

---

# EFFECT

## Exchange Format For Electronic Components and Texts

Technical Specifications  
Version 4.1 — September 1996

Author Paul Mostert  
Elsevier Science

---

Copyright © 1996, Elsevier Science. All rights reserved.

Copies for internal use only of these EFFECT Technical Specifications may be made by organisations collaborating with Elsevier Science or by vendors or other third parties with whom they are working. All other parties are allowed to make a single personal copy for evaluation purposes only. All copies must include this copyright statement in full.

Use of this documentation is free for organisations directly collaborating with Elsevier Science. All other parties are not allowed to apply these specifications without permission by Elsevier Science. Permission will usually be granted for non-commercial purposes provided that its use is identified in all material mentioning or referencing this documentation as the "Elsevier Science EFFECT Technical Specifications".

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.

These EFFECT Technical Specifications document may not be sold or reproduced in any other form, including electronic, or distributed in any other way without permission.

---

# Table of contents

<b>1</b>	<b>Introduction</b>	3
1.1	Conventions used in this manual	3
1.2	Overview of changes in EFFECT Specifications Version 4.1	4
1.3	EFFECT on the World Wide Web	5
<b>2</b>	<b>Dataset components</b>	6
2.1	Tagged Image File Format (TIFF) page image files	6
2.2	Raw text files	7
2.3	Standard Generalized Markup Language (SGML) files	7
2.4	Hypertext Markup Language (HTML) files	8
2.5	Artwork files (TIFF, JPEG, GIF, EPS)	9
2.6	Portable Document Format (PDF) files	9
2.7	Other files	10
<b>3</b>	<b>Dataset arrangement</b>	11
3.1	Directory structure	11
3.1.1	Issue based datasets	12
3.1.2	Item based datasets	13
3.2	dataset.toc	14
3.3	The repository directory	14
3.4	checkmd5.fil	15
<b>4</b>	<b>The <i>dataset.toc</i> file</b>	16
4.1	General notes	16
4.2	Hierarchy of dataset.toc	18
4.3	The "Dataset"-level (_t0)	19
4.4	The "Journal Title"-level (_t1)	21
4.5	The "Journal Issue"-level (_t2)	23
4.6	The "Publication Item"-level (_t3)	29
4.6.1	Manifestations in issue-based datasets	37
4.6.2	Manifestations in item-based datasets	38
<b>5</b>	<b>Dataset delivery</b>	43
5.1	Dataset delivery by CD-ROM	43
5.2	Dataset delivery by file transfer	44
<b>Appendix A</b>	<b>Greek and mathematical codes</b>	45
<b>Appendix B</b>	<b>Codes in dataset.toc in alphabetical order</b>	46
<b>Appendix C</b>	<b>Codes in dataset.toc in order of appearance</b>	48

---

# 1 Introduction

**E**FFECT - EXCHANGE FORMAT FOR ELECTRONIC COMPONENTS AND TEXTS is a standard to enable large-scale deliveries of electronic files. This standard has been developed by Elsevier Science to support comprehensive electronic journal/article distribution from production systems at the publisher to distribution servers either at local libraries or at a remote host organization. The EFFECT standard describes how large amounts of electronic files can be structured and encased in so-called *datasets*, and how the “packing list” (the file *dataset.toc*), which comes with the dataset, is structured.

The standard was developed in the course of the TULIP project, a five-year research project (1991-1995) on digital libraries by Elsevier Science and nine major universities in the USA. The material provided by Elsevier was used to create local current awareness and article delivery database systems.

This EFFECT Technical Specifications document is subdivided in the following chapters

- An overview of all possible technical components that can be embedded in EFFECT datasets.
- The overall directory structure of datasets.
- A comprehensive description of the *dataset.toc* file.
- How the material of the journals can be delivered.

## 1.1 Conventions used in this manual

Throughout this manual the following conventions are used:

- Text appearing within files is presented with a **non-proportional typeface**.
- Directory and file names are displayed in *bold-italic*.
- The EFFECT file/directory naming scheme is based on the ISO 9660 Mode 1 standard for CD-ROM's, which has been adopted because this is supported by virtually all computer platforms (MS-DOS, Apple Macintosh, OS/2, most UNIX-systems, etc.). This standard limits file- and directory names to a subset of MS-DOS conventions, i.e. file names of eight positions, a single dot and an extension of three positions, e.g. *filename.ext*. MS-DOS makes no distinction between uppercase and lowercase characters. MS-DOS file names appear in uppercase (e.g. *FILENAME.EXT*). In other computer platforms, ISO 9660 Mode 1 file names appear in all lowercase (e.g. *filename.ext*). Therefore, directory and file names in EFFECT datasets should be regarded as case insensitive. For example, the file names *FILENAME.EXT*, *filename.ext* and *FiLeNaMe.ExT* all refer to the same

file. In this manual and EFFECT datasets from version 4.1 and up, lowercase is used for directory and file names.

In some cases (typically non-MS-DOS) longer file names or file names with a distinction between uppercase and lowercase are required. In these cases the UNIX *tar* file format is used to collect those files together. This is noted where applicable.

- In these Technical Specifications the UNIX forward slash “ / ” convention for separating directory and file names is used. In MS-DOS applications this should be replaced by a backslash “ \ ” as the separator in path names.

## 1.2 Overview of changes in EFFECT Specifications Version 4.1

For clarity and readability the order of the chapters has been changed. We feel that this increased the useability of the specifications. There is no indication toward the original order of the chapters in previous versions.

The structural differences with EFFECT Technical Specifications version 4.0 are indicated by a vertical bar in the left margin (like this paragraph). Minor textual changes or corrections are not highlighted.

- Use of lowercase for directory and file names
- Extension of directory structure (see page 11)
- Introduction of new tags
  - **\_st** Structure Type (see page 20)
  - **\_cd** CODEN code (see page 22)
  - **\_ja** Journal abbreviation (see page 22)
  - **\_su** Subject area (see page 35)
- Removal of existing tags (because they were confusing)
  - **\_si** Status indicator
  - **\_cv** Cover page
- Corrections/Update of current tags
  - **\_ii** A change of the Item Identifier definition (see page 30)
  - **\_ty** Several new publication types added (see page 31)
  - **\_mf** Several changes to Manifestation Format (see pages 27 and 36)

### **1.3** *EFFECT on the World Wide Web*

More current information on the EFFECT standard, links to related documentation, and a collection of tools is available via the Elsevier Science home page at <http://www.elsevier.nl>, or its mirror <http://www.elsevier.com>, in the “General Information” section or directly by specifying <http://www.elsevier.nl/locate/effect>.

---

## 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 other order according to the particular arrangement between Elsevier and your organization.

In this chapter technical information is provided regarding the different components of EFFECT datasets such as page images, raw texts, SGML files, etc. Further in this manual is described how these components interrelate.

Please note that not every file format mentioned in this chapter appears in a particular dataset. For instance, dependent for a specific project TIFF page image files can be included, whereas another project expects electronic page equivalents in Portable Document Format (PDF).

### 2.1 Tagged Image File Format (TIFF) page image files

Page images occurring in EFFECT datasets are produced as follows: 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 image file per scanned page. Page images are standard black/white single-page Tagged Image File Format (TIFF) files with a scan resolution of 300 dots per inch (dpi). The file name extension for TIFF files in EFFECT datasets is *.tif*. The maximum scan size is European A4, i.e. 210 x 297 mm<sup>2</sup>.

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 with white background and black characters. Decompression could be performed in software, for instance with the programs HiJaak Pro, Image Alchemy, Kofax 910 software, Adobe Photoshop, 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 Pentium PC, range from 5 seconds to several minutes, whereas hardware decompression times range from tenths of seconds.

An average scientific publication 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*

Optionally, every page image file has a corresponding “raw” ASCII text file, indicated in EFFECT datasets by the file name extension *.raw*. These text files are produced as a result of Optical Character Recognition (OCR) procedures. The files are further called “raw” since no keyboarding/editing/spellchecking 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** *Standard Generalized Markup Language (SGML) files*

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

SGML is an international standard published by the International Standards Organisation (ISO) for the coding and presentation of text which allows the storage, transmission, display and editing of material through a descriptive environment. Rather than indicating in detail 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 formats and output media.

SGML documents conform to a Document Type Definition (DTD), a description of the structure of a particular type of publication, such as a full-length article. The DTD includes all the possible elements that such a document could contain, together with the hierarchical relationships between those elements and indications of optionality and repeatability.

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

EFFECT datasets can contain two different types of SGML files. These files can be recognized by their extensions:

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

Please refer to the EFFECT World Wide Web page for more information on SGML and Elsevier endorsed DTD's (see page 5).

## **2.4** *Hypertext Markup Language (HTML) files*

Hypertext Markup Language (HTML) is a standard for encoding information for the World Wide Web (WWW). This standard started as the vehicle for an internal information system in 1990-1991 at the Nuclear Physics Laboratories (CERN) in Geneva, Switzerland.

It became successful since the National Center of Supercomputer Applications (NCSA) of the University of Illinois created the popular graphical user interface Mosaic. Since then this encoding scheme is conquering the world, aided by the advent of popular World Wide Web browsers for most prevalent computer platforms such as NCSA's *Mosaic*, Netscape Communications *Navigator*, Microsofts *Internet Explorer*, and many others.

HTML has roots in SGML. It has a relative simple DTD, but deviates from the SGML concept in that it is more geared toward presentation than to structure.

HTML files in EFFECT datasets –indicated by the file name extension *.htm*– represent either full publication items or parts thereof. Opposed to publication items presented by a single SGML file, publication items in HTML are in most cases represented by a number of HTML files, each holding a separate part of the publication item, for instance, the citation with title, authors and abstract, each (larger) figure or table embedded in an HTML file with the annotation, the different chapters, etc.

Mostly HTML files are combined with small *.gif* files holding all elements for which there is no HTML equivalent (such as mathematic symbols and formulae) and larger *.gif* and *.jpg* files holding the figures.

## 2.5 Artwork files (TIFF, JPEG, GIF, EPS)

A large proportion of scientific texts is supplemented by artwork (figures, plates, maps, photographs, etc). There is a wide variation in graphics formats to represent different types of artwork. The most popular formats occurring in EFFECT datasets:

- Tagged Image File Format (TIFF) ITU T.6/Fax Group IV encoded files are utilized for black/white (monochrome) line art such as diagrams and drawings. This format is identical to the one used for TIFF page images (see above). TIFF files appear usually with the extension *.tif*
- Joint Photographer Expert Group (JPEG) encoded files are typically applied for grayscale and color artwork files such as photographs. This format has an excellent ratio of image quality versus image size. In EFFECT datasets, JPEG files are identified by the extension *.jpg*
- CompuServe Graphics Interchange Format (GIF) compressed files (represented with the *.gif* extension) are a popular format in the Internet and supported as a native format in all graphical World Wide Web browsers. GIF files have certain restrictions with regard to for instance color depth (no more than 256 colors possible) and are generally larger in storage size than the same images in the above-mentioned file formats.
- Encapsulated PostScript (EPS; file name extension is *.eps*) is a multipurpose file format, which is primarily used in cases where resolution independent vector-oriented graphics (e.g. diagrams) is applied or where bitmap parts are supplemented with vector elements (e.g. photographs with text and arrow annotations). Bitmap elements in EPS files are not compressed (TIFF and JPEG are better suited for this), therefore this format is mostly applied with vector-graphics elements only.

## 2.6 Portable Document Format (PDF) files

The Portable Document Format<sup>™</sup> (PDF) is a file format used to hold a document in its original presentation, independent of the application software, hardware, and operating system used to create it. The file name extension for PDF files in EFFECT datasets is *.pdf*

PDF was developed by Adobe Systems Incorporated specifically to aid in the transfer of documents across platforms. PDF enables users to exchange and view electronic documents independent of the environment in which they were created. PDF relies on the imaging model of the PostScript<sup>™</sup> language to describe text and graphics in a device- and resolution-independent manner. PDF defines more structure than used by most PostScript language programs. It also includes objects such as hypertext links and annotations, which are not part of the pages itself, but are useful for interactive viewing. PDF supports standard compression filters to help reduce file size for images, text, and graphics.

Adobe has created the *Acrobat*<sup>TM</sup> suite of programs, which are cross-platform tools for electronic publishing. With Acrobat it is possible to create (with *Distiller*, *Capture* or *Writer*), view and navigate (*Reader* or *Exchange*) documents in electronic form, print them, or distribute them electronically to various computer platforms. PDF files will retain with the same look and feel of the original document. *Acrobat* is device-independent, and is available for MS Windows (3.1x, 95 and NT), Apple Macintosh, OS/2, UNIX and MS-DOS platforms.

The Portable Document Format is an open standard available for third party software developers. *Adobe* has published the format and provides Application Program Interface (API) tools to assist developers.

## **2.7** *Other files*

The intention of the present Technical Specifications is to provide an open environment for electronic journal information. Therefore other file formats than the ones described above can be added easily as the need arises without introducing inconsistencies to this Technical Specifications. Supplier and receiver should agree on the particular file formats used before actual files are exchanged. See also the description of manifestation format at page 36.

Examples are:

- Audio/Video Interleave (AVI) multimedia files;
- Quicktime multimedia files;
- Computer Graphics Metafiles (CGM);
- WAV, MID or AU encoded sound files;
- Java applets;
- T<sub>E</sub>X (TEX) encoded files;
- etc.

---

## 3 Dataset arrangement

This chapter describes the directory arrangement of an EFFECT dataset and introduces the control and supporting components:

- Directory structure
- The *dataset.toc* file
- The *repostry* directory
- The *checkmd5.fil* files

### 3.1 Directory structure

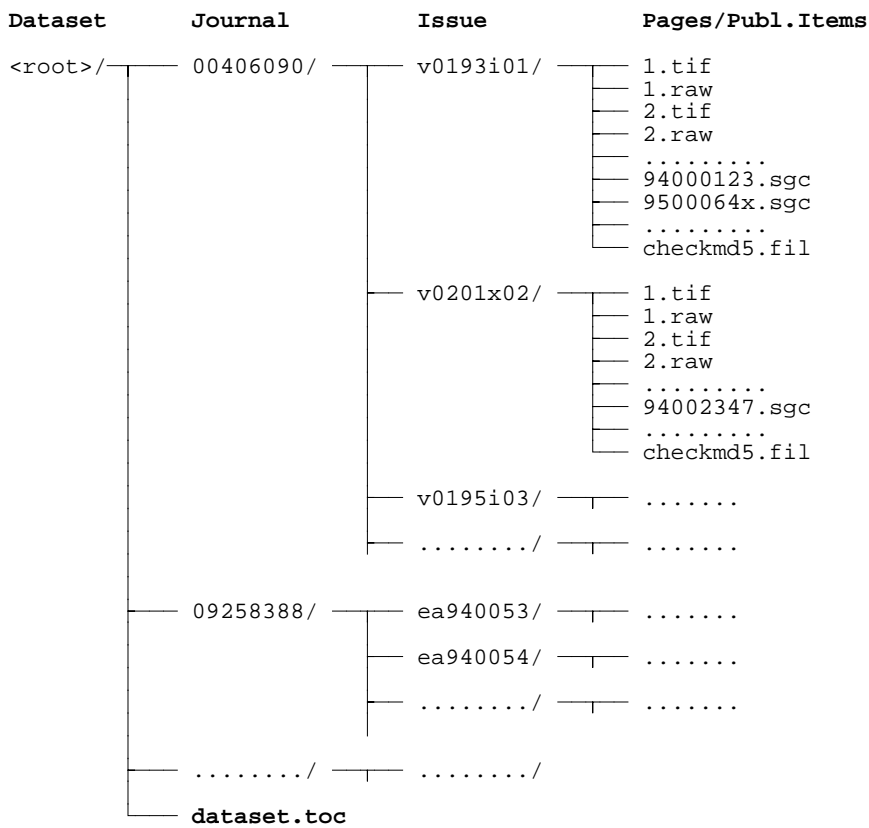
An EFFECT dataset has a directory structure which reflects the subdivision into journals, issues, publication items (articles) and pages. Two different schemes are possible:

- Issue based datasets, applied for after-print, scanned material (based on scanned page images). This directory structure has been utilized in previous editions of the EFFECT Technical Specifications.
- Item based datasets, utilized for pre-print, electronically prepared material (based on generic text in SGML or HTML) and artwork-files based systems. This is a new directory scheme, designed to deal with single article files, non- and multi-page files.

A dataset can exist in only one scheme. The *\_st* field in the *dataset.toc* file (see page 20) indicates the applied scheme for a particular dataset.

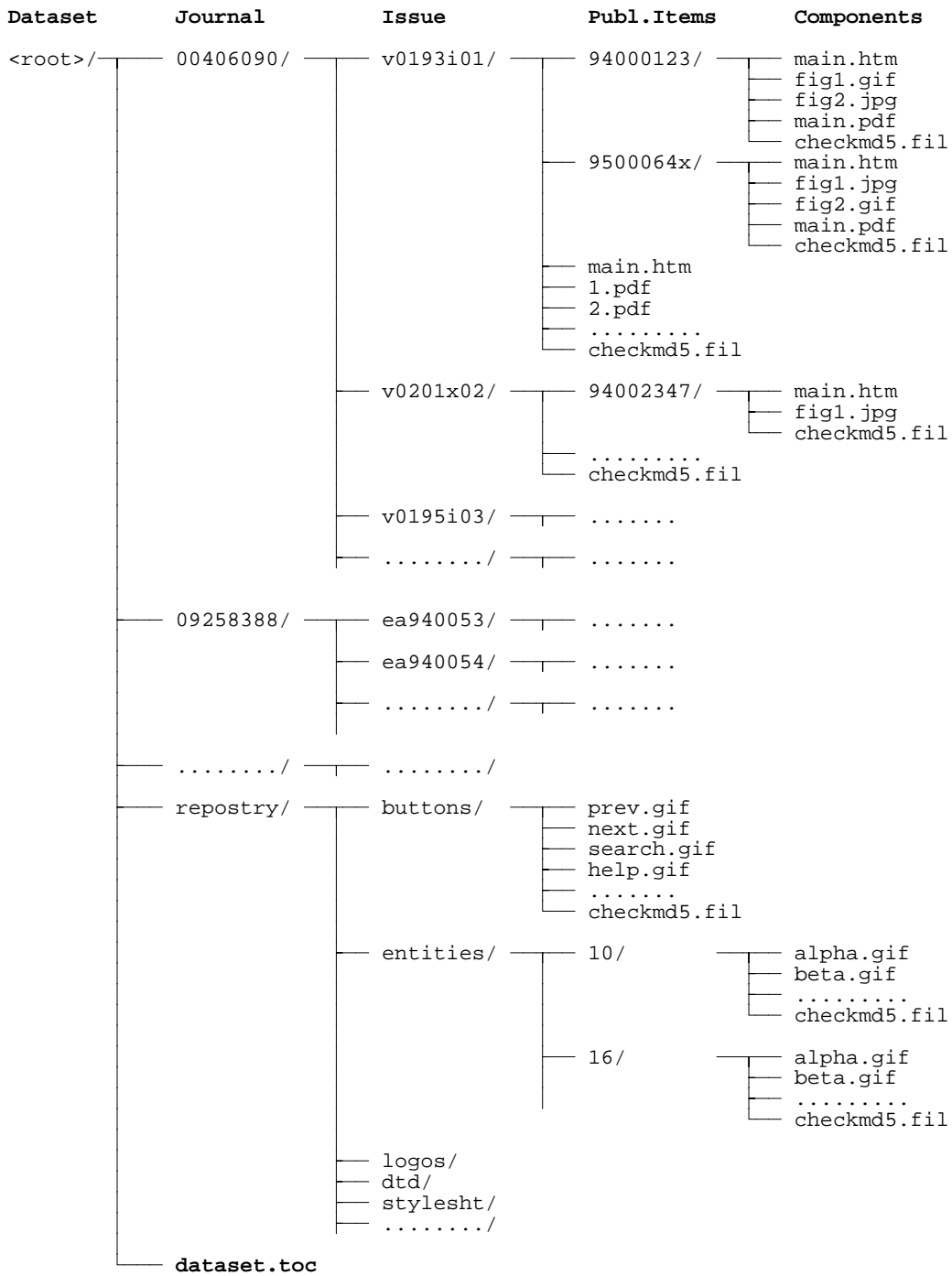
### 3.1.1 Issue based datasets

The structure can be represented in a semi-graphical fashion as follows:



### 3.1.2 Item based datasets

The structure can be represented in a semi-graphical fashion as follows:



## 3.2 *dataset.toc*

Each EFFECT dataset contains the file *dataset.toc* in the *<root>* directory in which all cross-indexing reference data is provided. It is the main entry-point to the dataset and contains all information on the actual datatransport itself (dates, origin, etc.) and/or information needed to reconstruct journal issues and publication items contained in them. Please refer to chapter 4 at page 16 for a full description.

## 3.3 *The repository directory*

In some cases EFFECT datasets carry a *repository* directory, directly available at the *<root>* level. This directory offers additional data to EFFECT datasets beyond the scope of a specific journal title. It can offer material to enhance the functionality of the accepting system and/or contains supporting artwork or other files for general use.

For example in an EFFECT dataset which is supplied for an HTML-based system, usually general artwork files exist with logos, navigational buttons (e.g. "Previous", "Next", "Help") or non-HTML-able symbols. HTML files refer to these symbols with constructs similar to

```
The symbol <IMG SRC="/repository/entities/10/alpha.gif"> shows the  
alpha in point size 10.
```

```
<IMG SRC="/repository/buttons/next.gif"> is the "Next" button.
```

If a *repository* directory is included, its contents should be copied to the receiving system in its entirety, preserving the directory structure. Possible earlier versions of *repository* may be overwritten. Note that the *repository* directory in the dataset works cumulative. Usually the *repository* contains additions and replacements. At the beginning of a project or in an early dataset the initial full *repository* directory is provided.

A special note: In some cases a file *repository.tar* is provided, instead of a *repository* directory, because the *repository* contains non-ISO 9660 compliant file names. To make this generic: "any" file with an extension of *.tar* should be considered as an implied directory structure. For instance, two different entities exist such as *Delta.gif* and *delta.gif* (for the greek delta in upper "Δ" and lowercase "δ"). It must be verified that the accepting system is a UNIX or equivalent file system which supports long file names and distinguishes between uppercase and lowercase characters.

### 3.4 *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 minimized. As a precautionary measure an additional checksum facility is available that could be used to verify the correctness of files.

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

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 pre-specified 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 cryptographic system 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 home page [www.rsa.com](http://www.rsa.com), reachable via the EFFECT Web page (see page 5).

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 and 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
```

Please note that file names are case insensitive. In some cases uppercased file names in *checkmd5.fil* should be translated to lowercase.

## 4 The *dataset.toc* file

Each EFFECT dataset contains the file *dataset.toc* in the *<root>* directory; in which all production data and cross-indexing reference data is provided. It is the main entry-point (or "packing list") to the dataset and contains all information needed to reconstruct journal issues and publication items included in the dataset.

Throughout these specifications, EFFECT datasets are regarded as a "logical" packages, not related to a particular physical medium with technical limitations, such as magnetic tape or CD-ROM. It is imaginable that a single EFFECT dataset is "spanned" across several CD-ROM's, or even that several datasets are delivered together in one physical medium. To avoid unnecessary confusion, use of EFFECT in this way is not recommended and should be considered with greatest caution.

The *dataset.toc* file contains some redundancy in order to have information easily readable and accessible to the human eye. Subsequent electronic processing could simply remove this redundant information.

### 4.1 General notes

General rules for the *dataset.toc* file are:

- The file contains only 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. For instance, *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  $\sqrt{x}$  appears as **@/x**. A real @ is coded as **@@**. See Appendix A on page 45 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 " " ", for instance,  $H_2O$  will appear as **H"2O** and  $I_{sum}$  as **I"s"u"m**. A real double quote is represented as **@'** and a real caret ^ appears as **@^**.
- Bold, italic and underlined characters will be converted to their normal representation. For instance **a**, *a*, a and  $\mathbf{a}$  are all converted to **a**.

- There is no limit to line lengths, but for human readability and compatibility with older ASCII editors, lines of text will in general not be longer than 80 positions. The rare exception is for very long words which may not be broken.
- The file *dataset.toc* is split up into records, which in turn are broken down into fields.
  - A new record starts on a new line with an underline “\_”, the character “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 alphanumeric characters and a space, followed by the contents of the field itself. Fields which are longer than 80 positions are wrapped to subsequent lines, so-called continuation lines.
  - A line starting with one or more spaces (preferred: four spaces) is a continuation of the previous line.
  - Some field tags may be followed by options or attributes. These are enclosed in square brackets “[ ]”. This use is explained in the sections describing the tags.
- Empty lines may be included for better human readability. These lines can be ignored for electronic processing.

Some tools are available in the EFFECT Web page (see page 5) to translate e.g. from @-codes and sub/superscripts to HTML equivalent codes.

## 4.2 Hierarchy of *dataset.toc*

The file *dataset.toc* reflects the directory structure depicted in the previous chapter, and has a subdivision as follows:

```
_t0 all data on the complete dataset
  . . . . .
_t1 the first journal title within the dataset
  . . . . .
_t2 the first journal issue within the first journal title
  . . . . .
_t3 the first publication item within the first journal issue
_t3 the second publication item within the first journal issue
_t2 the second journal issue within the first journal title
  . . . . .
_t3 the first publication item within the second journal issue
_t3 the second publication item within the second journal issue
    . . . . .
_t1 the second journal title within the dataset
  . . . . .
_t2 the first journal issue within the second journal title
  . . . . .
_t3 the first publication item within the first journal issue
    . . . . .
    etc
```

### 4.3 The “Dataset”-level (`_t0`)

Each dataset is identified by a unique identifier. This dataset identifier appears in the tag `_t0` in the file *dataset.toc*, which is present in the `<root>`-directory of each dataset. For instance, the twelfth dataset of the originator with code “xyz” is identified by the following:

```
_t0 xyz00012
_vn 4.1
_pd 19961231
_st ISS
```

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

- `_t0` *{Mandatory}* The dataset identifier, an alphanumeric field of fixed 8 positions which contains an originator code, and a sequence code. Only alphabetic (lowercase (a–z) or uppercase (A–Z); case insensitive) and numeric (0–9) characters are present. Each project gets a unique code and is 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, hexadecimal or alphabetic numbering methods are allowed. The length of both originator code and sequence code is free with the restriction that the total length of both codes should be eight positions. The dataset identifier should be unique for a certain project. 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 EFFECT 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 these 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 was generated on August 1, 1995
  - `_pd 199611302200` The dataset was generated on November 30, 1996 at 10 PM precisely.

- █ **\_st** Dataset Structure Type. This field indicates if this particular EFFECT dataset is structured according to items or issues (see also the explanation at page 11). Valid codes are:
- █ ISS Issue based datasets (default)
- █ ITM Item based datasets

## 4.4 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 fictitious example:

```
_t1 xyz00012 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: .....
_cr Elsevier Science S.A.: All rights reserved. No part of
this publication may be reproduced, stored .... etc.

_t2 etc.
.....
_t1 xyz00012 09258388
_jn Journal of Alloys and Compounds
_jo Journal of the Less Common Metals
_io 0012345x
_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* .....
_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, separated by a space, denoting the dataset identifier (see also **\_t0** on page 19), followed by the International Standard Serial Number (ISSN), without the dividing dash. ISSN's are numeric codes of seven positions. The last (eight) position is a check digit which could either be numeric or the character *x*.
- \_cd** The CODEN code, an alternative for the ISSN.
- \_jn** *{Mandatory}* The full name of the journal.
- \_ja** The officially abbreviated journal title. Especially for longer journal titles, there is a desire to abbreviate the name. E.g. the *Journal of Brain Research* is officially abbreviated to *J. Brain Res.* The **\_ja** field shows the preferred abbreviation of the publisher. It is only supplemental to the full journal name, provided in **\_jn**.
- \_jo** The former journal name. If a journal title changes its name (in the above example the *Journal of the Less-Common Metals* (with ISSN 0012-345X) changed its name into *Journal of Alloys and Compounds* with ISSN 0925-8388), then the **\_jo** field gives the old name. The **\_jo** field will appear for at least one year after the date of the 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 is related to the occurrence of the **\_jo** field and will also appear for at least 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 given 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 an **\_et** field.
- \_ia** *{Repeating}* The Instructions to Authors specification used by a particular journal title for its manuscripts.
- \_cr** The copyright notice of a journal.

## 4.5 The “Journal Issue”-level (*\_t2*)

The various journal issues which are present in the EFFECT dataset are available within subdirectories under the journal title (*\_t1*) level. It is mandatory that issue identifiers are unique within ISSN's. A physical journal issue is available in its entirety in one dataset, i.e. issues are not “spanned” across datasets.

An example of a fictitious *\_t2* part of a *dataset.toc* file:

```
_t2 xyz00012 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 19951215 December 15th, 1995
_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 xyz00012 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
_cf [name] Third Conference on the Flowering of Tulips - 1995
_cf [abbrev] TULIP '95 Conference information
_cf [number] 3
_cf [place] Amsterdam, The Netherlands
_cf [date] 19950828/0901
_cf [editor] Bulb, P.G.M.
_cf [editor] Leaf, M.M.
_xt Part 2 extra information
_dt 199523 Autumn, 1995
_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 each, separated with spaces, denoting
- the dataset identifier (see also **\_t0** on page 19),
  - the ISSN of the journal (see also **\_t1** on page 22) 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 (a–z) and numeric (0–9) characters are allowed. This string is unique within the ISSN to unambiguously identify 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 **\_v1** field below, in which the entire range is given. Some journals use a scheme based on publication year. In this case the year is taken, preceded by the character “v” (e.g. **v1996**).
  - The last three positions are for 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, in which the entire range is given. 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 a construction with the character “x” is used. For instance, **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 or compatibility reasons, or because there is another scheme in use than the volume/issue model. For those journal titles the following method is used. Issue identifiers start with a unique string of two characters, followed by the year in which the journal issue has appeared or is scheduled to appear (two positions, e.g. **94**, see also the **\_dt** field), and a sequence number within the year with leading zeroes (four positions, e.g. **0043**). This sequence number is a unique index number and 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 journal issues appearing earlier) and there *may* be gaps in the numbering scheme. Within one journal title this number is unique. Examples: **ea860123**, **qw910009**.

- Sometimes, 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** is used. In the *unassign* subdirectory publication items are present which will eventually reappear in complete, finished format in later datasets. It should be clearly stated within the confines of the particular project whether such unfinished material could be expected.

**\_v1** *{Mandatory}* The volume number(s) as it appears on the cover page or spine of the journal issue. If a physical journal issue involves several volumes, then the volume numbers are separated by a dash “ - ” (e.g. *Volume 5 and 6* = **\_v1 5-6**). Non-arabic numbering will be converted to arabic numbering (e.g. *Heft Drei und Vier* = **\_v1 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* = **\_v1 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 **\_v1 1995**).

**\_is** The issue number(s) as it appears on the cover page or spine of the journal issue. The rules are the same as for the **\_v1** 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 they appear on the cover page or spine, separated by a dash “ - ”. If there is more than one range on the cover page than the ranges are separated by a plus sign “ + ” or a comma “ , ”. Examples:

- *Pages 2345 upto 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** or **\_pr i-xii,1-250**

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

- [name]** *{Mandatory if there is a \_cf group for this journal issue}* The full name of the conference.
- [abbrev]** The official 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 location (city, region, etc.).
- [date]** *{Mandatory if there is a \_cf group for this journal issue}* 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 this issue.

See also the example at page 23.



In the fictitious 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 journal issue of twenty pages contains

- two unnumbered front cover pages (outside and inside, 2 x **nil**, which means “no printed page number”),
- six pages numbered in roman (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 regularly numbered pages (**105 – 108**),
- two plain unnumbered pages (2 x **nil**) and
- the inside and outside back covers (also 2 x **nil**).

There are exactly as many pages as the **\_np** field contains.

**\_ct** {Mandatory for issue structured datasets only (see **\_st ISS** at page 20) } The pages in the journal issue on which the table of contents is 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 **\_np** field.

Examples:

- **\_ct 3 4** means that the table of contents appears 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 starts at the outside of the physical back cover and is continued at the inside of the back cover.

**\_mf** Manifestation file information at journal issue level **\_t2**. See page 36 first for a detailed description of the **\_mf** field. The **\_mf** field at the journal issue level contains references to all files which cannot be “typified” (with **\_ty**; see page 31) as a formal publication item, but nevertheless are an integral part of a journal issue. The **\_mf** field provides cross links between all (logical) parts of an issue that cannot directly be related to a publication item and the (physical) files which hold the actual data. The related files are all stored in the appropriate directory at journal issue level.

The **\_mf** at the **\_t2** level is applied in the following cases:

- There is need to supply a “table of content”-like construct to enhance the access of the publication items contained at **\_t3** level. This could be done in World Wide Web-applications by for instance providing a table of content in HTML format. In this case the construct **\_mf [HTML version] main.htm** can be applied. Contrary to the use of **\_mf** in **\_t3** (see page 36 and further), the *main.htm* file is one position higher in the directory hierarchy. For instance:  
*/xyz00012/00406090/v0193i01/main.htm*

- In all cases where material is supported which is not part of publication item (such as advertisements, cover pages, bulletins, etc.) and their only reference is the page number on which they appear in the paper journal issue. In case the item level is the printed page, the file name is the physical page number. Please note that page number 1 is the front of the cover. Also note that page number zero (0) has been reserved for the spine (back) of the issue.

An example: Both

- `_mf [PDF 1.1 WRAPPED] 1 2 3 4 5 6` and
  - `_mf [PDF 1.1 WRAPPED] 1.pdf 2.pdf 3.pdf 4.pdf 5.pdf 6.pdf`
- mean that the issue directory mentioned in `_t2`, contains the files *1.pdf*, *2.pdf*, *3.pdf*, *4.pdf*, *5.pdf* and *6.pdf*

## 4.6 The "Publication Item"-level (\_t3)

A fictitious example of the \_t3 part of the *dataset.toc* file could look like this:

```
_t3 xyz00012 00406090 v0193i01 96000123
_ii S0040-6090(96)00456-X      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    the raw ASCII file names
  104 120
_mf [TIFF 5.0] 101 102 103    the TIFF file names
  104 120
_mf [SGML Cit] 94000123       the citation in SGML format

_t3 xyz00012 00406090 ea940053 96005646
_ii S0040-6090(96)00789-0
_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, separated by spaces, and denoting
- the dataset identifier (see also **\_t0** on page 19),
  - the ISSN of the journal (see also **\_t1** on page 22),
  - the issue identifier (see also **\_t2** on page 24) and
  - the internal EFFECT identifier of the specific publication item.

The combination of these four numbers is unique. The item identifier is a unique index code and has *no* correlation with the order in which the items appear in the journal issue; there *may* be gaps in the numbering scheme. Within one journal title this item identifier is unique.

Some examples:

- **\_t3 xyz00012 00406090 v0193i03 8900432s**
- **\_t3 tup00030 09258388 ea940054 9107945h**

- \_i** *{Mandatory, Repeating}* The official Publisher Item Identifier (PII) or other generally accepted article identification schemes.

In 1995 a collaboration to adopt a common document identifier has been initiated 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 the *Publication Item Identifier* (PII), which originated from earlier *Adonis number* and Elsevier Science's *Standard Serials Document Identifier* (SSDI) standards. Those earlier standards were restricted to serials items. The PII standard is extended to incorporate books, electronic media and other publication formats. From 1996 onward, item identifiers appear in small print on the first page of the articles published by the affiliated partners.

PII's are applied in two formats:

- An implicit (shortened) representation for use in computer applications. For instance, **s0040609096000123** represents an item received in 1996 for publication in the journal *Thin Solid Films* (ISSN = 0040-6090).
- An explicit (verbose) representation for printing on paper or other media that must be read by humans. The explicit representation is derived from the implicit one by adding punctuation (dashes and parentheses). For instance, the previous example in its explicit form is **s0040-6090(94)00012-3** These additional elements should not form part of a machine-readable string; instead they should be stripped out for machine-readable use and/or reinstated on printing. The explicit format is not mandatory, but optional.

Both PII formats could be applied in EFFECT datasets. There is no specific preference to use one or the other. More information on the Publisher Item Identifier (PII) and links to related standards can be found via the EFFECT Web page (see page 5).

The `_ii` field is not restricted to the PII identification scheme. Other item identification schemes are formulated. Examples include:

- The 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. The default is `_ii [PII]` Since it is possible that a particular item is supported in different schemes (e.g. journal items are supported in PII as well as in SICI), the `_ii` field is repeated for every identification scheme. Some examples:

- `_ii [PII] S1054-139X(96)01024-5` (the default)
- `_ii [SICI] 0040-6090(199501/03)175:1L.29:MEDL;1-`
- `_ii [URN] <urn:resolver:physics.elsevier.nl:report1234>`

A special note:

Previous versions of the EFFECT standard prescribed that the `_ii` field and the combined second and fourth field of `_t3` were equivalent. This requirement has been elevated because of the following reasons:

- The `_t3` field offers the directory names in which material is held. Therefore `_t3` is restricted by the ISO 9660 name scheme, in which directory names have a maximum length of eight positions and can only contain alphabetic (case insensitive) and numeric characters. As is shown in the above examples, generally accepted item identifier schemes mostly accommodate more than 16 positions.
- In paper publications, publication items appear only once and are typically directly linked to the journal title in which they appeared. Translated in EFFECT terminology: publication items are linked to the ISSN indicated in the `_t1` field. However, when electronic publishing emerged, items were linked to paper *and* electronic publications, each potentially with different ISSN's. Also, republishing of publication items in another journal or as part of a compilation or a handbook could create clutter, because the ISSN originally was an integral part of the item identifier.

`_ty` {Mandatory} The type of the publication 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.
- ADV** ADvertisement: Commercial statement, with or without illustration.

- ANN** ANNouncement: Informative statement with a scope within the context of the publication in which it appears.
- BRV** Book ReView: Review of one or more books. See also *PRV*.
- CAL** Meetings CALENDAR: Chronologically ordered list of forthcoming meetings relevant to the publication in which the list is published.
- 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 illustrative 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.
- LIT** LIterature alert: Publication item containing information on relevant literature.
- NWS** NeWS: Publication item containing new information relevant to the audience of the publication..
- OCN** Other CoNtent: Contents list of related journal.
- PNT** PateNT Report: Report on newly developed patents.
- PRP** Personal RePort: Item reporting news about one or more persons, e.g. obituaries and anniversaries; also including personal historical overviews, reminiscences, etc.
- PRV** Product ReView: Review of one or more products other than books, e.g. software. See also *BRV*.
- PUB** PUBlisher's note: Message from the Publisher to the readers of a publication.
- REQ** REQuest for assistance: Publication item containing the description of a problem together with an appeal to the audience for the submission of a solution.
- 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 publication items which don't fit in any of the item types mentioned and which don't merit introduction of a new item type.

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, acknowledgments, bulletins, dedications, diaries and quizzes. In issue-based datasets, pages which do not contain "real" publication items are still available for browsing, although they are not cross-indexed.

A special case is formed by the table of contents and indexes. Except from the `_ct` field (see also page 27), 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. Also indexes (subject, author, etc.) are not supported. This was done for the following reasons:

- The text version of the table of contents can easily be generated by taking the fields `_su`, `_ti`, `_au` and `_pg` for every publication item. Similarly, author indexes can be produced by selecting `_au` and `_pg` fields, and subject indexes from `_su`, `_kw` and `_pg` tags. In this way it is possible to create standard layouts to present to end-users.
- Not every single publication item (especially the small ones such as errata, product reviews, publication 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 and indexes. It is better to rely on the "real" page indicators than on these mentioned in the printed table of contents and indexes.

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

`_ti` The full title of the item in English, if available.

`_tf` The foreign language title of the item in the `_li` language. 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. Please note that if the publisher didn't translate the foreign language title into English, a `_tf` field exists without an accompanying `_ti` field.

Example: The French item with the title "*Un essai au sujet de fleurir des tulipes*", together with the translated title 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` *{Repeating}* The full English abstract of the item.

**\_la** {Mandatory for each **\_ab** code, Repeating} The language of the abstract. If more than one abstract is present, the **\_la** code is available for every abstract. If no abstract is present, then **\_la** is absent. Valid language codes are the same as those mentioned with code **\_li** (see above).

```
_ab This is the english abstract
_la EN
_ab Ceci c'est le sommaire francaise
_la FR
```

**\_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**

**\_su** The SUBJECT area of the publication item. This field is especially used for larger journal titles covering broader areas. The **\_su** field is used to group related publication items together in automatically generated overviews such as sections in tables of contents, or hitlists.

An example of its use: Consider a journal issue which carries three articles. Two articles fall in subject area *General aspects*, the other one in subject area *Specific aspects*. Selected parts of *dataset.toc* are as follows:

```
_t3 xyz00012 00406090 v0193i01 96000001
_ti Article One
_su General aspects

_t3 xyz00012 00406090 v0193i01 96000002
_ti Article Two
_su General aspects

_t3 xyz00012 00406090 v0193i01 96000003
_ti Article Three
_su Specific aspects
```

A formatted table of contents could be presented as follows:

```
General aspects
● Article One
● Article Two

Specific aspects
● Article three
```

**\_pg** The printed page numbers on which the item 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 “ + ” or a comma “ , ”. 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** or alternatively **\_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 30). This field is repeated as often as there are identified backward references for the item. The receiving system should be able to use the backward reference data for adding a forward reference to the original item information (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 29:

```
_t3 ea000016 00406090 v0194i03 96000034
_ii S0040-6090(96)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 S0040-6090(96)00012-3
_mf [Raw ASCII] 103
_mf [TIFF 5.0] 103
```

**\_mf** *{Repeating}* Manifestation format. A manifestation is a particular instance of a publication item for a certain application, which often is composed as a group of related files. The **\_mf** tag is directly followed by an identifier indicating the specific manifestation format, appearing within square brackets “[ ]”.

A distinction exists between issue- and item-based EFFECT datasets (see also page 11):

- In issue-based datasets (identified by “**\_st ISS**”, see page 20) all material (bibliographic data, raw text, page images, etc.) from a particular journal issue and its publication items is held in a single (issue) directory.
- Item-based datasets (with “**\_st ITM**”) issues and their publication items are subdivided into several subdirectories within the journal issue directory.

Dependent of the value given in the **\_st** field, the **\_mf** fields should be interpreted differently. This is worked out in further detail in this section.

#### 4.6.1 Manifestations in issue-based datasets

Issue-based datasets include the following manifestation formats:

##### [TIFF 5.0]

Followed by the image files which denote all page images on which the particular item was printed. If not supplied explicitly, the extension *.tif* should be added to obtain the full file names.

Example: Both

- `_mf [TIFF 5.0] 101 102 103 104 120` and
- `_mf [TIFF 5.0] 101.tif 102.tif 103.tif 104.tif 120.tif`

mean 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 ...*”.

##### [Raw ASCII]

Followed by the ASCII files with the full, unedited text of the item (result of Optical Character Recognition). If not supplied explicitly, the extension *.raw* should be added to obtain the full file names.

Example: Both

- `_mf [Raw ASCII] 101 102 103 104 120` and
- `_mf [Raw ASCII] 101.raw 102.raw 103.raw 104.raw 120.raw`

mean that the five files *101.raw*, *102.raw*, *103.raw*, *104.raw* and *120.raw* contain the raw text of the publication item.

##### [SGML Cit]

Followed by the file name which denotes the file which holds the bibliographic data (only title, author(s), abstract, etc.) of the publication item in Standard Generalized Markup Language (SGML) format. If not supplied explicitly, the extension *.sgc* should be added to obtain the full file name. Document Type Definitions (DTD's) which correspond with SGML formats are available separately via the EFFECT Web page or upon request.

Example: Both

- `_mf [SGML Cit] 9601234` and
- `_mf [SGML Cit] 9601234.sgc`

mean that the file *96001234.sgc* contains the SGML formatted bibliographic data of the item.

All files mentioned in the above formats reside in the journal issue directory, for instance */xyz00012/00406090/v0193i01*

## 4.6.2 Manifestations in item-based datasets

Data contained in item-based EFFECT datasets is generally “richer” in nature than that of issue-based datasets. The range of file types is wider and the diverse files offer more information, are structured and are of higher quality. Publication items are broken down and offered in several components. For instance, the text is available in a well-defined structure and the figures are offered as separate files, linked to the text. Also, multimedia components such as video and sound clips –impossible to represent in paper format– are possible to encase.

To enable the inclusion of more filetypes, the manifestation format identifier between square brackets “[ ]” denotes a particular manifestation and specific attributes, details or scheme for this manifestation.

### [SGML *dtd* {*version*}]

This tag specifies material in Standard Generalized Markup Language (*SGML*) format. The attribute *dtd* following **SGML** denotes the particular Document Type Document (DTD) and DTD version to which the SGML codes adhere. The optional *version* attribute indicates the particular version of the DTD.

Note that EFFECT does not specify the location of the DTD; this should be agreed for each project by the partners.

The manifestation format identifier is followed by the file names which hold the SGML code. SGML files are identified in EFFECT datasets by the file extension *.sgm*. At least the file *main.sgm* is available, which is the “starting point” for all other files that “belong” to the publication item. All possible other SGML files in this directory are referenced in the *main.sgm* file. If not supplied explicitly, the extension *.sgm* should be added to obtain the full file names.

Example: Both

- `_mf [SGML ART 3.0.0] main` and
- `_mf [SGML ART 3.0.0] main.sgm`

mean that the file *main.sgm* in the subdirectory mentioned in the last field of `_t3` contains the SGML formatted text of the item. The SGML tagging conforms to the **ART** Document Type Definition version **3.0.0**

**[HTML {version}]**

This tag indicates material coded in HyperText Markup Language (*HTML*) format, the generally accepted “language” of World Wide Web applications. The optional *version* attribute indicates the version of HTML to which the files adhere.

The manifestation format identifier is followed by the file names which hold the HTML code. In EFFECT datasets, HTML files are marked with the file extension *.htm*. At least the file *main.htm* is available, which is the “starting point” for all other files that “belong” to the publication item. All possible other HTML files in this directory are referenced in the *main.htm* file. If not supplied explicitly, the extension *.htm* should be added to obtain the full file names.

Example: Both

- `_mf [HTML 3.2] main sect1 sect2 fn1` and
- `_mf [HTML 3.2] main.htm sect1.htm sect2.htm fn1.htm`

mean that the files *main.htm*, *sect1.htm*, *sect2.htm* and *fn1.htm* in the subdirectory mentioned in the last field of `_t3` contain the HTML version 3.2 formatted text of the item.

Note that, apart from the *main.htm* file, the naming of all other *.htm*-files is *not* specified in EFFECT. The file names in the above and in other examples are only for explanatory purposes.

**[TIFF {version} {type}]**

**[JPEG {version}]**

**[EPS {version}]**

**[GIF {version}]**

These manifestation format identifiers denote artwork files (figures, plates, maps, photographs, etc) in several graphics formats to represent different types of artwork (see also chapter 2.5 at page 9). The optional *version* and *type* attributes specify more detail on the particular graphics format.

The manifestation format identifier is followed by the file names which contain the particular figures. Since artwork files are supplemental to the text, there is no comparable *main* construct such as described in [SGML] and [HTML]. All artwork files are referenced in the SGML or HTML files. Note that the naming of artwork files is *not* further specified in EFFECT. The file names in the above and in other examples are only for explanatory purposes. If not supplied explicitly, the extensions *.tif*, *.jpg*, *.eps* and *.gif* should be added respectively to obtain the full file names.

Some examples:

- Both  
\_mf [TIFF 5.0 FaxIV] fig1 fig2 and  
\_mf [TIFF 5.0 FaxIV] fig1.tif fig2.tif  
signify that the two files *fig1.tif* and *fig2.tif* are Tagged Image File Format *TIFF* files, compressed according to the ITU T.6/Fax Group IV compression scheme.
- Both  
\_mf [GIF 89a] fig3 fig4 and  
\_mf [GIF 89a] fig3.gif fig4.gif  
mark the two files *fig3.gif* and *fig4.gif* as Compuserve GIF files following the 89a version.
- Both  
\_mf [JPEG] fig5 fig6 fig7 and  
\_mf [JPEG] fig5.jpg fig6.jpg fig7.jpg  
denote the three JPEG encoded files *fig5.jpg*, *fig6.jpg* and *fig7.jpg*

[PDF {version} {type}]

Followed by the file names which denote the file in Portable Document Format (*PDF*), The file name extension *.pdf* is used for indicating these files in EFFECT datasets. At least the file *main.pdf* is available, which is the “starting point” for all other files that “belong” to the publication item. If not supplied explicitly, the extension *.pdf* should be added to obtain the full file names.

The optional *version* attribute shows the particular version of PDF used.

The optional *type* attribute is an indicator how particular PDF files where created, information which is not easily recognizable from the “outside”. The following type possibilities exist:

- **WRAPPED** The PDF file is created by “wrapping” one or more scanned page images in a PDF file. Wrapped PDF files lack the text search functionality.
- **INTERPRETED** The PDF file created by both wrapping one or more scanned pages, and an OCR-interpreted text part, typically prepared with Acrobat Capture™. Actually it is a “wrapped” file, with full text search functionality.
- **DISTILLED** When the PDF file is derived from an original PostScript file with Acrobat Distiller™.
- **OPTIMIZED** The PDF file is optimized for faster network transfer and viewing.
- **BOOKMARKED** The PDF file contains bookmarks.

It is possible that more than one type possibility is applied.

A few examples:

- `_mf [PDF 1.1 DISTILLED OPTIMIZED] main.pdf` means that the PDF 1.1 file *main.pdf* in the subdirectory mentioned in the last field of `_t3`, is created by Acrobat Distiller from an original PostScript file, and subsequently is optimized for Internet/Web transfer.
- `_mf [PDF 1.1 WRAPPED] main.pdf` indicates the PDF file *main.pdf* which contains one or more scanned page images wrapped in the PDF format.
- `_mf [PDF 1.1 DISTILLED] main.pdf add1.pdf add2.pdf` denotes three interconnected PDF file, of which *main.pdf* is the starting point.

Some combinations will never occur, like **WRAPPED** and **DISTILLED**. Other type possibilities are never used alone, such as **BOOKMARKED** with **DISTILLED**.

The article at page 29 in an item-based dataset for exchange of material in SGML and related formats:

```
_t3 xyz00012 00406090 v0193i01 96000123
_ii S0040-6090(96)00012-3      the item identifier *)
_ty FLA                        the item type

      ....{part removed for clarity, please compare with page 29}....

_pg 501-504+520                the range of pages as they
                                appear in a citation

_mf [SGML ART 3.0.0] main.sgm   the full text in SGML format in accordance
                                with ARTicle DTD version 3.0.0

_mf [TIFF 5.0 FaxIV] fig1.tif   two black/white line-art TIFF files
                                representing figures 1 and 3
      fig3.tif

_mf [JPEG] fig2.jpg fig4.jpg    three JPEG files: figures 2, 4 and 6
      fig6.jpg

_mf [EPS] fig5.eps              figure 5 is an Encapsulated PostScript file

_mf [PDF 1.1 DISTILLED] main.pdf the Portable Document Format file
                                representing the entire article
```

\*) Note: text in italic only for clarity

The same article in for exchange of material for World Wide Web applications, delivered in HTML 3.2 and artwork in GIF format:

_t3 xyz00012 00406090 v0193i01 96000123	
_ii S0040-6090(96)00012-3	<i>the item identifier *)</i>
_ty FLA	<i>the item type</i>
<i>.....[part removed for clarity, please compare with page 29].....</i>	
_pg 501-504+520	<i>the range of pages as they appear in a citation</i>
_mf [HTML 3.2] main.htm sect1.htm sect2.htm sect3.htm sect4.htm refer.htm fn1.htm fn2.htm fn3.htm appendx1.htm appendx2.htm	<i>the full text in HTML format version 3.2 the text is divided into four sections, a References section, three footnotes and two appendices</i>
_mf [GIF 89a] fig1.gif fig2.gif fig3.gif fig4.gif fig5.gif fig6.gif frmla1.gif frmla2.gif frmla3.gif constr1.gif constr2.gif	<i>a number of GIF files representing figures 1 through 6 three mathematical formulas which cannot be rendered with HTML tags two "other" non-HTML-able constructs</i>
_mf [PDF 1.1 DISTILLED] main.pdf	<i>the Portable Document Format file representing the entire article</i>

\*) Note: text in italic only for clarity

---

## **5 Dataset delivery**

EFFECT datasets can be transferred with “any” medium: CD-ROM, tape or Internet FTP. The standard format described in the previous sections is a “logical” format, not related to a particular medium. It is easily possible to switch to another medium without changing the standard. The layout of a dataset can be represented as a standard MS-DOS or UNIX directory tree. The following assumptions have been made in drawing up the present Technical Specifications:

- The emphasis is on archiving and in transporting large quantities of information.
- The different parts of the information should be easily discernable, so that they can 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, and included in internal database and/or image bank systems.
- The receiving organizations make the data available to their staff using their own hardware/software infrastructure.

Because of the “logical” nature of EFFECT datasets, it is possible that a single dataset is “spanned” across several CD-ROM’s, or even that several datasets are accumulated together in one physical medium. To avoid unnecessary confusion, use of EFFECT in this way is not recommended and should be considered with greatest caution.

### **5.1 Dataset delivery by 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 so-called CD-Recordable (CD-R) device, a write-once CD-player in which single CD-ROMs 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, for instance a MS-DOS PC with the Microsoft CD Extensions (MSCDEX). Fetching material from a CD-ROM simply involves copying from it.

## 5.2 Dataset delivery by file transfer

For network delivery of material, entire EFFECT datasets with their directory structure and files can be collected in so-called container files, which can easily be transported over the Internet or other networks and subsequently disassembled into their original structure at the receiving system. EFFECT does not specify the transfer mechanism for this, only the structuring method is depicted. Several methods exist, which could be applied for a particular project:

- *TAR* is a popular packaging format for UNIX systems, in which entire directory trees are enveloped. File names end with the extension *.tar*, for instance *xyz00012.tar*. A variation exist in which the different files are also decreased in size via the *GZIP* compression scheme. To indicate *GZIP*-ped / *TAR*-red EFFECT datasets, these container file names end with the extension *.tgz*, for instance *xyz00012.tgz*.
- Several combined compression/enveloping schemes exist for MS-DOS such as *PK ZIP* (extension is *.zip*), *ARJ* (extension is *.arj*) and *LZH* (extension is *.lzh*). These may be applied according to the demand of a particular project.

## "@"-symbols

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
@W	Ω Omega
@X	Ξ Xi
@a	α alpha
@b	β beta
@c	γ gamma
@d	δ delta
@e	ε epsilon
@f	φ phi
@g	χ chi
@h	η eta
@i	ι iota
@j	ψ psi
@k	κ kappa
@l	λ lambda
@m	μ mu
@n	ν nu
@p	π pi
@q	θ theta
@r	ρ rho
@s	σ sigma
@t	τ tau
@u	υ upsilon
@w	ω omega
@x	ξ xi
@z	ζ zeta

Code	Symbol
@6	↑ arrow up
@7	↓ arrow down
@8	≡ identical
@K	≈ approximately equal
@/	√ square root
@!	∫ integral
@A	∅ circle with slash, diameter
@+	± plus minus
@&	∞ infinite
@%	‰ promilla
@9	Å Ångström
@\$	£ English pound
@M	♂ male
@V	♀ female
@*	° degree
@'	" double quote
@^	^ caret-symbol
@@	@ at-symbol
@?	any special character not in this list

## Other typographic symbols

Code	Symbol
->	→ arrow right
<-	← arrow left
<->	↔ left over right arrow
=<	≤ smaller or equal
>=	≥ larger or equal
<>	≠ not equal
~	~ similar, varies linearly with
1/4	¼ one quarter
1/2	½ one half
3/4	¾ three quarter

## Appendix B

## Codes in dataset.toc in alphabetical order

<b>_ab</b>	Abstract of publication item	33
<b>_au</b>	Author name	33
<b>_br</b>	Backward reference	36
<b>_ca</b>	Correspondence for publication item	33
<b>_cd</b>	CODEN code	22
<b>_cf</b>	Conference details	25
<b>_ci</b>	City of publisher	22
<b>_cr</b>	Copyright notice	22
<b>_ct</b>	Content pages of journal issue	27
<b>_dt</b>	Date of publication of journal issue	26
<b>_em</b>	Editorial Board member name	22
<b>_et</b>	Editorial Board title	22
<b>_ia</b>	Instructions for authors	22
<b>_if</b>	ISSN of full set	22
<b>_ii</b>	Item identifier	30
<b>_im</b>	Publisher imprint	22
<b>_io</b>	Former journal ISSN	22
<b>_is</b>	Issue number	25
<b>_ja</b>	Journal name (official abbreviation)	22
<b>_jf</b>	Full set journal name	22
<b>_jn</b>	Journal name (full)	22
<b>_jo</b>	Former journal name	22
<b>_kw</b>	Keywords of publication item	34
<b>_la</b>	Language of abstract of publication item	34
<b>_li</b>	Language of the publication item text	33
<b>_mf</b>	Manifestation files of journal issue and of publication item	27, 36
<b>_np</b>	Number of pages of journal issue	26
<b>_pd</b>	Production date of dataset	19
<b>_pg</b>	Page range of publication item	35
<b>_pn</b>	Actual page names of journal issue	26
<b>_pr</b>	Page range of journal issue	25
<b>_pu</b>	Publisher name	22
<b>_st</b>	Structure type of dataset	20
<b>_su</b>	Subject area of publication item	35
<b>_t0</b>	Identification of dataset	19
<b>_t1</b>	Identification of journal title	22
<b>_t2</b>	Identification of journal issue	24
<b>_t3</b>	Identification of publication item	30
<b>_tf</b>	Title of publication item in foreign language	33
<b>_ti</b>	Title of publication item	33

<b>_ty</b>	Type of publication item . . . . .	31
<b>_v1</b>	Volume number of journal issue . . . . .	25
<b>_vn</b>	Version number of dataset . . . . .	19
<b>_xt</b>	Extra information to identify journal issue . . . . .	26

## Appendix C

## Codes in dataset.toc in order of appearance

<b>_t0</b>	Identification of dataset	19
<b>_vn</b>	Version number of dataset	19
<b>_pd</b>	Production date of dataset	19
<b>_st</b>	Structure type of dataset	20
<b>_t1</b>	Identification of journal title	22
<b>_cd</b>	CODEN code	22
<b>_jn</b>	Journal name (full)	22
<b>_ja</b>	Journal name (official abbreviation)	22
<b>_jo</b>	Former journal name	22
<b>_io</b>	Former journal ISSN	22
<b>_jf</b>	Full set journal name	22
<b>_if</b>	ISSN of full set	22
<b>_pu</b>	Publisher name	22
<b>_ci</b>	City of publisher	22
<b>_im</b>	Publisher imprint	22
<b>_et</b>	Editorial Board title	22
<b>_em</b>	Editorial Board member name	22
<b>_ia</b>	Instructions for authors	22
<b>_cr</b>	Copyright notice	22
<b>_t2</b>	Identification of journal issue	24
<b>_v1</b>	Volume number of journal issue	25
<b>_is</b>	Issue number	25
<b>_pr</b>	Page range of journal issue	25
<b>_cf</b>	Conference details	25
<b>_xt</b>	Extra information to identify journal issue	26
<b>_dt</b>	Date of publication of journal issue	26
<b>_np</b>	Number of pages of journal issue	26
<b>_pn</b>	Actual page names of journal issue	26
<b>_ct</b>	Content pages of journal issue	27
<b>_mf</b>	Manifestation Files of journal issue	27

_t3	Identification of publication item . . . . .	30
■ _ii	Item identifier . . . . .	30
■ _ty	Type of publication item . . . . .	31
_li	Language of the publication item text . . . . .	33
_ti	Title of publication item . . . . .	33
_tf	Title of publication item in foreign language . . . . .	33
_au	Author name . . . . .	33
_ca	Correspondence for publication item . . . . .	33
_ab	Abstract of publication item . . . . .	33
_la	Language of abstract of publication item . . . . .	34
_kw	Keywords of publication item . . . . .	34
■ _su	Subject area of publication item . . . . .	35
_pg	Page range of publication item . . . . .	35
_br	Backward reference . . . . .	36
■ _mf	Manifestation files of item . . . . .	36