128 characters of International Standard ISO 646:1991, augmented by another 128, defined on additional code pages. This character set will be expanded or modified to deal with future CD-ROMs comprising bibliographic and text data in various languages (Hungarian, Polish, etc.).

For its future mixed-mode CD-ROM a different approach will most likely be adopted:

- International Standard ISO 646:1991 (International Reference Version) will be the base code page
- Characters in the data records that cannot be defined with International Standard ISO 646:1991 will be coded/named with the entity references used in the SGML International Standard ISO 8879:1986, Annex D, “Character entities.”

That annex contains character sets for accented characters, Greek, Cyrillic, mathematical, etc. Characters receive mnemonic names (always placed between an ampersand and a semicolon e.g., “ü” is character ü).

The mnemonic names are system independent and can be made displayable by software routines dependent on the display facilities of the hardware being used; characters that are not displayable are made visible in a descriptive way.

Future enhanced character display sets can be adopted simply by changing the conversion table.

2. *Proposed by the Japanese Patent Office*

The JPO is of the opinion that WIPO Standard [ST.31](#) is not appropriate for a CD-ROM since WIPO Standard [ST.31](#) requires an escape sequence every time when switching code pages. The JPO feels that it is difficult to gain random access to process data when escape sequences are intermingled with the data.

In a usual Japanese document, 1-byte code characters are intermingled with 2-byte code characters so that it is strongly desirable to use no escape sequence between 1-byte code and 2-byte code in order to process Japanese document easily.

The JPO would like to propose CHARACTER SETS for CD-ROM as follows:

- (a) Coding of characters shall conform with International Standard ISO 2022:1986;
- (b) Control functions and Graphic character set which are defined in International Standard 646:1991 shall be designated as CO set and GO set according to International Standard ISO 2022:1986;
- (c) Other character sets necessary for a language shall be designated as G1 set (G2 set and G3 set will be additionally used if necessary) and invoked in area of column 10 to column 15 according to International Standard ISO 2022:1986.

According to the proposal, a character of which 8th bit is “0” corresponds to an International Standard ISO 646:1991 character and a character of which 8th bit is “1” is designated depending on the necessity of a language.

The JPO plans to designate JIS-X-0208-1990 character set as G1 set. JIS-X-0208-1990 is a Japanese standard which is named “Code of the Japanese graphic character set for information interchange.” This character set will be internationally registered according to International Standard ISO 2375:1985.

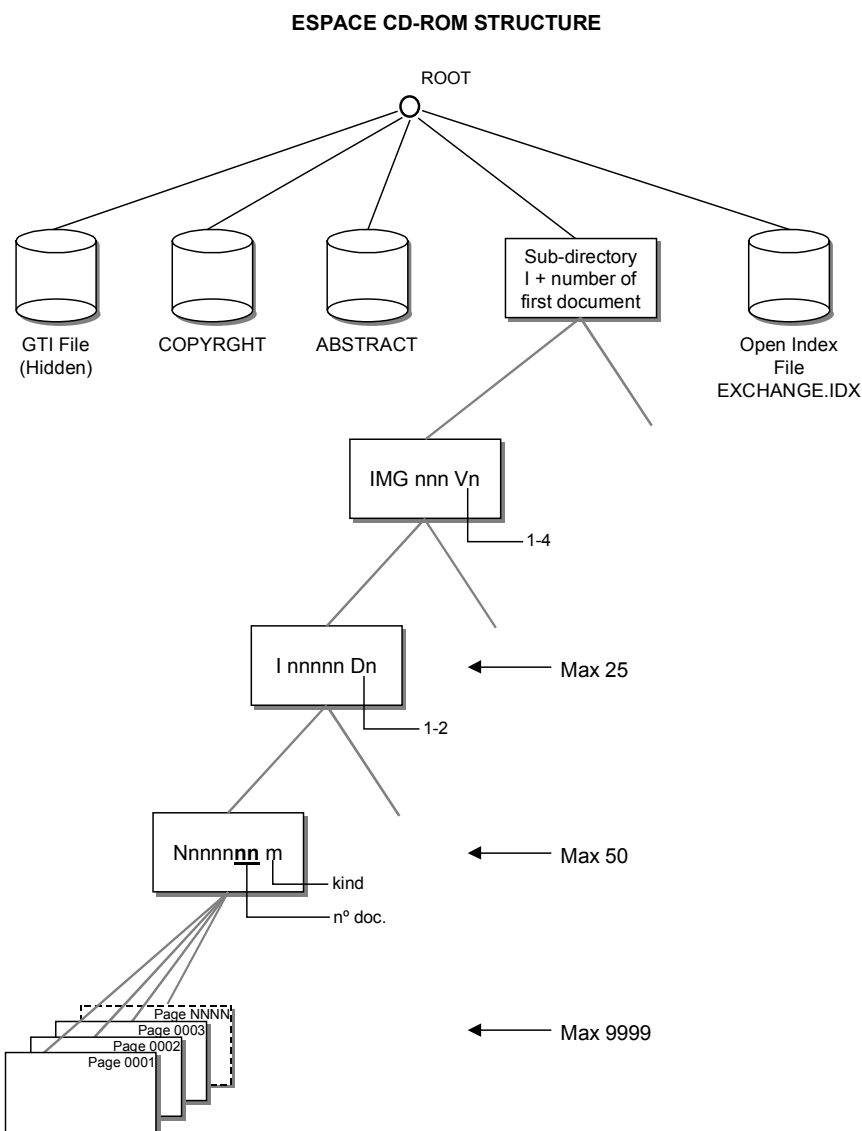
[Annex D follows]



ANNEX D

CURRENT OR PROPOSED PRACTICES IN RESPECT OF DIRECTORY STRUCTURE AND DOCUMENTARY FILES

1. Used by the European Patent Office



The file called COPYRGHT on ESPACE contains the copyright notice.

The file called ABSTRACT on ESPACE gives the name of the publishing office, the type of documents on the disc and their number range.

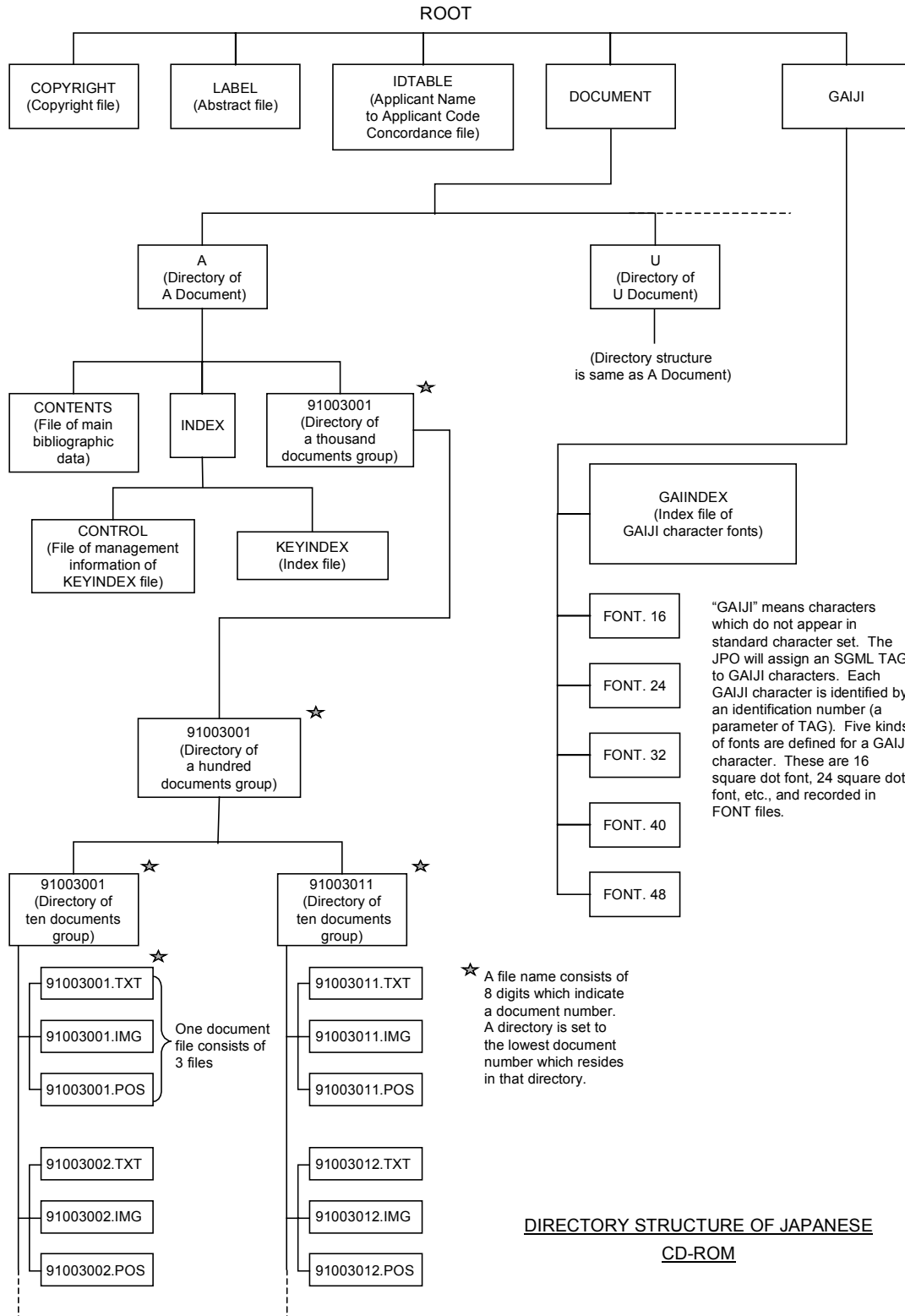
For a description of the file EXCHANGE.IDX, see Annex F.



HANDBOOK ON INDUSTRIAL PROPERTY INFORMATION AND DOCUMENTATION

Annex D, page 2

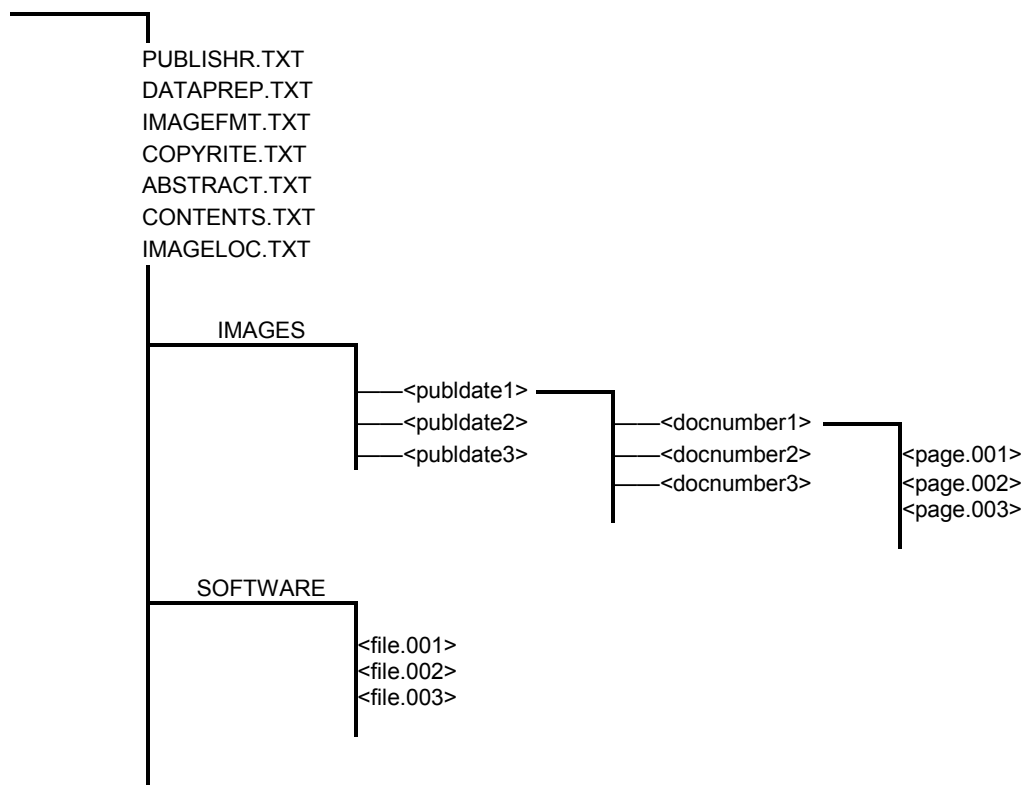
2. Proposed by the Japanese Patent Office





Annex D, page 3

3. Proposed by the United States Patent and Trademark Office



The file PUBLISHR.TXT shall be in the root directory and shall record the full name and address of the office issuing the patent documents recorded on the CD-ROM.

The file DATAPREP.TXT shall be in the root directory and shall record the full name of the office or its agent preparing the data for recording on the CD-ROM.

The file IMAGEFMT.TXT shall be in the root directory and shall record the image file format of the image data on the CD-ROM, either WIPO Standard [ST.33](#) or TIFF. In case the format is WIPO Standard [ST.33](#), the WIPO Standard [ST.33](#) version number and date shall be recorded. In case the format is TIFF, the TIFF version number and date and a complete but concise specification of the field tags used shall be recorded.

Note: It is likely that at some future date a set of patent-specific tags for use with the TIFF file format may be agreed upon, at which time the appropriate standard number may be referenced instead of enumerating and describing the field tags.

The file COPYRITE.TXT shall be in the root directory and shall record the full name of the owner(s) of proprietary rights held in the patent documents, retrieval software or other non-patent information contained on the disc.

The file ABSTRACT.TXT shall be in the root directory and shall record disc identification information of the CD-ROM.

The file CONTENTS.TXT shall be in the root directory and shall record a brief description of the contents of the CD-ROM.

[Annex E follows]



ANNEX E

CURRENT OR PROPOSED PRACTICES IN RESPECT OF IMAGE FILE FORMATS

1. Used by the European Patent Office

The image file is composed of a header of 256 bytes followed by the compressed image according to the recommendation T6 of CCITT (fax Group IV). Its length is variable (averaging around 50 kB) and is given in positions 37-46 of the EXCHANGE.IDX file.



The header is constructed in accordance with WIPO Standard [ST.33](#), Appendix 2 (where "Prefix" means "header") as follows (see WIPO Standard [ST.33](#) for a complete explanation):



HANDBOOK ON INDUSTRIAL PROPERTY INFORMATION AND DOCUMENTATION

Ref.: Standards – ST.40

page: 3.40.19

Annex E, page 2

PREFIX DEFINITION

M/D	ITEM No.	NAME	BYTES	TYPE
M	0	RECORD LENGTH	4	B
M	1	RECORD LENGTH	5	C
M	2	PUBLICATION OFFICE	2	C
M	3	KIND OF DOCUMENT CODE	2	C
M	4	DOCUMENT NUMBER	8	C
M	5	PAGE NUMBER	4	C
M	6	FRAME NUMBER	4	C
M	7	RECORD SEQUENCE NUMBER	2	B
*M	8	EMPEROR'S YEAR CODE	1	C
M	9	OTHERS (EXCHANGE USE)	19	C
D	10	OTHERS (DOMESTIC USE)	20	C
M	11	ORIGINATING OFFICE	2	C
M	12	DATE OF DRAW UP	6	C
M	13	RECORD STATUS	1	C
D	14	TOTAL PAGES	4	C
M	15	END OF FRAME NUMBER	4	C
M	16	TOTAL RECORDS	2	B
D	17	DISTINCTION OF REVISORY BULLETIN	1	C
D	18	SIZE OF DOCUMENT HEIGHT	3	C
D	19	SIZE OF DOCUMENT WIDTH	3	C
M	20	OTHERS (EXCHANGE USE)	20	C
D	21	OTHERS (DOMESTIC USE)	20	C
M	22	EXISTENCE OF BIBLIOGRAPHIC DATA	1	C
M	23	EXISTENCE OF CLAIM	1	C
M	24	EXISTENCE OF DRAWING	1	C
M	25	EXISTENCE OF AMENDMENT	1	C
M	26	EXISTENCE OF DESCRIPTION	1	C
M	27	EXISTENCE OF ABSTRACT	1	C
M	28	EXISTENCE OF SEARCH REPORT	1	C
M	29	OTHERS (EXCHANGE USE)	20	C
D	30	OTHERS (DOMESTIC USE, E.G., IPC SYMBOLS)	20	C
M	31	DATA TYPE	1	C
M	32	COMPRESSION METHOD OF IMAGE DATA	2	C
M	33	K-FACTOR CODE	2	C
M	34	RESOLUTION	2	C
M	35	SIZE OF FRAME HEIGHT	3	C
M	36	SIZE OF FRAME WIDTH	3	C
M	37	NUMBER OF LINES OF FRAME HEIGHT	4	C
M	38	NUMBER OF LINES OF FRAME WIDTH	4	C
D	39	ROTATION CODE	1	C
M	40	FRAME LOCATION X-AXIS COORDINATES	4	C
M	41	FRAME LOCATION Y-AXIS COORDINATES	4	C
M	42	FRAME STATUS	1	C
M	43	OTHERS (EXCHANGE USE)	19	C
D	44	OTHERS (DOMESTIC USE)	20	C
M	45	LENGTH OF IMAGE DATA	2	B
M	46	IMAGE DATA	V	B

M: Mandatory

D: Desirable

* M: Mandatory for data concerning Japanese documents only

B: Binary

C: Character

V: Variable

All format fields should be present, but only mandatory ones must contain data.

[Annex F follows]



ANNEX F

CURRENT OR PROPOSED PRACTICES IN RESPECT OF INDEXES

I. OPEN INDEX

1. *Used by the European Patent Office*

The EXCHANGE.IDX file is an editable MS-DOS ASCII file enabling a user to gain direct access to the documents on ESPACE CD-ROMs without using the ESPACE software. It identifies and gives the location of the individual pages of each document, but does not indicate the contents of those pages.

The file comprises a succession of blocks (one block being one line of the file). Each block has a length of 101 bytes. A document of 13 pages is described by 13 blocks in the EXCHANGE.IDX file.

The positions within each block are counted in bytes, starting with one. Each byte corresponds to one item of information about the document (except for the reserved spaces).

Block Structure

Position	1 – 2:	Country Code (i.e., EP for the European Patent Office)
	3 – 4:	Type of document ("KIND"): A0, A1, A2...
	5:	(Reserved Space)
	6 – 13:	Publication Number, eight digits
	15 – 18:	Page number of document (also in positions 88-91)
	19 – 22:	(Reserved Space)
	23 – 26:	Decimal number written in the form "0000" to "0127" indicating the presence of sub-documents. Each type of sub-document is represented by a number. The number present is the sum of the numbers representing all the sub-documents present:

<i>number</i>	<i>sub-document</i>
64:	bibliographic data
32:	claim
16:	drawing
8:	amendment
4:	description
2:	abstract
1:	search report

Examples of positions 23-26:

number	positions 23-26	signifies:
66	0066	bibliographic data and abstract
36	0036	claim and description
4	0004	description



HANDBOOK ON INDUSTRIAL PROPERTY INFORMATION AND DOCUMENTATION

Ref.: Standards – ST.40

page: 3.40.21

Annex F, page 2

- Position 27 – 36: address of the image file (ten digits). This is the address of the first block.
- 37 – 46: Length of the image file (ten digits).
- 47 – 91: Description of the MS-DOS access path. This is a string of four directory names, each of up to eight characters. These names are derived from the publication number and type of the document concerned:
- publication number: xxxyyzz
 - kind: k
- 47 – 55: First directory. This is the number of the first document on the disk preceded by I: e.g., I0360792.
- 56 – 64: The second directory name always starts with "IMG," followed by the three first digits of the documents in it. After this there is the character "V" and a single digit, which is connected to the identifier of the third directory, e.g., IMG036V1.
- The final digit is determined by the name of the third directory. It has the value of either 1, 2, 3 or 4 depending upon the value of the positions 70-71 (yy).
- | | |
|-------|----------------|
| yy | V followed by: |
| 00-24 | 1 |
| 25-49 | 2 |
| 50-74 | 3 |
| 75-99 | 4 |
- The principle is to limit the number of subdirectories to less than 25.
- 66 – 73: The third directory starts with "I" followed by five digits xxxyy, then the character "D" and a digit which is either 1 or 2. The final digit is determined by the fourth directory (zz):
- | | |
|-------|----------------|
| zz | D followed by: |
| 00-49 | 1 |
| 50-99 | 2 |
- The principle is to limit the number of subdirectories to less than 50.
- 74 – 82: The name of the fourth directory consists of the publication number of the document (same as positions 7-13) and the digit indicating the kind: e.g., xxxyyzzk.
- 83 – 91: Each page of the document is stored in a file. The name of the file is "PAGE" followed by the number of the page: e.g., PAGE 0001 or PAGE0120.
- 92 – 99: (Reserved Space).
- 100: Carriage Return (CR).
- 101: Line-feed (LF).

Brief Resume of the Access Path for ESPACE and FIRST Documents

To minimize the number of documents in any given directory, the path is divided into successive levels of subdirectories.

Each document in ESPACE is an individual directory at the fourth level from the root. These documents are themselves grouped in directories containing no more than 50 documents (third level). The second level comprises at the most 1250 documents (25 directories of 50 documents). The first level bears the number of the first document on the CD-ROM.



HANDBOOK ON INDUSTRIAL PROPERTY INFORMATION AND DOCUMENTATION

Ref.: Standards – ST.40

page: 3.40.22

Annex F, page 3

For ESPACE, the file EXCHANGE.IDX describes each page of each document, the fourth level of directories on the CD-ROM each containing as many files as there are pages in the document.

FIRST, on the other hand, only contains the first pages of each application (and occasionally the second). In this case, the access path is the same as that in ESPACE. For a document with a single page, the access path is only via three subdirectories leading to a file containing the document. A document is thus no longer represented by a directory, but by a file.

Access Path on ESPACE:	4 successive subdirectories + N files
Access Path on FIRST:	3 successive subdirectories + 1 file (exceptionally, same as ESPACE).



HANDBOOK ON INDUSTRIAL PROPERTY INFORMATION AND DOCUMENTATION

Ref.: Standards – ST.40

page: 3.40.23

Annex F, page 4

EPA1 00 3813 00010000006600060456970000034066\I0443813\IMG044V2\I04438D1\044381
31\PAGE0001
EPA1 00443813 00020000000400060805130000106542\I0443813\IMG044V2\I04438D1\044381
31\PAGE0002
EPA1 00443813 00030000000400061890570000114068\I0443813\IMG044V2\I04438D1\044381
31\PAGE0003
EPA1 00443813 00040000000400063037450000091108\I0443813\IMG044V2\I04438D1\044381
31\PAGE0004
EPA1 00443813 00050000000400063959050000035720\I0443813\IMG044V2\I04438D1\044381
31\PAGE0005
EPA1 00443813 00060000000400064327690000026162\I0443813\IMG044V2\I04438D1\044381
31\PAGE0006
EPA1 00443813 00070000003200064593930000023862\I0443813\IMG044V2\I04438D1\044381
31\PAGE0007
EPA1 00443813 00080000000100064839690000029130\I0443813\IMG044V2\I04438D1\044381
31\PAGE0008
EPA2 00443814 00010000006600065146890000052642\I0443813\IMG044V2\I04438D1\044381
42\PAGE0001
EPA2 00443814 00020000000400065679370000103722\I0443813\IMG044V2\I04438D1\044381
42\PAGE0002
EPA2 00443814 00030000000400066723850000104316\I0443813\IMG044V2\I04438D1\044381
42\PAGE0003
EPA2 00443814 00040000000400067768330000106236\I0443813\IMG044V2\I04438D1\044381
42\PAGE0004
EPA2 00443814 00050000000400068833290000104072\I0443813\IMG044V2\I04438D1\044381
42\PAGE0005
EPA2 00443814 00060000003600069877770000103044\I0443813\IMG044V2\I04438D1\044381
42\PAGE0006
EPA2 00443814 00070000003200070922250000075918\I0443813\IMG044V2\I04438D1\044381
42\PAGE0007
EPA2 00443814 00080000003200071700490000013714\I0443813\IMG044V2\I04438D1\044381
42\PAGE0008
EPA2 00443814 00090000001600071843850000033838\I0443813\IMG044V2\I04438D1\044381
42\PAGE0009
EPA2 00443814 00100000001600072192010000035158\I0443813\IMG044V2\I04438D1\044381
42\PAGE0010
EPA2 00443814 00110000001600072560650000035462\I0443813\IMG044V2\I04438D1\044381
42\PAGE0011
EPA2 00443814 00120000001600072929290000028786\I0443813\IMG044V2\I04438D1\044381
42\PAGE0012
EPA2 00443814 00130000001600073236490000026220\I0443813\IMG044V2\I04438D1\044381
42\PAGE0013
EPA2 00443814 00140000001600073502730000016802\I0443813\IMG044V2\I04438D1\044381
42\PAGE0014
EPA1 00443815 00010000006600073687050000027942\I0443813\IMG044V2\I04438D1\044381
51\PAGE0001
EPA1 00443815 00020000000400073973770000063058\I0443813\IMG044V2\I04438D1\044381
51\PAGE0002
EPA1 00443815 00030000000400074608650000062072\I0443813\IMG044V2\I04438D1\044381
51\PAGE0003
EPA1 00443815 00040000003600075243530000023850\I0443813\IMG044V2\I04438D1\044381
51\PAGE0004
EPA1 00443815 00050000000100075489290000032048\I0443813\IMG044V2\I04438D1\044381
51\PAGE0005
EPA2 00443816 00010000006600075816970000032864\I0443813\IMG044V2\I04438D1\044381
62\PAGE0001
EPA2 00443816 00020000000400076165130000098528\I0443813\IMG044V2\I04438D1\044381
62\PAGE0002
EPA2 00443816 00030000000400077168650000098904\I0443813\IMG044V2\I04438D1\044381
62\PAGE0003
EPA2 00443816 00040000000400078172170000106140\I0443813\IMG044V2\I04438D1\044381
62\PAGE0004
EPA2 00443816 00050000003600079237130000093204\I0443813\IMG044V2\I04438D1\044381
62\PAGE0005
EPA2 00443816 00060000003200080179210000067304\I0443813\IMG044V2\I04438D1\044381
62\PAGE0006



HANDBOOK ON INDUSTRIAL PROPERTY INFORMATION AND DOCUMENTATION

Ref.: Standards – ST.40

page: 3.40.24

Annex F, page 5

2. *Proposed by the United States Patent and Trademark Office as one of the possible practices*

An open index (to be named the General Purpose Index) will be present on each CD-ROM, which will provide the physical location of each image stored on the disc. The index file will be named IMAGELOC.TXT and shall be included in the root directory of the CD-ROM. It will provide for each patent document on the CD-ROM the following information in records which will be separated by a carriage return/line feed pair:

	Bytes:	Contents:
1	1-2	Two-letter office code (INID 19)
2	3-4	Kind-of-document code (INID 13)
3	5-12	Document number (INID 11)
4	13-16	WIPO Standard ST.33 files: <page number>; TIFF files: 0000
5	17-20	WIPO Standard ST.33 files: <frame number>; TIFF files: 0000
6	21-24	WIPO Standard ST.33 files: <subdoc ID>; TIFF files: 0000
7	25-34	Image location (absolute starting byte number)
8	35-44	Image length
9	45-(X)	Full path (ccccccc ccccccc ccccccc ...)
10	(X+1)-(Y)	File name (ccccccc.ccc)
11	(Y+1)-(Y+2)	Record separator (carriage return, line feed)

Fields 3-8 shall be padded on the left with the number zero.

In case the image file format is WIPO Standard [ST.33](#), fields 4-6 shall contain the page number, frame number, and subdoc ID. In case the image file format is TIFF, fields 4-6 shall contain the number 0000.

The path name (field 9) shall consist of the relevant directory names in top-down hierarchical order, each directory name preceded by the “ ” character.

The file name (field 10), preceded by the “ ” character, shall immediately follow the path. The characters appearing between the last “ ” and the record separator (carriage return, line feed) will therefore be the file name.

II. DOCUMENT LOCATION INDEX

1. *Used by the European Patent Office*

On discs published by the EPO, the EXCHANGE.IDX file (described above as their open index) serves the purpose of a document location index.

2. *Proposed by the United States Patent and Trademark Office*

As recommended in paragraph 20 of this Recommendation.

III. CUMULATIVE DOCUMENT LOCATION INDEX

1. *Used by the European Patent Office*

The ESPACE/ACCESS product constitutes a cumulative document location index for ESPACE/EPA and ESPACE/FIRST. ESPACE/ACCESS, published quarterly, contains searchable bibliographic information for all European patent applications from 1978 to date. For each patent application, a reference is included indicating on which ESPACE/EPA and ESPACE/FIRST disc the same application appears.

2. *Proposed by the United States Patent and Trademark Office*

As recommended in paragraph 21 of this Recommendation.

[Annex G follows]



ANNEX G

CURRENT OR PROPOSED PRACTICES IN RESPECT OF CLIENT/SERVER MODELS

The SilverPlatter Exchange

Volume 4, No. 1

A publication for and about CD-ROM users

May 1991

Meeting the Challenge: CD-ROM Interchangeability

by Peter Ciuffetti, Vice President, Network Publishing Division

The access of electronic publications through CD-ROM is entering its seventh year. Existing standards have simplified the introduction of new titles, which, in turn, has encouraged industry growth and stability. The 1985 Yellow Book Standard made it possible for any manufacturer's CD-ROM drive to play any disc. The 1988 ISO 9660 file system standard made it possible for any operating system to find any data file on the disc. These facilities have fostered the introduction of thousands of CD-ROM titles with numerous new publishers entering the industry.

As new publishers enter the market they attempt to develop the definitive user interface, adding to the collection of software to be learned by the researcher. The attraction of searching databases on CD-ROM becomes diluted when the searcher is faced with having to know and understand many interfaces. As CD-ROM publishers, we have an obligation to come up with a solution to this dilemma.

We cannot expect that all publishers converge to a standard interface. There is no such thing as a definitive user interface; it is a personal choice on the part of the researcher. We need to continue to encourage innovation, invention, and the entrepreneurial spirit in CD-ROM development. And we need to let the researcher have the choice of interface, regardless of the database to be searched.

The Next CD-ROM Standard

The technique already exists to give the researcher choice of interface. It is called client/server architecture and it has been around since the computer was invented. Simply, it involves separating the database search engine (server) from the user interface (client) and using a messaging system through which they communicate with each other. It is the messaging system which follows a well-defined specification called a client/server protocol.

Essentially, all that each interface (client) and each search engine (server) need to understand is how to communicate with the messaging system. They do not have to know how each other performs their respective duties. So, in theory, a searcher can use any interface to access any database which conforms to the standard messaging system protocol.

The protocol needs to be well defined and standardized, and most important, accepted and implemented by vendors. A standard messaging protocol can open opportunities which transcend those offered by the current standards in the industry. Foremost, one user interface can be used to search any database, regardless of vendor. It gives the researcher the choice.

Continued ...

**Opportunities for Growth and Economy**

The client/server architecture provides opportunities for individual vendors as well as researchers. Most databases sold today are available on only one platform, usually a PC compatible running DOS. Even when there is a choice of platforms, the choice is usually limited. Many researchers have a different kind of machine on their desktop than the one supported by the database vendor. For a vendor to make their data-

base available on additional platforms can require expensive programming efforts to port the retrieval software to each new type of machine. Extensive porting is typically not profitable because of a limited installed base of a particular machine in a given market.

As a contrast to PCs, most UNIX-based desktop machines are networked. In a network environment, the client and the server can pass messages among various desktop machines not of the same type. To support this environment, vendors may only need to make the server (search engine) compatible with the host machine. Initially, they may not need to create a user interface because there may be a conforming interface already available from another vendor. And, even if a user interface is needed for various types of desktop machines, the effort is less than half when compared to supporting both interface and search engine on different machines.

Perhaps the most promising opportunity introduced by the client/server architecture is the possibility of querying diverse information sources with a single user interface. The definition of the client/server protocol does not need to be limited to bibliographic CD-ROM databases. Many

information resources, such as full text databases, catalogs, and mainframe databases could be fitted with a front-end that understands the queries defined by the protocol. This front-end would also package the results in standard response messages to be displayed by the client in its own fashion.

This level of integration would help dissipate the artificial barriers among information sources today.

There are currently three standards under various stages of development that could be an appropriate client/server standard. They are: the Information Retrieval Protocol (Z39.50) sponsored by the library community; the Structured Full-Text Query Language (SFQL) sponsored by the aerospace industry; and the CD-ROM Read-only Data Exchange (CD-RDX) sponsored by the government's Information Handling Committee. Most vendors would find that these protocols need more work before they are well enough defined to be implemented.

The completion of an intelligent standard is important, though not as important as the need for each database publisher to be committed to the adherence to standards at this level and to participate in its development. The user community should stress the importance of these standards with each vendor they do business with, since it is the user's voice which is most heard. SilverPlatter's success has resulted from a commitment to standards and we will continue that strategy as long as there remains an aspect of research which warrants standardization.

Note: SFQL was distributed as PCIP/P 918/90 Rev.7, Annex 33 and CD-RDX was distributed as PCIP/P 918/90 Rev.8, Annex 39. The most recent version of Z39.50 will be distributed to the Working Group at a later date.

(Reprinted with kind permission of the publisher)

[Annex H follows]



ANNEX H

BIBLIOGRAPHIC DATA: FIELD NAMES AND FIELD FORMATS

- (a) Number of the document (INID Code 11)
Field name: PN
– User permitted by the software to enter one or other, or both, of the following formats:
- (i) Fixed format of eight digits. No hyphens or slashes, or other separators, to be used. If the number as published by the office has less than eight digits, user infills zeros at the beginning to make the eight digits.
 - (ii) Exact publication number as used by the office on the document. Software should recognize less than eight digits and insert filling zeros as necessary, removing hyphens or slashes, or other separators between digits forming the publication number. Check digits are not considered part of the publication number;
- (b) Kind-of-document code (INID Code 13)
Field name: KD
– The one or two character code according to WIPO Standard [ST.16](#);
- (c) Office publishing the document (INID Code 19)
Field name: PC
– Two-letter code according to WIPO Standard [ST.3](#);
- (d) Application number (INID Code 21)
Field name: AP
– Eight characters which present the minimum significant part as defined in WIPO Standard [ST.10/C](#), right adjusted with infilling zeros. PCT application numbers should be entered using the format YYWQAANNNNNN, where YY = last two digits of the year, AA = two-letter code of the receiving office, NNNNNN = application number;
- (e) Application date (INID Code 22)
Field name: AD
– YYMMDD, i.e., last two digits of the year, followed by the two digits of the month, followed by the two digits of the day, infilling zeros are used as necessary;
- (f) Priority application number(s) (INID Code 31)
Field name: PRN
– Eight characters which present the minimum significant part as defined in WIPO Standard [ST.10/C](#), right adjusted with infilling zeros;
- (g) Priority date(s) (INID Code 32)
Field name: PRD
– YYMMDD, i.e. last two digits of the year, followed by the two digits of the month, followed by the two digits of the day, infilling zeros are used as necessary;
- (h) Priority country or organization (INID Code 33)
Field name: PRC
– Two-letter code according to WIPO Standard [ST.3](#);
- (i) Priority data (INID Code 30)
Field name: PR
– If the software permits the search of the full priority information, e.g., the combination of (f), (g) and (h), the data should be entered using the format YYAANNNNNNNN, where YY = last two digits of the year of priority, AA = priority country, NNNNNNNN = priority application number;
- (j) Publication date (INID Code 40)
Field name: PD
– YYMMDD, i.e., last two digits of the year, followed by the two digits of the month, followed by the two digits of the day, infilling zeros are used as necessary;



HANDBOOK ON INDUSTRIAL PROPERTY INFORMATION AND DOCUMENTATION

Ref.: Standards – ST.40

page: 3.40.28

Annex H, page 2

- (k) International Patent Classification (INID Code 51)
Field name: IC
– ANNANN|NNNN, where A = alpha, N = numeric, | = slash or colon as appropriate. In the three positions before the slash or colon, the digits should be right adjusted and zero filled as necessary. At least two, and no more than five, numerals may be given after the slash or colon, depending on the symbol. In some instances, a further character, as defined in WIPO Standard [ST.8](#), may be given serving to characterize the use of an IPC symbol;
- (l) Words from title (INID Code 54)
Field name: TI. Where offices publish titles in two or more languages, the following field names| should be used
ET – English title
FT – French title
GT – German title
ST – Spanish title
– Each word from the title, also including words with digits, e.g., chemical names, except common words which appear on a stop list. The stop list should be published in the documentation supplied to users. Words taken from titles in different languages should be stored in separate fields for each language;
- (m) Words from abstract (INID Code 57)
Field name: AB
– Each word from the abstract, also including words with digits, e.g., chemical names, except common words which appear on a stop list. The stop list should be published in the documentation supplied to users. Words taken from abstracts in different languages should be stored in separate fields for each language;
- (n) Name of applicant(s) (INID Codes 71, 75, 76)
Field name: PA
– Full name and single words from full name up to at least 30 characters. Personal names should be expressed as family name, space, first name, space, initial as necessary;
- (o) Name of inventor(s) (INID Code 72)
Field name: IN
– Full name and single words from full name up to at least 30 characters. Full names should be expressed as family name, space, first name, space, initial as necessary;
- (p) Name of attorney or agent (INID code 74)
Field name: AT
– Full name and single words from full name up to at least 30 characters. Full names should be expressed as family name, space, first name, space, initial as necessary;
- (q) Language in which the application is published (INID Code 26)
Field name: LA
– The code of the language as provided for in International Standard ISO 639:1988;
- (r) Designated States (INID Codes 81 and 84)
Field name: DS
– Two-letter code according to WIPO Standard [ST.3](#);
- (s) Domestic or national classification (INID Code 52)
Field name: NC
– The codes and symbols as given in the domestic or national classification.
- Note:* The preferred order of writing dates as given in subparagraphs (e), (g) and (j), above, is in accordance with International Standard ISO 8601:1988 and follows WIPO Standard [ST.30](#). The use of any other order of expressing dates shall only be permitted if the software prompts a user to input days, months and years, in any order.

[End of Annexes and of Standard]