“EFFECT”
Exchange Format For
Electronic Components and Texts

Technical Specifications
Version 4.0 — October 1995

Author	Paul Mostert Elsevier Science
Mail	Molenwerf 1 1014 AG Amsterdam The Netherlands
Telephone	+31 (0)20 - 485 3574
Fax	+31 (0)20 - 485 3706
E-mail	P.Mostert@Elsevier.NL

Elsevier Science makes no representations concerning the suitability of these EFFECT Technical Specifications for any particular purpose. It is provided “as is” without express or implied warranty of any kind.

Table of contents

1 Introduction

THIS TECHNICAL SPECIFICATIONS MANUAL describes the structure and format of electronic journals by Elsevier Science, which are delivered to your organization. These electronic journals can be used by your organization for your own local Current Awareness and Article Delivery database systems. This Manual describes

- The overall structure of the material and the individual technical components.
  - Structured bibliographic data of all editorial items (e.g. title, authors, abstract, etc.) in plain tagged text format.
  - Structured bibliographic data of selected editorial items in SGML format.
  - Full unedited body text material of the editorial items.
  - Image files of all pages in the journal issues.
  - Files in other electronic formats.
- How the material of the journals on which a subscription is running will arrive at your organization.

1.1 *A special note for TULIP and EASE projects*

The information in this manual combines specifications used in different projects performed by Elsevier Science and others. The "*EFFECT* Technical Specifications Version 4.0" hereby supersedes the "*TULIP* Technical Specifications Version 2.2", dated March 1994 and "*EASE* Specifications Version 3.0" used for the EASE project (Elsevier research project with Tilburg University, The Netherlands). To avoid conflicts with version numbering the first *EFFECT* Technical Specifications manual starts with Version 4.0. For clarity the division into chapters is reorganized. The structural differences with earlier Specifications are indicated by a vertical bar in the left margin (like this paragraph). Minor textual changes or corrections are not highlighted.

Overview of changes in *EFFECT* Specifications Version 4.0:

- Introduction of new tags
  - **`_io`** Former ISSN (see page 15)
  - **`_jf`** Full set journal name (see page 15)
  - **`_if`** Full set journal ISSN (see page 15)
  - **`_cr`** Copyright notice (see page 16)
  - **`_si`** Status indicator/message (see pages 12, 15, 18 and 22)
  - **`_cf`** Conference details (see page 19)

- Corrections / Update of current tags
    - ○ **_t0** Dataset Identifier (no real change but sharpening of definition; see page 12)
    - ○ **_t2** Issue Identifier (no real change but sharpening of definition; see page 17)
    - ○ **_vl** Volume Number of Journal Issue (minor change “ **/** ” to “ **-** ”; see page 18)
    - ○ **_is** Issue Number (minor change “ **/** ” to “ **-** ”; see page 18)
    - ○ **_ii** Item identification (change to enable future possibilities; see page 22)
    - ○ **_kw** Keyword (change to enable future possibilities; see page 26)
    - ○ **_mf** **[SGML]** Full item text in Standard Generalized Markup Language and accompanying artwork / figure files in a variety of formats (sharpening of definition; see page 27).
- Change from checksum algorithm based on UNIX “sum” to a checksum algorithm based on RSA’s “MD5” (see page 7).
- The Chapter “Feedback to Elsevier” has been removed from this Technical Specifications and is separately available.

2 Dataset Components

Electronic journal material is bundled into so-called *Datasets*. A Dataset is a collection of several journal issues from a selection of journal titles, organized by period, subscription or any order according to the particular arrangement between Elsevier and your organization.

In this chapter technical information is provided regarding the different components of Dataset such as page images, raw texts, SGML files, etc. The next chapter describes how these components interrelate.

2.1 *Page images*

Every page in a journal issue is scanned by means of a high-volume scanning machine or converted from other electronic files. This results in one page image file per scanned page. Page images are standard black / white single-page TIFF 5.0 files with a scan resolution of 300 dots per inch (dpi). The maximum scan size is European A4, i.e. 210 x 297 mm^2.

The files are compressed according to the international standard ITU T.6 encoding scheme, formerly known as CCITT Fax Group IV, on average achieving a compression ratio of 8%. A typical 1 Megabyte image is compressed to ± 80 Kilobytes. Pages with large contiguous white areas gain a better ratio, pages with photographs on them score worse.

All software supporting TIFF with Fax Group IV compression functionality should be able to manipulate the image files. Images will be in standard TIFF 5.0 with white background and black characters. Decompression could be performed in software, for instance with the programs HiJaak, Image Alchemy, Kofax 910 software, PaintShop Pro, TIFFLIB and a variety of commercial and public domain toolkits. Alternatively, hardware accelerators with special Application Specific Integrated Circuits (ASICs) are available which decompress images in realtime such as Kofax or Xionics decompression boards. Typical software decompression times, on a standard 80486 PC, rate from 15 seconds to several minutes, whereas hardware decompression times rate tenths of seconds.

An average editorial item contains about 8 pages. This rates for long full length research articles and short ones such as review articles, letters, abstracts of conference papers, errata, book reviews, other reviews and editorials.

2.2 Raw text files

Every page image file has a corresponding “raw” ASCII file. These text files are produced as a result of Optical Character Recognition (OCR) procedures. The files are further called “raw” since no keyboarding / editing / spell-checking is performed on them.

The text files conform to the ASCII code, hence contain only plain ASCII-characters 32 to 126. Lines are of variable length and are in “stream” mode. The end of a line of text will be denoted with the MS-DOS character-combination ASCII 13,10 (Carriage Return, Linefeed). In other operating environments than MS-DOS this should be translated into a similar native End-Of-Line combination.

Raw text files are intended to provide a basis for creating searchable indexes. They should not be used to show to end users.

2.3 SGML files

The text of editorial items is available structured according to Standard Generalized Markup Language (SGML) rules.

SGML is an international standard within the International Standards Organisation for the coding and presentation of text which allows the storage, transmission, display and editing of material through a descriptive environment. Rather than indicating detail of how a document should be presented, SGML describes the document structure in a standard way. Separating the document structure from a particular representational style opens up new possibilities for various presentation and output medium formats.

SGML documents are based on a Document Type Definition (DTD), a description of the structure of a particular type of publication, such as a full length article. It includes all the possible elements that such an article could contain, together with the hierarchical relationships between those elements and the mandatory / optional attribute.

SGML files are coded in plain ASCII. Structure codes are embedded in the text and are identified by angle brackets, “ < ” for open and “ > ” for close.

Datasets can have two different types of SGML files. These are recognizable as files with two different extensions:

- Files with the extension **.SGM** hold the full content (bibliographic information, full article text and references) of editorial items.
- Files ending in **.SGC** contain only the bibliographic information of an editorial item.

The DTD and its description which is used for editorial items is available separately from Elsevier’s Internet FTP server “FTP.Elsevier.NL” or upon request.

2.4 Other files

The intention of the Technical Specifications is to provide an open environment for electronic journal information. Therefore other files than the ones described above can be added as the need arises without friction with the Technical Specifications. Agreements between supplier and receiver on a particular file format should exist. See also the description of manifestation format on page 27. Examples are:

- ○ Adobe Acrobat™ Portable Document Format (PDF) files.
- ○ Encapsulated PostScript (EPS) files.
- ○ Joint Photographer Expert Group (JPEG) encoded files.
- ○ Hypertext Markup Language (HTML) files.
- ○ Compuserve Graphics Interchange Format (GIF) compressed files.
- ○ TEX (TEX) encoded files.

2.5 CHECKMD5.FIL

Experiences in delivering large quantities of data either through network transfer or CD-ROM technology has shown that technologies which are supposed to have sufficient error correction mechanisms sometimes falter in rare occasions. Because one incorrect bit in a binary file such as a page image corrupts the entire file, the chance that such an event occurs should be minimalized. As a precautionary measure an additional checksum facility is available that could be used to verify the correctness of files.

In earlier projects for this UNIX's “sum” command has been used, but incompatibilities between different versions of UNIX and MS-DOS induced rejection of “sum”-based algorithms. The current algorithm is the RSA Data Security, Inc. MD5 Message-Digest Algorithm.

The MD5 algorithm takes as input a message of arbitrary length and produces as output a 128-bit “fingerprint” or “message digest” of the input. It is conjectured that it is computationally infeasible to produce two messages having the same message digest, or to produce any message having a given prespecified target message digest. The MD5 algorithm is intended for digital signature applications, where a large file must be “compressed” in a secure manner before being encrypted with a private (secret) key under a public-key cryptosystem such as RSA. The MD5 algorithm works identically in all known UNIX and MS-DOS environments, regardless of byte order. RSA placed the MD5 algorithm in the public domain for review and possible adoption as a standard. The description and the source for this program is available from RSA's FTP server (ftp.rsa.com).

Every directory except the “root” directory of a Dataset, which contains page images, raw text files or other files related to items or pages, has a file *CHECKMD5.FIL* that contains a computed checksum for every file in the directory, except for *CHECKMD5.FIL* itself. Every line contains a 32-position checksum (the hexadecimal representation of the 128-bit “fingerprint”) followed by the file name, separated by at least one blank.

Example: In a particular directory the six files *1.TIF*, *2.TIF*, *1.RAW*, *2.RAW*, *94000138.SGC* and *CHECKMD5.FIL* could be found. The file *CHECKMD5.FIL* has the following contents:

```
06d51a03a5d608a2dc8463a4838af18d 1.RAW
ca37dce1c5a8e1630a74cf49212758a0 1.TIF
1895cbee155e70cad5f5ab453770a253 2.RAW
f06ddb8bd1ecf20f0ed245aa99daad1b 2.TIF
a91618c0ccd3b662459a3ab72344da34 94000138.SGC
```

2.6 DATASET.TOC

Each Dataset contains the file *DATASET.TOC* in the “root”-directory, in which all cross-indexing reference data is provided. This file is the main entry point to the Dataset and contains all information to reconstruct journal issues and editorial items contained in it. *DATASET.TOC* contains some redundancy in order to have information easily readable and accessible to the human eye. Subsequent electronic processing could simply strip off this redundant information.

General rules for the *DATASET.TOC* file are:

- The file does only contain plain ASCII-characters (characters 32 to 126). The end of a line of text will be denoted in MS-DOS with the character-combination ASCII 13,10 (Carriage Return, Linefeed). In other operating environments this should be translated into a similar native End-Of-Line combination.
- Diacritical characters with accents, such as appearing in foreign names or titles, are converted to their non-diacritical form. E.g. *é*, *è*, *ê* and *ë* are all converted to “ **e** ”.
- The greek alphabet and a restricted number of mathematical characters are converted to a combination of the @-character and a character or number. E.g. α-*melamine* will appear as **@a-melamine** and $a \leq b$ appears as **a@<b**. See Appendix A on page 30 for a complete overview.
- Superscripted characters will be preceded by a caret “ **^** ”, e.g. a^2 appears as **a^2** and b^{123} as **b^1^2^3**. Subscripted characters are constructed by a double quote “ **"** ”, e.g. H_2O will appear as **H"2O** and I_{sum} as **I"s"u"m**.
- Bold, italic and underlined characters will be converted to their normal representation.
- A line of text will in general not be longer than 80 positions. The rare exception is for very long words which may not be broken.

- *DATASET.TOC* is split up into records, which are broken down into fields.
    - A new record starts on a new line with an underline “ _ ”, the character “**`t`**” or “**`T`**”, a single digit (**`0`**, **`1`**, **`2`** or **`3`**) and a space, followed by the content of the field. E.g. **`_t0`**, **`_t3`**.
    - A new field starts on a new line with an underline “ _ ” , two alphanumerical characters and a space, followed by the contents of the field itself. Fields which are longer than 80 positions are wrapped to following lines.
    - A line starting with four spaces is a continuation of the previous line.
- White lines are included for better human readability. These lines can be ignored for electronic processing.

3 Dataset Structure

The structure of the Dataset follows a directory structure which reflects the subdivision into journals, issues, editorial items (articles) and pages. Please note that further in this Technical Specifications the MS-DOS conventions for directory and file names are used (for instance a backslash “ \ ” as separator between directory names instead of the UNIX forward slash “ / ”). The structure can be represented as follows:

```
Dataset   Journal        Issue         Pages/Ed.Items     Components

T:\ ──┬── 00406090\ ──┬── V0193I01\ ──┬── 94000123\ ───┬── MAIN.SGM
      │               │               │                ├── FIG1.TIF
      │               │               │                ├── FIG2.JPG
      │               │               │                └── CHECKMD5.FIL
      │               │               ├── 9500064X\ ───┬── MAIN.SGM
      │               │               │                ├── FIG1.JPG
      │               │               │                └── CHECKMD5.FIL
      │               │               ├── 1.TIF
      │               │               ├── 1.RAW
      │               │               ├── 2.TIF
      │               │               ├── 2.RAW
      │               │               ├── 94000123.SGC
      │               │               ├── 94000123.PDF
      │               │               ├── 9500064X.SGC
      │               │               ├── 9500064X.PDF
      │               │               ├── .........
      │               │               └── CHECKMD5.FIL
      │               ├── V0201X02\ ──┬── 94002347\ ───┬── MAIN.SGM
      │               │               │                ├── FIG1.JPG
      │               │               │                └── ....
      │               │               ├── 1.TIF
      │               │               ├── 1.RAW
      │               │               ├── 2.TIF
      │               │               ├── 2.RAW
      │               │               ├── 94002347.SGC
      │               │               ├── 94002347.PDF
      │               │               ├── .........
      │               │               └── CHECKMD5.FIL
      │               │
      │               ├── V0195I03\ ──┬── .......
      │               │
      │               └── ........\ ──┬── .......
      │
      ├── 09258388\ ──┬── EA940053\ ──┬── .......
      │               │
      │               ├── EA940054\ ──┬── .......
      │               │
      │               └── ........\ ──┬── .......
      │
      ├── ........\ ──┬── ........\
      │
      ├── DATASET.TOC
      ├── 00406090.GIF
      ├── 09258388.GIF
      └── ......
```

The *DATASET.TOC* file reflects the directory structure depicted above and has a subdivision as follows:

```
_t0 all data on the complete Dataset
_t1 . all data on a specific journal title
_t2 .. all data on a specific journal issue within _t1
_t3 ... the first editorial item within the issue
_t3 ... the second editorial item within the issue
_t2 .. the second journal issue
_t3 ... the first editorial item within this journal issue
_t1 . another journal title with its _t2 and _t3 fields
_t2 ..
_t3 ...
```

In the following chapters each “level” of *DATASET.TOC* will be worked out in detail.

3.1 The “Dataset”-level (_t0)

Each Dataset will be identified by a unique identifier. This Dataset Identifier will appear under the tag **_t0** in the file *DATASET.TOC*, which is present in the root-directory of each Dataset. For instance, the thirteenth Dataset of the customer with code “EA” will be identified by the following:

```
_t0 EA000013
_vn 4.0
_pd 19951231
```

Description of the field (*{Mandatory}* denotes a mandatory field):

_t0 *{Mandatory}* The Dataset Identifier, an alphanumeric field of fixed 8 positions which contains a customer code or publishing centre code, and a sequence code. Only alphabetic (uppercase A—Z) and numeric (0—9) characters are present. Each publishing centre or customer will have a unique code and will be assigned its own sequence numbering. This sequence number does *not* have to be in a chronological order (e.g. more recent Datasets could have lower sequence numbers than earlier produced Datasets) and there *might* be gaps in the numbering scheme. It is not required that the sequence numbering is numerical, hexidecimal or alphabetic numbering methods are allowed. The length of both customer / publishing centre code and sequence code is free with the restriction that the total length of both codes should be eight positions. The Dataset Identifier will be unique. Examples are: **EA000123**, **TUP000A7**, **CUSTM009**.

_vn *{Mandatory}* The Version Number of the particular Dataset. It is expected that the standard format of the Datasets will evolve over time with future requirements. In order to be able to process “old” Datasets, it is necessary to identify the version of the standard with which the Dataset is written. The Version Number will be correlated to the version number of the cover page of this manual. Please retain preceding versions of this Technical Specifications in your archive for future reference.

_pd *{Mandatory}* The production date of the Dataset. This is a number in the format *YYYYMMDD*, where *YYYY* denotes the year, *MM* the month (**01**=January, **02**=February, **03**=March, ..., **12**=December) and *DD* the day in the month. It is possible that a production time is included, in which case the number has the format *YYYYMMDDhhmm*, in which *hh* indicates the hour in 24-hour format ranging from **00** to **23**, and *mm* denotes the minute within the hour, ranging from **00** to **59**. Examples:

- **_pd 19950801** The Dataset has been generated on August 1, 1995
- **_pd 199611302200** The Dataset has been generated on November 30, 1996 at 10 PM precisely.

_si Status indicator / message at Dataset level. In some projects it is necessary to specify a certain status or action to an entire Dataset or only to certain parts, e.g. to replace older material in case of errors or to indicate that material is derived from some unfinished stage of production. The **_si** field could be found on each level (Dataset, Journal Title, Journal Issue and Editorial Item), depending on the nature of a certain status or message. It is imperative to check every incoming Dataset for the occurrence of the **_si** field. Please read also the additional information regarding **_si** on pages 15, 18 and 22.

The first code in **_si** specifies the status indicator between square brackets " **[]** ". Please note, that status indicator codes are dependent of the particular project for which they are used. Each project in which **_si** is used can have its own set of status indicator codes per level. Valid codes should be communicated separately from these Technical Specifications within the confines of a project.

The remainder of the field can be used for additional data, a message or a comment, depending on the status indicator defined for the particular project.

Some examples for use of **_si** at the Dataset level:

○ **_si [DELETE] _t3 EA000013 00406090 V0193I03 8900432S** This example indicates that the editorial item identified by **_t3** should be removed from your collection. This code can only be applied if it is certain that Datasets are always enrolled in the same chronological sequence in which they were created. In this case the information provided in the **_t3** field is not related to *this* Dataset, but to an older one.

○ **_si [REPLACE] _t0 CUMUL002** This example is applicable in situations in which Datasets are generated cumulative, i.e. every new Dataset has new data, but also holds all material from previous Datasets, similar to publishing a new edition/version of a reference work.

3.2 The “Journal Title”-level (_t1)

One **_t1** entry appears in *DATASET.TOC* for each journal title from which one or more journal issues occur in the particular Dataset. A fictional example:

```
_t1 EA000013 00406090
_jn Thin Solid Films
_pu Elsevier Science S.A.
_ci Lausanne, Switzerland
_im Elsevier
_et Editor-in-Chief
_em Prof. J.E. Greene, Urbana IL, USA
_et Editorial Board
_em C.J. Adkins, Cambridge, UK
_em L.N. Aleksandrov, Novosibirsk, Russia
_em D.E. Aspnes, Red Bank NJ, USA
_em ....
_ia *Types of Contribution* - Original papers
    not previously published - Review articles -
    Letters, 600 - 800 words - Announcements, reports
    on conferences, news
_ia *Submission of Papers* Three copies of letters
    or full papers should be sent to: ......
_cv \00406090.GIF
_cr Elsevier Science S.A.: All rights reserved. No part of
    this publication may be reproduced, stored .... etc.

_t2 etc.
...........................................................
_t1 EA000013 09258388
_jn Journal of Alloys and Compounds
_jo Journal of the Less Common Metals
_io 00123456
_pu Elsevier Science B.V.
_ci Amsterdam, The Netherlands
_im Elsevier
_et Editor-in-Chief
_em Ch. J. Raub, Schwaebisch Gmuend, Germany
_et Editors
_em H.F. Franzen, Ames, IA, USA
_em K.H.J. Buschow, Waalre, The Netherlands
_et Honorary Editor
_em J.W. Christian, Oxford, UK
_et Editorial Advisory Board
_em G.-Y. Adachi, Osaka, Japan
_em A.V. Andreev, Ekaterinburg, Russia
_em ................
_ia *Types of Contribution* ........
_cv \09258388.GIF
_cr Elsevier Science B.V.: All rights reserved. .... etc.

_t2 etc.
......
```

Description of the fields (*{Mandatory}* denotes a mandatory field, if non-mandatory fields are empty, then they are not included at all; *{Repeating}* indicates a field which could appear more than once for the item):

`_t1` *{Mandatory}* Two strings of 8 positions, divided by a space, denoting the Dataset Identifier (see also **`_t0`** on page 12), followed by the International Standard Serial Number (ISSN), without the dividing dash.

`_si` Status indicator / message at Journal Title level. Page 12 provides an introduction of **`_si`** while additional information can be found on pages 18 and 22. An example for use of **`_si`** at this level:

○ **`_si [NEWJOURNALTITLE]`** This is the first occurrence of this ISSN or journal title in the project, because it is added to the collection or because there is a change in the ISSN (indicated in the **`_io`** field described below).

`_jn` *{Mandatory}* The full name of the journal.

`_jo` The former journal name. If a journal title changes its name (in the above example the *Journal of the Less-Common Metals* changed its name into *Journal of Alloys and Compounds*), then the **`_jo`** field will present the old name. The **`_jo`** field will appear at least for one year after the date of name change.

`_io` The former journal ISSN without the dividing dash. If the journal's ISSN changes, usually due to a change in the journal name or due to another editorial decision, then the **`_io`** field will offer the old ISSN. The **`_io`** field will appear at least for one year after the date of ISSN change.

`_jf` The full set journal name. Sometimes the particular journal title is a subordinate part of a so-called larger journal set. In this case the "parent" journal name is given here.

`_if` The full set journal ISSN without the dividing dash. If the full journal set name is presented in the **`_jf`** field, the **`_if`** field holds the ISSN of the full set.

`_pu` The publisher (e.g. Elsevier Science).

`_ci` The publisher's city and country or full address.

`_im` The imprint of the particular journal title (e.g. Pergamon, Elsevier, North-Holland). An imprint in the publishing industry is comparable with brand names in other branches of industries (e.g. General Motors produces motorcars with brand names such as Cadillac, Chevrolet, Opel).

`_et` *{Repeating}* Together with the **`_em`** field it constitutes the Editorial Board members and their titles. The **`_et`** field will specify the heading under which the members are grouped. It is followed by one or more **`_em`** fields.

`_em` *{Repeating}* The name and address (at least city) of a particular Editorial Board Member. The **`_em`** field is always preceded by a **`_et`** field.

`_ia` *{Repeating}* The Instructions to Authors specification used by a particular journal title for its manuscripts.

`_cv` The colour front cover page of a journal. The cover page constitutes an essential part of the "identity" of that journal. Such a cover page (if available) is added in a "thumbnail" colour image file. These files could be displayed on a computer screen in an interactive Article Delivery system or printed as part of a cover letter with each printout. The **_cv** tag is followed by the full specification of the colour image file. These file normally appear in the root of the Dataset e.g. T:\00406090.GIF. Those GIF tagged files are colour CompuServe **G**raphics **I**nterchange **F**iles, a popular format for colour images.

_cr The copyright notice of a journal.

3.3 *The “Journal Issue”-level (_t2)*

The several journal issues which are present in the Dataset are available within subdirectories under the Journal Title level. Issue Identifiers are unique within the ISSN. One physical journal issue is completely available in one Dataset. No “spanning” of issues over several Datasets occurs. An example of a fictional **_t2** part of the *DATASET.TOC* file:

```
_t2 EA000013 00406090 V0193I01   journal issue identifier *)
_vl 193                         Volume 193
_is 1-2                         Issue numbers 1 to 2
_pr 501-786                     the page range as it appeared on the spine
_dt 19941215                    December 15th, 1994
_np 300                         the physical number of pages
_pn nil nil nil nil i   ii  iii iv  v   vi  vii viii
    501 502 503 504 505 506 507 508 509 510 ....
    ... ... 783 784 785 786 nil nil
_ct 300 299                     the table of contents was printed
                                on the back cover and continued on
                                the inner back cover

_t3 ....                        The first item of this journal issue

..........................................................
_t2 EA000013 00406090 V0201X02   another issue in this Dataset
_vl 201-202                     Combined volume 201 and 202
_pr 309-512                     the page range as it appeared on the spine
_xt Supplement 2                extra information
_dt 199423                      Autumn, 1994
_np 208                         the physical number of pages
_pn nil  nil  nil  nil
    L309 L310 L311 L312 L313 L314 315 316
    .... .... 509  510  511  512  nil nil
_ct 3 4

_t3 etc.
......
```

**) Note: text in italic only for clarity*

Description of the fields (*{Mandatory}* denotes a mandatory field; if non-mandatory fields are empty, then they are not included at all):

`_t2` *{Mandatory}* Three strings of 8 positions, divided by spaces, denoting

- the Dataset Identifier (see also **`_t0`** on page 12),
- the ISSN of the journal (see also **`_t1`** on page 15) and
- the Issue Identifier of the specified journal issue.

The Issue Identifier consists of one string of 8 positions indicating the physical journal issue. Only alphabetic (uppercase A—Z) and numeric (0—9) characters are present. This string is unique within the ISSN to identify unambiguously the journal issue. There is no strict correlation with actual volume, issue or publication date information. The following possibilities exist which are used to construct Issue Identifiers:

- If journal issues are identified by official volume, issue information or publication dates, then the following setup is used:
    - The first five positions are the volume number, padded with the character "**`V`**" and zeroes (for instance *Volume 51* appears as **`V0051`**). If the physical journal issue involves more volumes (e.g. *Volume 101—103*) only the first one is taken (e.g. **`V0101`**). Please note the difference with the **`_vl`** field below. Some journals use a scheme based on publication year. In this case the year is taken, preceded by the character "**`V`**" (e.g. **`V1995`**).
    - The last three positions is the issue number of the journal issue, padded with the character "**`I`**" and zeroes (e.g. *Issue 23* appears as **`I23`**). If the physical journal issue involves more issues (e.g. *Issue 2—3*) only the first one is taken (e.g. **`I02`**). Please note the difference with the **`_is`** field below. If no issue number is given at all, "**`I00`**" is used. But if the physical journal issue is for instance *Part A* or *Supplement 2* of a multi volume publication (see also the **`_xt`** field below), then the constructions **`X0A`** or **`X02`** are used respectively.

  Examples: **`V0123I02`**, **`V1994I12`**, **`V0345X0B`**.
- For a number of journal titles it is not possible to use the above mentioned Issue Identifier construction method, because of historical/compatibility reasons or because there is another scheme in use than the volume/issue model. For those the following method is used. Issue Identifiers start with two characters specifying the publishing centre (in the example **`EA`**), followed by the year in which the journal issue has appeared or is scheduled to appear (2 positions, e.g. **`94`**, see also the **`_dt`** field) and 4 positions with a sequence number within the year with leading zeroes (e.g. **`0043`**). This sequence number is a unique index number and it has no correlation with actual volume, issue or publication date information. The sequence number also does *not* have a chronological order within the year (e.g. more recent journal issues could even have lower sequence numbers than earlier appearing journal issues) and there *might* be gaps in the numbering scheme. Within one journal title this number is unique. Examples: **`EA860123`**, **`QW910009`**.
- For special projects unfinalized editorial material is included that has not yet been assigned to a particular journal issue. In such cases the special code **`UNASSIGN`** will be used. In the **`UNASSIGN`** subdirectory editorial items are present which will eventually reappear in complete, finalized format in later Datasets. It should be

clearly stated within the confines of the particular project whether such unfinalized material could be expected.

_si Status indicator/message at Journal Issue level. Page 12 provides an introduction on **_si** while additional information can be found on pages 15 and 22. Some examples for use of **_si** at this level:

- ○ **_si [REPLACE] _t2 EA000011 00406090 V0201I02** This indicates a replacement of an erroneous journal issue which appeared in a previous Dataset. In this situation the Issue Identifier and the Item Identifiers are *identical* to the incorrect ones. So there is an *intentional* duplicate journal Issue Identifier and there are *intentional* duplicate Item Identifiers. If you “skip” the Dataset number you simply overwrite/replace the files in the old directory. If you don’t do this you should take note and make the old journal issue invalid. If you are not inputting Datasets in chronological order because of e.g. a backlog situation, you should take extra notice to not overwrite correct data by older incorrect material.
- ○ **_si [S200] 19951120** The journal issue is not yet finished entirely, but is in final proofing stage (it has passed step “200” in the production cycle) and the included material should be regarded as a “Preview” version, which will be replaced by a final version in a later Dataset. The final version of this issue is scheduled to be completed no later than November 20, 1995. It is beyond the scope of this Technical Specifications if and how claiming of “overdue” material should be performed.

_vl *{Mandatory}* The volume number(s) as it appears on the cover page or spine of the journal issue. If a physical journal issue involves more volumes, then the volume numbers are separated by a dash “ **–** ” (e.g. *Volume 5 and 6* = **_vl 5-6**). Non-arabic numbering will be converted to the arabic numbering scheme (e.g. *Heft Drei und Vier* = **_vl 3-4**). If a range longer than 2 volumes is given, only the first and last numbers appear (e.g. *Volume 3 up until and including 6* = **_vl 3-6**). Some journals use a scheme based on publication year. In this case the year is taken (e.g. all issues of *1995* will appear with **_vl 1995**).

_is The issue number(s) as it appears on the cover page or spine of the journal issue. The rules follow the same as described for the **_vl** field. If no issue information is given, then the **_is** field does not appear. This mostly occurs in multi-volume journal issues such as special Proceedings issues. Examples:

- ○ *Issue 52* = **_is 52**
- ○ *Issues 5, 6 and 7* translates to **_is 5-7**

_pr The starting and ending page numbers, such as appearing on the cover page or spine, divided by a dash “ **–** ”. If there is more than one range on the cover page than the ranges are separated by a plus sign “ **+** ”. Examples:

- ○ *Pages 2345 up until and including 2478* are translated to **_pr 2345-2478**
- ○ *Page ranges i to xii and 1 to 250* translate to **_pr i-xii+1-250**

_cf This field contains specific details in the case the particular physical issue is devoted on publishing the proceedings of a conference, symposium or another event. The **_cf** tag is followed by a code enclosed in square brackets characterizing the individual details. Valid codes are:

[name] The full conference name.

[abbrev] The officially abbreviation of the conference name (if any).

[number] The conference number if the conference is part of a repeating series of conferences.

[place] The conference place (city, region, etc.).

[date] The date of the conference in the format *YYYYMMDD*, in which *YYYY* denotes the year, *MM* the month and *DD* the day in the month. See also the description of the **_dt** field below for date ranges.

[editor] *{Repeating}* The editor(s) of the conference proceedings. This field is repeated as often as there are editors for the issue.

An example:

```
_cf [name] Third Conference on the Flowering of Tulips - 1995
_cf [abbrev] TULIP ’95
_cf [number] 3
_cf [place] Amsterdam, The Netherlands
_cf [date] 19950828/0901
_cf [editor] Bulb, P.G.M.
_cf [editor] Leaf, M.M.
```

_xt Extra information for the particular journal issue that appears on the cover page or spine. This usually is the case with special issues (other than conference proceedings) such as supplements, indexes, memorials, multipart volumes, etc. Examples:

- **_xt Supplement 23**
- **_xt Index 99**
- **_xt Dedicated to the Commemoration of ....**
- **_xt Part C**

_dt *{Mandatory}* The date field contains the issue date of the journal issue. It is converted to a number in the format *YYYYMMDD*, in which *YYYY* denotes the year, *MM* the month, season or quarter and *DD* the day in the month.

Date-conventions for *MM* are:

- **01**=January, **02**=February, **03**=March, ..., **12**=December,
- **21**=Spring, **22**=Summer, **23**=Autumn, **24**=Winter,
- **31**=1st Quarter, **32**=2nd Quarter, **33**=3rd Quarter, **34**=4th Quarter.

If a range of dates occurs, then only the start and end dates are taken, separated by a forward slash “ **/** ”, omitting duplicate elements.

Examples:

- **_dt 19861215** = The journal issue dated *December 15, 1986*.
- **_dt 199004** = The *April 1990* issue.
- **_dt 1988/1989** = The journal issue containing the years *1988 and 1989*.
- **_dt 199009/11** = The *September - November 1990* issue.
- **_dt 19951030/1103** = The issue for the week *October 30 - November 3, 1995*.
- **_dt 199412/199502** = The issue covering the period *December '94 - February '95*
- **_dt 195523** = The *Autumn 1955* issue.

_np *{Mandatory}* The total number of physical pages the journal issue including the covers.

_pn *{Mandatory}* To facilitate browsing, the actual printed page numbers of the journal issue, in the order in which they appear in the journal issue, are included, separated by spaces. This is different from the **_pr** field in that a typical journal issue contains front and back matter which does not follow normal numbering schemes. A journal issue has a front cover and a back cover (both double-sided, but generally not numbered), parts that are numbered with roman numerals (iii, xxii, etc.), pages that have no numbers at all (such as advertisements) and pages with regular numbering schemes. The parts of fold-out pages appear as separate images. If the numbering scheme of the journal issue is discontinued at a fold-out page then the naming scheme should reflect this discontinuation [1].

In the fictious example

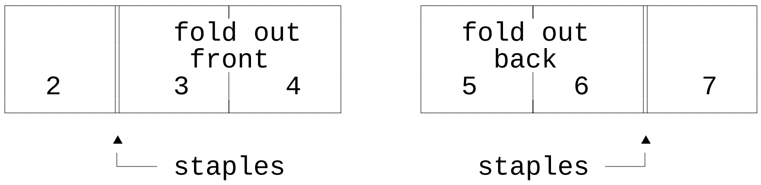
```
_np 20
_pn nil nil i ii iii iv v vi L101 L102 L103 L104
    105 106 107 108 nil nil nil nil
```

it means that this 20 page journal issue contains

- two unnumbered front cover pages (outside and inside, 2 x **nil**),
- six roman numbered pages (pages **i** — **vi**),
- four pages with a special meaning (in this case for a particular journal, e.g. the *Journal of Molecular Catalysis*, the *Letters to the Editor* section is numbered with an "L" printed together with the page number; pages **L101** — **L104**),
- four regular numbered pages (**105** — **108**),
- two plain unnumbered pages (2 x **nil**) and
- the inside and outside back covers (also 2 x **nil**).

There are precisely as much pages as the **_np** field states.

[1] Elsevier does actually number all fold-out pages. Example:



_ct The pages in the journal issue on which the table of contents was printed, relative to the physical page sequence, in which the front cover page is 1. The page numbers are in the range described in the **_pn** field.

Examples:

- **_ct 3 4** means that the table of contents appeared on the third and fourth physical page of the journal issue.
- **_ct 300 299** (in combination with **_np 300**) means that the table of contents started at the outside of the physical back cover and was continued at the inside of the back cover.

3.4 The “Editorial Item”-level (_t3)

A fictional example of the **_t3** part of the *DATASET.TOC* file could look like:

```
_t3 EA000013 00406090 V0193I01 94000123
_ii 0040-6090(94)00012-3        the item identifier *)
_ty FLA                         the item type
_li EN                          the language of the full item
_ti Growth of epitaxial thin films in the KTiOPO"4 family of
    crystals                    the title
_au Cheng, L.K.                 the authors
_au Bierlein, J.D.
_au Foris, C.M.
_au Ballman, A.A.
                                the correspondence address
_ca Prof. C.M. Foris, CR&D Department, E.I. du Punt de
    Nemours & Co, Experimental Station, P.O. Box 80306,
    Wilmington, Delaware 19880-0306, USA
_ab We report the growth of thin epitaxial films in an
    environment ........        the abstract
_la EN                          the language of the abstract
_kw thin films                  the keywords the authors supplied
_kw crystals
_pg 501-504+520                 the range of pages as they
                                appear in a citation
_mf [Raw ASCII] 101 102 103 104 120    the raw ASCII file names
_mf [TIFF 5.0] 101 102 103 104 120     the TIFF file names
_mf [SGML Cit] 94000123                the citation in SGML format

_t3 EA000013 00406090 EA940053 94005646
_ii 0040-6090(94)00564-6
_ty ......
......
```

**) Note: text in italic only for clarity*

Description of the fields (*{Mandatory}* denotes a mandatory field; if non-mandatory fields are empty, then they are not included at all; *{Repeating}* denotes a field which could appear more than once for the item):

_t3 *{Mandatory}* Four strings of 8 positions, divided by spaces, denoting

- the Dataset Identifier (see also **_t0** on page 12),
- the ISSN of the journal (see also **_t1** on page 15),
- the Issue Identifier (see also **_t2** on page 17) and
- the identifier of the specified item.

The combination of these four numbers is unique. This Item Identifier is a unique index code and it has *no* correlation with the order in which the items appear in the journal issue and there *might* be gaps in the numbering scheme. Within one journal title this Item Identifier is unique.

Some examples:

- **_t3 EA000013 00406090 V0193I03 8900432S**
- **_t3 TUP0003A 09258388 EA940054 9107945H**

_si Status indicator / message at Editorial Item level. Page 12 provides an introduction of **_si** while additional information can be found on pages 15 and 18. Some examples for use of **_si** at this level:

- **_si [EXPIRE] 19950801** This editorial item is of temporary nature, e.g. an announcement for an event. Its expiration date is August 1, 1995, after which it is no longer valid or appropriate.
- **_si [S050] 19960101** This particular editorial item is not yet completely finalized (it has passed step “50” in the production sequence) and is still subject to change, but it is provided for use in an “early warning” service. The final version of this item is scheduled to be completed no later than January 1, 1996. Please note the possible relationship with the **UNASSIGN** code in the **_t2** field (see also page 17).

_ii *{Mandatory, Repeating}* Item Identifier. There are two non exclusive possibilities for this field, the unspecified (default) version and the specified (tagged) version.

- The unspecified Item Identifier in its explicit punctuated format is included for convenience. Item identifiers according to the scheme as it is used in the **_t3** field (by taking the second and fourth field) were developed by Adonis for use in the Adonis system, in which they are known as Adonis-identifiers. Elsevier Science adopted this scheme for identifying editorial items in publications and refers to them as to the *Standard Serials Document Identifier* (SSDI). From 1993 onward Item Identifiers appear in small print at the first page of Elsevier Science’s editorial items. The SSDI’s are useable in two formats, an implicit representation for use in computer applications and an explicit representation when the SSDI is visually presented on paper or other media that must be read by humans. The 16 digit implicit code, consists of the following parts:
  - The ISSN without the hyphen separator (8 positions),
  - The last two digits of the year of actual receipt of the item by the publisher **or** the year of receipt of this item by the scanning bureau (could be different from the year of publication), e.g. **89**, **92**, **95**, etc.

  - A unique identification code of five numeric positions within ISSN and year assigned by the publisher or by the scanning bureau, e.g. **00123**, **80045**, **78006**, **90001**, etc. and
  - Lastly a check digit (one position). The C-source of the program to calculate this checkdigit is available for download from Elsevier's Internet FTP Server "FTP.Elsevier.NL" or upon request.

For example, the identifier

```
           Year  Checkdigit
            ┌┐     ┬
    0040609094000123
    └──────┘  └───┘
      ISSN    Number
```

represents an item received in 1994 for publication in the journal *Thin Solid Films* (ISSN = 0040-6090).

The explicit representation is derived from the implicit one by adding punctuation (dashes and parentheses). These additional elements should not form part of a machine readable string, they should be stripped out for machine readable use and/or reinstated on printing. The explicit format is optional, not mandatory.

```
           Year  Checkdigit
            ┌┐       ┬
  0040-6090(94)00012-3
  └───────┘    └───┘
    ISSN       Number
```

Hence:

  - The ISSN with the dividing dash,
  - The year between parentheses,
  - The Unique code within ISSN and year, and
  - a dash followed by the check digit.

- An identification scheme for digital (electronic) use which is based on printed source items (serials, books, reports) has limitations. It is recognized that at present documents do not stand alone, but are a subset of a publication type (e.g. articles are part of a journal, chapters are part of a book). This relationship may not always hold in the future. *EFFECT* has been defined as a generic "envelope" format for digital electronic information. Therefore other publication item types and their associated identifiers should be supported. Examples include:
  - In 1995 a collaboration to adopt a common document identifier has been inititiated by the American Chemical Society (ACS), the American Institute of Physics (AIP), the American Physical Society (APS), the Institute of Electrical and Electronics Engineers (IEEE) and Elsevier Science. The working name for this identifier is Publisher Item Identifier (PII), which is similar, but not identical to the Adonis identifiers and SSDI's since it also incorporates books.
  - Serials Item and Contribution Identifier (SICI), also referred to as the National Information Standards Organization (NISO) Z39.56 code. It was previously known as the SISAC code.

- Universal Resource Name (URN) on the Internet (generalizing from Universal Resource Locators — URL's).
- Universal Data Identifier (UDID).
- International Standard Book Number (ISBN).
- International Standard Music Number (ISMN).
- International Standard Work Code (ISWC).
- Proprietary identification schemes.

Different identification schemes are supported by the inclusion of the scheme name within square brackets " **[]** " after the **_ii** tag, followed by the identifier. Because it is possible that a particular item is supported in different schemes (e.g. journal items are supported in SSDI/PII as well as in SICI), the **_ii** field is repeated for every identification scheme. Some examples:

- **_ii [PII] S 1054-139X(96)01024-5**
- **_ii [SICI] 0040-6090(199501/03)175:1L.29:MEDL;1-**
- **_ii [URN] <urn:resolver:physics.elsevier.nl:report1234>**

_ty *{Mandatory}* The type of the item. Valid type-codes are:

ABS ABStract only: Abstract of a paper or oral presentation, published as a separate item.

ADD ADDendum: Publication item giving additional information regarding another publication item.

BRV Book ReView: Review of one or more books. See also *PRV*.

COR CORrespondence: Letter to the editor or reply to letter.

DIS DIScussion: Argumentative communication. May be a perspective, commentary, discussion, etc.

EDI EDItorial: Note (in the general meaning of the word) by the Editor of the publication, usually of adstructive nature.

ERR ERRatum: Message reporting errors for which the publisher was responsible in items which were previously published in the same journal.

FLA Full Length Article: Complete report on original research, containing sections on, e.g., Methods, Results, Discussion, References.

PRV Product ReView: Review of one or more products other than books, e.g. software. See also *BRV*.

REV REView article: Substantial overview of original research, usually with a comprehensive bibliography.

SCO Short COmmunication: Short report or announcement on research, usually claiming certain results, with a short publication time compared to other papers in the same publication. Other names for "short communications" are letter papers, preliminary notes, notes, etc.

SSU Short SUrvey: Short, mini- or microreview (perspective) of original research published elsewhere, which may reflect personal opinion or experience.

MIS MIScellaneous: All editorial items which don't fit in any of the item types mentioned and which don't merit introduction of a new item type. Examples of these are Announcements, News sections, Obituaries and Calendars.

Note that not all editorial material is "itemized". For instance, material of volatile character or non-scientific nature will not be tagged separately. Examples of such items are Cover pages, Advertisements, Acknowledgments, Author and Subject Indexes, Bulletins, Dedications, Diaries and Quizzes. Pages which do not contain “real” editorial items are still available for browsing, although they are not cross-indexed.

A special note is given on the Table of Contents. Except from the **_ct** field (see also page 21), which points to the page images on which the Table of Contents of the particular Journal issue was printed, no item for the text version of the Table of Contents is directly available. This is disregarded because of the following reasons:

- ○ The text version can easily be generated by taking the fields **_ti**, **_au** and **_pg** from every editorial item. In this way you could work out your own standard layout to present to your end users.
- ○ Not every single editorial item (especially the small ones such as errata, product reviews, editorial items) is mentioned in the journal’s printed Table of Contents.
- ○ Not every journal issue contains a Table of Contents. Especially smaller journal issues sometimes have only one article, in which it is considered useless to print a Table of Contents.
- ○ Journal issues have typographic errors in their printed Table of Contents. It is better to rely on the “real” page indicators than on these mentioned in the printed Table of Contents.

_li *{Mandatory}* The language of the item. Valid codes are: **EN**=English, **FR**=French, **DE**=German (Deutsch), **RU**=Russian, **ES**=Spanish (Español), **PT**=Portugese.

_ti The full title of the item in the english language.

_tf The foreign title of the item. If the original title in the printed issue appears in a foreign language, then the original foreign language title appears in the **_tf** field. If available, the translated title in english appears in the **_ti** field. Example: The item with a title in french “*Un Essai au sujet de fleurir des tulipes*” would result in:

```
_li FR
_ti An Essay about the flowering of tulips
_tf Un Essai au sujet de fleurir des tulipes
```

_au *{Repeating}* The author(s) of the item. This field is repeated as often as there are authors in the item, such as is shown in the example.

_ca The full address for correspondence with the author(s).

_ab The full english abstract of the item.

_la *{Mandatory, Repeating}* The language of the abstract. If more than one abstract is present, the **_la** code is repeated for every abstract. Note however that only the english abstract (if present) appears in the **_ab** field. If no abstract is present, then **_la** is absent. Valid codes are the same as those mentioned with code **_li**.

_kw *{Repeating}* The keyword(s) which apply to the item. This field is repeated as often as there are keywords for the item, as is shown in the above example.

The **_kw** field is available in two different formats, non-specific and controlled-dictionary.

- The non-specific format specifies the keywords the author has supplied with the original item. This is the default format for this field, in which the **_kw** tag is followed by a keyword. Example: the author supplied the term *crystals:* **_kw crystals**
- The controlled-dictionary format has a code between square brackets after the **_kw** tag, which identifies the particular dictionary, thesaurus or classification scheme which holds the term. Allowed dictionary codes will be separately arranged. Example: the publisher associates the term *crystallography* from the controlled dictionary *PhysicsThesaurus* with this particular article:
  _kw [PhysicsThesaurus] crystallography

_pg The pages on which the item physically appeared in the journal issue, as it would be printed in a citation. Page ranges are divided by a dash, discontinued pages are separated by a plus sign " + ". If an item does not have an actual page number then this **_pg** field does not appear. Some examples:

- The item is spread across the pages 2, 3, 4, 7, 8 and 20 = **_pg 2-4+7-8+20**
- The editorial appears on pages iii, iv, v and vi = **_pg iii-vi**
- A letter to the editor appears on pages L20, L21 and L22 = **_pg L20-L22**

_br *{Repeating}* The item has an identified backward reference to another item (not necessarily falling within the scope of journal article types in this Specification). For instance:

- An erratum points to the original article.
- A product review refers to the evaluated book, software, motion picture, etc.
- A correspondence paper, normally a reply to a letter, mentions that letter.
- Review or full length articles cite several publications.

The **_br** tag is followed by the item identifier of that backward referenced item and follows the same rules as described in the **_ii** format including the possibility to reference other specified identification schemes (see page 22). This field is repeated as often as there are identified backward references keywords for the item. The receiving system at your organization should be able to use the backward reference data for amending the original item information with a forward reference (e.g. an article should reference forward to errata, which appeared in a later journal issue).

An example of an erratum referring back to the example on page 21:

```
_t3 EA000016 00406090 V0194I03 95000569
_ii 0040-6090(95)00056-9
_ty ERR
_ti Erratum on @'Growth of epitaxial thin films in the
    KTiOPO"4 family of crystals@' by Cheng, L.K., Bierlein,
    J.D. et. al.
_pg 655
_br 0040-6090(94)00012-3
_mf [Raw ASCII] 103
_mf [TIFF 5.0] 103
```

`_mf` *{Repeating}* The manifestation file information which forms the item. All material (bibliographic material, text, images, etc.) from a particular journal issue and its editorial items is held in a single directory plus possible subdirectories. The **`_mf`** field provides cross references between (logical) editorial items and the (physical) files which hold the actual data. Within an issue directory, file names follow regular MS-DOS conventions. Such a file name is restricted to an 8 character name part and a 3 character extension part, divided by a period (e.g. *nnnnnnnn.eee*).

- ○ The 3 character extension of the file denotes the file-type.
  - ■ *RAW* stands for a "raw", non-edited ASCII-file.
  - ■ *TIF* means the file is a black / white image file in Tagged Image File Format (TIFF) 5.0 format with the ITU T.6 compression scheme, formerly known as CCITT Fax Group IV
  - ■ *SGM* applies to a file holding the full text of an editorial item in Standard Generalized Markup Language (SGML) format.
  - ■ *SGC* applies to a file which only contains bibliographic data in SGML format.
  - ■ *PDF* addresses a file in Portable Document Format (PDF), such as used in Adobe Acrobat^{TM}.
  - ■ *JPG* constitutes a greyscale or colour file in the Joint Photographers Expert Group (JPEG) lossy compression format.
  - ■ *EPS* applies to an Encapsulated PostScript (EPS) file.
- ○ The name parts of the TIF and RAW files describe the physical page numbers, in which the front cover is page 1. Multi-paged, item-based files, such as SGML or PDF files, have names which correspond with their unique item identifier (see also **`_ii`** on page 22).

Examples:

- ○ File *242.TIF* corresponds with a TIFF page image of the 242nd physical page in the issue (relative to the front cover which is the first page).
- ○ *242.RAW* contains the full, unedited "raw" text of this page.
- ○ File *94000123.PDF* matches the Portable Document Format version of a particular item

An item can have zero or more manifestation fields. Each manifestation has its own **`_mf`** entry. Each **`_mf`** entry is divided into separate fields. If the **`_mf`** field is not available for a particular journal title or item, then your organization has no subscription on this journal or the particular manifestation is not available.

Explanation of the fields:

- ○ The first field is the descriptive name of the manifestation. These names appear between square brackets. The options are:

`[Raw ASCII]` Followed by the ASCII files with the full, unedited text of the item (result of Optical Character Recognition). The extension .*RAW* should be added to these files.

Example: **`_mf [Raw ASCII] 101 102 103 104 120`** means that the five files 101.RAW, 102.RAW, 103.RAW, 104.RAW and 120.RAW contain the raw text of the editorial item.

[TIFF 5.0] Followed by the image files which denote all page images on which the particular item was printed. The extension *.TIF* should be added to these files.

Example: **_mf [TIFF 5.0] 101 102 103 104 120** means that the five image files 101.TIF, 102.TIF 103.TIF, 104.TIF and 120.TIF constitute the page images of the item. Please note the possibility of gaps such as between pages 104 and 120. In this case the article is spread across a non-contiguous page range, normally with visible clues such as *"Continued on page ..."*.

[SGML Cit] Followed by the file name which denotes the file which holds the bibliographic data (only title, author(s), abstract, etc.) of the editorial item in Standard Generalized Markup Language *(SGML)* format. The extension **.SGC** should be added to this file. The Document Type Definition *(DTD)* which corresponds with the SGML format is available separately via Elsevier's Internet FTP server "FTP.Elsevier.NL" or upon request.

Example: **_mf [SGML Cit] 9412345X** means that the file *9412345X.SGC* contains the SGML formatted bibliographic data of the item.

[SGML] Followed by a directory name which denotes the subdirectory which holds the full text of the editorial item in Standard Generalized Markup Language *(SGML)* format and all accompanying artwork files. In this subdirectory —one level below the **_t3** level directory— a file *MAIN.SGM* is available. This file holds the full text of the editorial item. All other files in this directory are referenced in the *MAIN.SGM* file. These are usually artwork files, holding the illustrations, figures and photographs for the editorial item.

Example: **_mf [SGML] 9412345X** means that the file *MAIN.SGM* in subdirectory *9412345X* contains the SGML formatted text of the item. In this file itself, reference is made to three artwork files *FIG1.JPG*, *FIG2.TIF* and *FIG3.EPS*, also present in subdirectory *9412345X*.

[PDF] Followed by the file name which denotes the file in Portable Document Format *(PDF)*, such as defined by Adobe Corp. in its Acrobat^{TM} line of products, which constitutes the full text of the editorial item. The extension **.PDF** should be added to this file.

Example: **_mf [PDF] 9412345X** means that the file *9412345X.PDF* contains the Portable Document Format version of the editorial item.

○ The subsequent parts are the full file names without extensions, which form the manifestation of the item.

4 Dataset Delivery

Datasets will arrive at your organization in the medium of your choice: CD-ROM, tape or Internet FTP. It is not relevant for the standard format described earlier which media are applied. It is easily possible to switch to another medium without changing the standard. The layout of a Dataset could be represented as a standard MS-DOS or UNIX directory tree.

A note on file names: throughout this Technical Specifications manual file names are displayed in full capitals. Because of file name restrictions instigated by MS-DOS, actual file names in Datasets appear in some UNIX systems in all lowercase. File names in Datasets are case insensitive. The file names *"DATASET.TOC"*, *"dataset.toc"* and *"DaTaSeT.tOc"* all refer to the same file.

The following assumptions have been taken into account:

- The emphasis is on archiving and in transporting large quantities of information.
- The different parts of the information should be easily discernable to be processed in a batch oriented environment in which internal database files will be created.
- The data should be parsed and indexed by the receiving organizations, to be included in internal database and/or image bank systems.
- The receiving organizations make available the data to their own staff using their own hardware/software infrastructure.

4.1 Delivery through CD-ROM

The structure and format of a Dataset on CD-ROM fully complies with the ISO 9660 Mode 1 format, similar to regular MS-DOS conventions for file names and directory structures. The CD-ROMs will be produced by means of a socalled CD-Recordable (CD-R) device, a write-once CD-player in which single or small sets of standard CD-ROMs are produced. Any CD-ROM player will be able to read CD-ROMs produced by such a machine. Fetching material from a CD-ROM involves simply copying from it.

Requirement:

- ○ Standard CD-ROM drive conforming to ISO 9660 Mode 1, using the Microsoft CD Extensions (MSCDEX) or equivalent.

Appendix A Greek and mathematical codes

Code		Symbol	
@C	=	Γ	GAMMA
@D	=	Δ	DELTA
@F	=	Φ	PHI
@J	=	Ψ	PSI
@L	=	Λ	LAMBDA
@P	=	Π	PI
@Q	=	Θ	THETA
@S	=	Σ	SIGMA
@U	=	Υ	UPSILON
@W	=	Ω	OMEGA
@X	=	Ξ	XI
@a	=	α	alpha
@b	=	βϐ	beta
@c	=	γ	gamma
@d	=	δ	delta
@e	=	εε	epsilon
@f	=	ϕφ	phi
@g	=	χ	chi
@h	=	η	eta
@i	=	ι	iota
@j	=	ψ	psi
@k	=	κϰ	kappa
@l	=	λ	lambda
@m	=	μ	mu
@n	=	ν	nu
@p	=	π	pi
@q	=	θϑ	theta
@r	=	ρϱ	rho
@s	=	σς	sigma
@t	=	τ	tau
@u	=	υ	upsilon
@w	=	ω	omega
@x	=	ξ	xi
@z	=	ζ	zèta

Code		Symbol	
@6	=	↑	arrow up
@7	=	↓	arrow down
@[	=	→	arrow right
@]	=	←	arrow left
@#	=	⇆	left over right arrow
@<	=	≤	smaller or equal
@4	=	≰	not smaller or equal
@>	=	≥	larger or equal
@5	=	≱	not larger or equal
@=	=	≠	not equal
@8	=	≡	identical
@K	=	≃	approx. equal
@O	=	~	similar, varies linearly with
@/	=	√	square root
@!	=	∫	integral
@A	=	∅	circle with slash, diameter
@+	=	±	plus minus
@1	=	¼	one quarter
@2	=	½	one half
@3	=	¾	three quarter
@&	=	∞	infinite
@%	=	‰	promilla
@9	=	Å	Ångström
@$	=	£	English pound
@M	=	♂	male
@V	=	♀	female
@*	=	°	degree
@’	=	"	double quote
@@	=	@	at-symbol

@? = any character not in this list

Appendix B Codes in DATASET.TOC

	_ab	Abstract of Item	25
	_au	Author Name	25
	_br	Backward Reference	26
	_ca	Correspondence for Item	25
▌	**_cf**	Conference details	19
	_ci	City of Publisher	15
▌	**_cr**	Copyright Notice	16
	_ct	Content Pages of Journal Issue	21
	_cv	Cover Page of Journal Title	15
	_dt	Date of Publication of Journal Issue	19
	_em	Editorial Board Member name	15
	_et	Editorial Board Title	15
	_ia	Instructions for Authors	15
▌	**_if**	ISSN of full set	15
	_ii	Item Identifier	22
	_im	Publisher Imprint	15
▌	**_io**	Former Journal ISSN	15
	_is	Issue Number	18
▌	**_jf**	Full set journal name	15
	_jn	Journal Name	15
	_jo	Former Journal Name	15
	_kw	Keywords of Item	26
	_la	Language of the Abstract of the Item	25
	_li	Language of the Item text	25
	_mf	Manifestation Files of Item	27
	_np	Number of Pages of Journal Issue	20
	_pd	Production date of Dataset	12
	_pg	Page Range of Item	26
	_pn	Actual Page Names of Journal Issue	20
	_pr	Page range of Journal Issue	18
	_pu	Publisher name	15
▌	**_si**	Status indicator/message	12, 15, 18, 22
	_t0	Identification of Dataset	12
	_t1	Identification of Journal Title	15
	_t2	Identification of Journal Issue	17
	_t3	Identification of Item	22
	_tf	Title of Item in Foreign language	25
	_ti	Title of Item	25
	_ty	Type of Item	24
	_vl	Volume Number of Journal Issue	18
	_vn	Version Number of Dataset	12
	_xt	Extra information to identify Journal Issue	19